

**SYNTHÈSE DES PHRASES À EXTRACTION  
EN FRANÇAIS CONTEMPORAIN  
(DU RÉSEAU SÉMANTIQUE À L'ARBRE SYNTAXIQUE)**

**Sylvain KAHANE\* et Igor MEL'ČUK\*\***

**Résumé - Abstract**

Cet article décrit la synthèse de deux types de phrases à extraction en français — les phrases avec des relatives et des interrogatives indirectes. Nous présentons, dans le cadre de la théorie Sens-Texte, le passage de la structure sémantique (un réseau) d'une phrase à extraction à sa structure syntaxique profonde (un arbre de dépendance). La nature des extractions — le caractère discontinu et non local des règles d'extraction — et le fait que notre structure sémantique soit un réseau posent des problèmes linguistiques et procéduraux, que nous résolvons par l'introduction de deux concepts : le nucléus et le pointeur. D'autre part, notre étude apporte un éclairage nouveau sur la formalisation des règles linguistiques de correspondance entre la structure sémantique et la structure syntaxique profonde et sur l'organisation de l'ensemble de ces règles, qui constituent le module sémantique d'un modèle Sens-Texte.

This paper describes the synthesis of two types of sentences containing extractions in Modern French—sentences with a relative or an indirect interrogative. We present, in the Meaning-Text framework, the transition from the semantic structure (a network) of a sentence containing extractions to its deep-syntactic structure (a dependency tree). The fact that our semantic structure is a network and the nature of extractions themselves—specifically, the discontinuity and non-locality of extraction rules—lead to certain linguistic and procedural problems, which we try to resolve by introducing two notions—the nucleus and the pointer. At the same time, our research sheds some new light on the formalization of the linguistic correspondence rules that ensure the transition from the semantic to the deep-syntactic level and on the over-all organization of the set of these rules (= the semantic module of a Meaning-Text model).

**Mots Clefs - Keywords**

Interface sémantique-syntaxe, génération de texte, théorie Sens-Texte, arbre de dépendance, réseau sémantique, formalisme grammatical, transducteur graphe-arbre.

Semantic-syntax interface, text generation, Meaning-Text theory, dependency tree, grammatical formalism, semantic network, network-to-tree transducer.

---

\* TALaNa, Université Paris 7, UFRJL, Case 7003, 2 place Jussieu, 75251 Paris cedex 05, et Univ. Paris 10 - Nanterre (sk@ccr.jussieu.fr).

\*\* Collège de France et Observatoire de Linguistique Sens-Texte, Univ. de Montréal, Département de linguistique et de traduction, C.P. 6128, Centre-Ville, Montréal H3C 3J7, Canada (melcuk@ere.umontreal.ca).

© copyright ATALA

## INTRODUCTION

Les **phrases à extraction** sont les phrases comprenant des subordonnées relatives, interrogatives indirectes, clivées . . . , d'où un élément (en gras ci-dessous) est “extrait” :

- (1) a. *Je prends **le livre que** Jean est en train de lire.*
- b. *Je me demande **quel livre** Jean est en train de lire.*
- c. *C'est **ce livre que** Jean est en train de lire.*

Nous nous concentrerons sur les deux types principaux de phrases à extraction : celles comprenant des relatives avec antécédent (1a) et celles comprenant des interrogatives indirectes (1b).

Le terme “extraction” est hérité des générativistes, pour lesquels les phrases à extraction sont produites, à partir de phrases sous-jacentes, par la transformation dite d'extraction (Chomsky 1977). Dans notre approche, l'extraction n'existe pas en tant que transformation. Nous procédons de façon plus immédiate : une phrase à extraction possède un sens particulier, et elle est synthétisée directement à partir de son sens, c'est-à-dire à partir de sa **représentation sémantique**, sans transformer quoi que ce soit. Néanmoins, nous retenons le terme de phrases à extraction parce qu'il dénote un regroupement de phénomènes syntaxiques assez différents, mais partageant des propriétés importantes (Chomsky 1977) : l'usage des **pronoms subordonnatifs**<sup>1</sup>, l'antéposition de l'élément extrait et les contraintes linguistiques qui bloquent certaines extractions.

Nous montrerons comment, à partir d'une représentation de leur sens, on peut synthétiser les phrases à extraction au moyen d'un ensemble de règles générales. Nous commençons par une description de la correspondance STATIQUE entre les phrases à extraction et leur sens. Cette correspondance n'est pas orientée et peut, au moins en théorie, être utilisée pour la synthèse comme pour l'analyse. Elle est donnée par une liste de règles linguistiques formulées comme bidirectionnelles. Nous ébaucherons ensuite une procédure de synthèse capable de mettre ces règles en jeu.

Notre étude se place dans le cadre général de la théorie Sens-Texte (cf., par exemple, Žolkovskij & Mel'čuk 1967, Mel'čuk 1974, 1988, 1997). C'est pour nous l'occasion de développer des aspects de cette théorie jusqu'ici peu étudiés, notamment le rôle fondamental de la structure communicative (§1.1.2), la structure du module sémantique (§2.1) et l'aspect procédural de la synthèse (§3.1).

---

1. Un pronom subordonnatif est soit un pronom relatif (*l'homme **qui** est venu*), soit un pronom interrogatif direct et indirect (***Qui** est venu ? ; Je veux savoir **qui** est venu*), soit un pronom pseudo-relatif (*J'embrasse **qui** je veux*). On notera que le pronom pseudo-relatif *qui* ne peut pas être considéré comme un pronom relatif, car il joue, au sein de la proposition subordonnée, le rôle d'un complément d'objet direct et que le pronom relatif objet est *que* — comme dans *J'embrasse la personne **que** je veux (embrasser)*. D'autre part, ce “pronom”, à la différence du pronom relatif, est une lexie pleine, comme le pronom interrogatif (cf. §1.1.3).

Nous prenons pour acquise la division des représentations linguistiques en niveaux prévue par la théorie Sens-Texte. La synthèse des textes s'effectue à partir du niveau sémantique en passant par les niveaux syntaxiques (profond et de surface), morphologiques (profond et de surface) et phonologiques (également profond et de surface). A chaque niveau, la phrase possède une **représentation** propre à ce niveau, composée de différentes structures : la **structure centrale** et des **structures périphériques**, qui la précisent. Notre étude se limite aux deux premiers niveaux de représentation : sémantique et syntaxique profond. La structure centrale du niveau sémantique, la **structure sémantique** [= SSém], est un réseau (= un graphe orienté) dont les nœuds sont étiquetés par des **sémantèmes**<sup>2</sup>, tandis que la structure centrale du niveau syntaxique profond, la **structure syntaxique profonde** [= SSyntP], est un arbre de dépendance dont les nœuds sont étiquetés par des **lexies pleines**. Le passage de la représentation sémantique à la représentation syntaxique profonde et vice versa est assuré par un ensemble de règles appelé le **module sémantique** (d'un modèle Sens-Texte).

Ayant précisé notre cadre théorique, indiquons maintenant deux problèmes qui ont particulièrement retenu notre attention.

• **Formulation des contraintes linguistiques.** Les règles de synthèse des phrases à extraction doivent répondre à des contraintes linguistiques complexes. Par exemple, la phrase (2a), bien qu'agrammaticale, correspond à un sens normal qui peut être exprimé en français par (2b-c) :

- (2) a. \* *Je me demande quelle est la personne que tu lis ce livre en regardant.*  
 b. *Tu lis ce livre en regardant une personne ; je me demande laquelle.*  
 c. *Je me demande quelle est la personne que tu regardes en lisant ce livre.*

La phrase (2a) doit donc être interdite au niveau syntaxique profond.

Les contraintes limitant l'extraction ont fait l'objet de nombreuses études dans différentes langues et ont été formalisées dans des cadres d'études variés (Chomsky 1965, 1973, 1977, 1981, Ross 1967, Rizzi 1978, Gazdar & al. 1985, Kroch 1987, Godard 1988, Kaplan & Zaenen 1989, Pollard & Sag 1994, Sag & Fodor 1996, ...). A l'inverse des descriptions en terme de contraintes, notre approche propose une description positive de la distribution des extractions, reposant sur deux concepts linguistiques nouveaux : le **nucléus verbal** et le **nucléus nominal** (§2.3.2).

• **Discontinuité non bornée (dans la correspondance SSém ↔ SSyntP).** Lorsqu'on considère la SSém et la SSyntP d'une même phrase, il y a entre les nœuds de la SSém et les nœuds de la SSyntP — c'est-à-dire entre les sémantèmes et les lexies — une correspondance évidente (les lexies expriment les sémantèmes). Les dépendances sémantiques et syntaxiques se correspondent également en général : quand il y a entre deux sémantèmes une dépendance sémantique, il y a entre les lexies qui les expriment une dépendance syntaxique

---

2. Un sémantème est le signifié d'une lexie, c'est-à-dire une acception lexicographique désambiguïsée.

(SyntP)<sup>3</sup>. Comme on le sait (Mel'čuk 1964, 1988 : 118-28), il existe néanmoins des cas où dépendance sémantique et dépendance syntaxique ne se correspondent pas. On parlera alors de **discontinuité** dans la correspondance SSém ↔ SSyntP. Exemple : dans la phrase *Pierre boit un verre de vin*, la lexie VERRE dépend syntaxiquement de la lexie BOIRE, alors que c'est le sémantème 'vin' qui dépend sémantiquement du sémantème 'boire' (ce que Pierre boit est le vin et pas le verre). Les phrases à extraction présentent également des discontinuités. Dans la phrase *Je cherche le livre que Pierre veut lire*, le verbe principal VOULOIR de la relative dépend syntaxiquement de LIVRE (l'antécédent de la relative), alors qu'il n'y a pas de dépendance sémantique entre 'vouloir' et 'livre'. Pour traiter les discontinuités, nous devons introduire des règles de correspondance de type nouveau, les **règles discontinues**.

De plus, la discontinuité dans les phrases à extraction est **non bornée**<sup>4</sup>, c'est-à-dire que deux nœuds Sém liés par une dépendance DIRECTE se trouvent en correspondance avec deux nœuds SyntP liés par une CHAÎNE de dépendances dont la longueur est potentiellement non bornée : *le livre que Pierre lit*, *le livre que Pierre commence à lire*, *le livre que Pierre veut commencer à lire* ... (où le lien Sém direct entre 'livre' et 'lire' correspond à une chaîne de liens SyntP entre LIVRE et LIRE de plus en plus longue). Les règles d'extraction seront donc des **règles (discontinues) non locales**, c'est-à-dire des règles qui prennent en compte des nœuds liés par une chaîne non bornée de dépendances (sémantiques ou syntaxiques). À l'inverse, une **règle locale** opère localement, c'est-à-dire met en rapport une portion bornée de la SSém avec une portion bornée de la SSyntP. Il y a ainsi trois types majeurs de règles de correspondance envisageables : les **règles continues** (nécessairement locales), les **règles discontinues locales** (non considérées dans cet article) et les **règles discontinues non locales**, que sont les règles d'extraction. La mise en jeu de règles discontinues (et qui plus est, non locales) pose certaines difficultés pour la procédure de synthèse. Nous les résolvons en utilisant le concept algorithmique de **pointeur** (§2.1.1).

L'article est divisé en trois sections : la Section 1 est consacrée à la représentation sémantique et syntaxique des phrases à extraction, la Section 2 aux règles (statiques) de correspondance entre les deux représentations et la Section 3 à la procédure (dynamique) de passage de la représentation Sém à la représentation SyntP.

La présentation s'articulera autour de la phrase de référence suivante :

(3) *Je prends le livre dont Pierre sait quelle page il a arrachée.*

La phrase (3) combine les deux types d'extraction que nous étudions, la relativisation et l'interrogation indirecte. Plus précisément, elle illustre le cas de la relativisation d'un complément d'une proposition interrogative enchâssée

---

3. Dans le cas d'un modifieur, le sens de la dépendance s'inverse entre les niveau Sém et SyntP : dans *un livre rouge*, 'livre' est un dépendant Sém de 'rouge', mais ROUGE est un dépendant SyntP de LIVRE.

4. La discontinuité peut être bornée, comme dans la fameuse construction latine du type *ab urbe condita*, litt. 'à-partir-de [la] ville fondée', qui veut dire 'à partir de la fondation de la ville'.

## SYNTHÈSE DES PHRASES À EXTRACTION

(Pierre sait quelle page il a arrachée **de ce livre**  $\Rightarrow$  le livre **dont** Pierre sait quelle page il a arrachée). On peut donner d'autres exemples de la même construction :

- (4) a. *les affaires que Pierre ne sait pas où ranger*  
b. *une amie à qui je me demande ce que Pierre a donné*  
c. *un navire sur lequel on ne sait quel marin accepterait de s'embarquer*  
d. *des arguments dont Pierre se demande lequel tu peux conserver.*

Les phrases de ce type illustrent bien le croisement des deux phénomènes qui nous intéressent.

Pour faciliter la lecture de notre texte, nous donnons dès maintenant la liste des abréviations et notations que nous utiliserons par la suite :

### Abréviations et notations

Comm	: communicative
DEC	: <i>Dictionnaire Explicatif et Combinatoire</i> (= lexique Sens-Texte)
FR	: focus rhématique
$\mathcal{L}$	: langue donnée
L	: lexie
L( $\langle s \rangle$ )	: lexie dont le signifié est le sémantème $\langle s \rangle$
-P	: profond
R	: règle de correspondance linguistique
<b>R</b>	: rhème (marqueur communicatif de)
R	: relation syntaxique profonde
r	: relation sémantique
R-	: représentation
S-	: structure
-S	: de surface
<b>T</b>	: thème (marqueur communicatif de)
$\langle s \rangle$	: sémantème ou configuration de sémantèmes
$\langle \mathfrak{s} \rangle$	: nœud communicativement dominant d'un sous-réseau sémantique
Sém	: sémantique
Synt	: syntaxique
$\equiv$	: équivalence d'éléments linguistiques d'un même niveau
$\Leftrightarrow$	: correspondance entre éléments linguistiques de deux niveaux adjacents

## 1. LA REPRÉSENTATION SÉMANTIQUE ET SYNTAXIQUE PROFONDE DES PHRASES À EXTRACTION

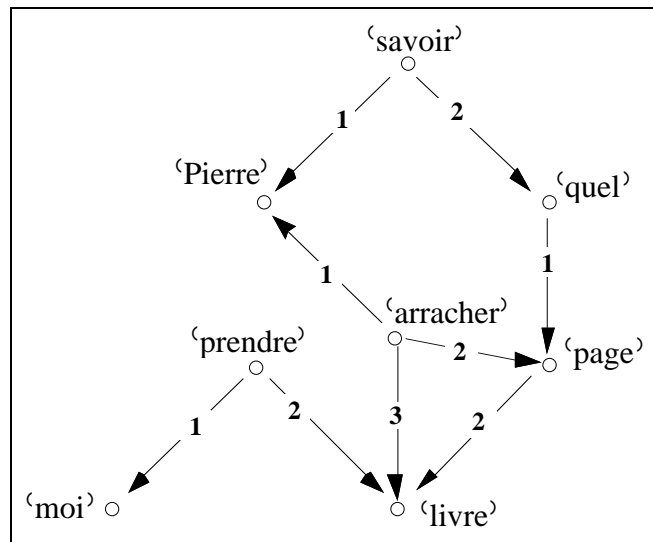
### 1.1. La représentation sémantique

Nous allons étudier la structure sémantique, qui est la composante centrale de la représentation sémantique (§1.1.1), et la structure communicative sémantique, qui en est une composante périphérique, primordiale dans notre optique (§1.1.2). Nous nous intéresserons ensuite à la façon de représenter le sens des pronoms interrogatifs (§1.1.3).

### 1.1.1. La structure sémantique

Dans un modèle Sens-Texte, l'outil central de la description du sens langagier est un objet formel appelé structure sémantique. Une SSém vise le sens propositionnel d'une phrase, c'est-à-dire cette partie de son contenu sémantique qui reflète un état des choses dans le monde et qu'on peut décrire par des propositions logiques. (Le sens propositionnel s'oppose au sens communicatif, voir §1.1.2.)

Comme nous l'avons déjà dit, une SSém est formellement un **réseau** dont les nœuds sont étiquetés par des **sémantèmes** de la langue et dont les arcs représentent des relations prédicat-argument et sont numérotés de façon à distinguer les différents arguments d'un même sémantème prédictif. La SSém de la phrase de référence (3) est donnée dans la Fig. 1. (Attention : il s'agit d'une SSém simplifiée : nous n'avons pas indiqué les sens flexionnels — le temps-mode-aspect des verbes, ni le nombre et la détermination des noms. Ces éléments sémantiques ne sont pas pertinents pour notre étude.)



**Figure 1. La SSém de la phrase de référence (3)**  
*Je prends le livre dont Pierre sait quelle page il a arrachée*

La bonne formation d'une SSém de la langue L est contrôlée principalement par le lexique de L. Celui-ci indique, entre autres, pour chaque sémantème prédictif  $\langle s \rangle$ , exprimé par la lexie  $L(\langle s \rangle)$ , quelle est sa **valence**, c'est-à-dire le nombre de ses **actants sémantiques** (= arguments du prédicat correspondant) et pour chacun d'eux son rôle vis-à-vis du prédicat. Ainsi, dans la Fig. 1 :

- les sémantèmes  $\langle \text{prendre} \rangle$  et  $\langle \text{savoir} \rangle$  sont des prédicats binaires :  $\langle X \text{ prend } Y \rangle$  et  $\langle X \text{ sait que } Y \rangle$  ;
- le sémantème  $\langle \text{arracher} \rangle$  est un prédicat ternaire :  $\langle X \text{ arrache } Y \text{ de } Z \rangle$  ;
- le sémantème  $\langle \text{page} \rangle$  est un quasi-prédicat binaire :  $\langle [\text{quelque chose qui est}] \text{ une page de } Y \rangle$ . (Nous traitons le sémantème  $\langle \text{page} \rangle$  à l'instar de tout ce qui est partie de quelque chose —  $\langle X \text{ est une partie de } Y \rangle$  —, sauf que le sémantème

‘page’ ne réalise pas son premier argument : ce dernier est ici “incorporé” dans le prédicat.)

Pour chaque sémantème (qui n’est pas une primitive sémantique), le lexique d’un modèle Sens-Texte donne sa **décomposition sémantique** : une expression du sémantème par une configuration de sémantèmes plus simples que lui. La correspondance entre un sémantème et sa décomposition sémantique est donnée par une **règle de décomposition sémantique**. Il s’agit d’une correspondance entre deux éléments de même niveau, notée  $\equiv$  ; la correspondance entre des éléments de niveaux adjacents est notée  $\Leftrightarrow$ .

Par exemple, la Fig. 2 représente la décomposition du sémantème ‘arracher’, que l’on peut verbaliser de la façon suivante :

- (5) ‘X arrache Y de Z’  $\equiv$  ‘X, en tirant sur Y, détache Y de Z, ce qui porte atteinte à Z’

On parle d’**expansion** si on utilise une règle de décomposition sémantique pour remplacer un sémantème par sa décomposition, et de **réduction** dans le cas inverse.

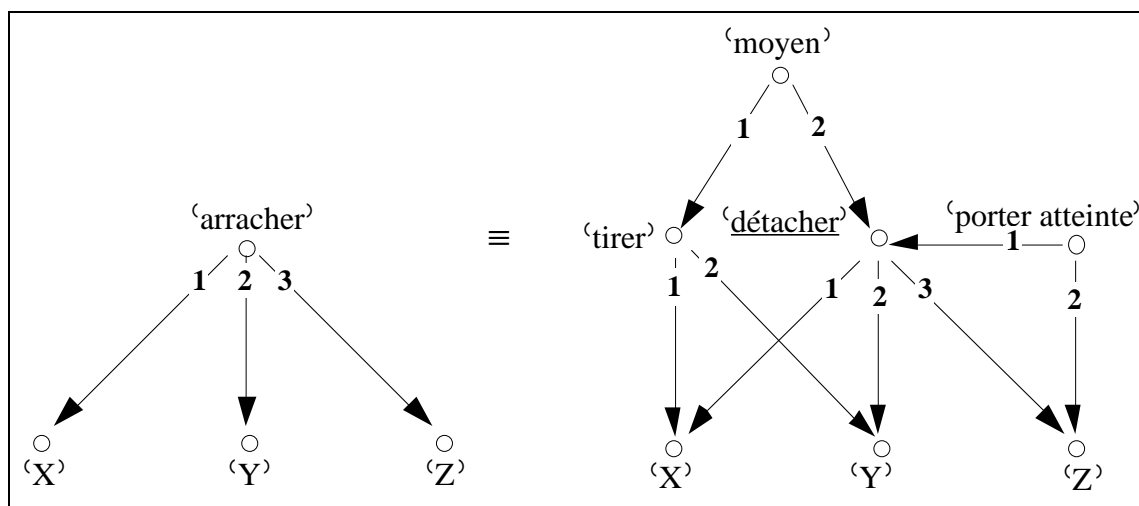


Figure 2. La règle de décomposition sémantique du sémantème ‘arracher’

[Le soulignement du nœud ‘détacher’ dans la partie droite de la règle indique que c’est le nœud communicativement dominant du sous-réseau correspondant, c’est-à-dire le sens auquel le sous-réseau peut se réduire (voir §1.1.2 et Annexe A).]

Chaque règle de décomposition sémantique constitue une partie importante de l’article de dictionnaire d’une lexie ; plus précisément, c’est la définition lexicographique de cette lexie. Un dictionnaire spécialisé contenant de tels articles, appelé *Dictionnaire Explicatif et Combinatoire* [= DEC], est en cours de développement, notamment pour le français (Mel’čuk *et al.* 1984, 1988, 1992, 1999 et Mel’čuk, Clas & Polguère 1995) ; il constitue le lexique d’un modèle Sens-Texte.

Les règles de décomposition sémantique définissent une **relation d’équivalence sur les SSém** (et par extension sur les représentations sémantiques) :

deux SSém sont équivalentes si on peut passer de l'une à l'autre par l'application de règles de décomposition sémantique utilisées dans le sens de l'expansion comme de la réduction. La mise en jeu de ces règles, qu'on peut appeler des **règles sémantiques d'équivalence**, est contrôlée par des **métarègles d'équivalence sémantique** (voir Annexe A). En appliquant les règles sémantiques d'équivalence dans le sens de l'expansion, on obtient un réseau plus décomposé ; en les appliquant dans le sens de la réduction, on obtient un réseau moins décomposé. L'expansion peut être poussée, en principe, jusqu'aux primitives sémantiques. La réduction, elle, est limitée par le lexique de la langue en question : par exemple, en français, 'faire manger' peut être réduit à 'nourrir', mais 'faire boire' n'admet pas de réduction (par contre, en russe, on aurait pu réduire 'faire boire' à un seul sémantème, puisque le russe possède le verbe POIT' 'faire boire'). Parmi les réductions sémantiques, on distingue les **réductions obligatoires** et les **réductions optionnelles**. La plupart des réductions sont obligatoires : elles doivent être effectuées si aucun facteur sémantique ou syntaxique ne les en empêche ; par exemple, la SSém 'X, en tirant sur Y, détache Y de Z, ce qui porte atteinte à Z' doit être réduite en 'X arrache Y de Z' avant le passage à la représentation SyntP. Dans les cas où, dans la langue, une lexie coexiste avec une expression synonyme plurilexicale, les réductions des SSém sont optionnelles : par exemple 'le gendre de X' et 'le mari de la fille de X', 'nourrir' et 'faire manger', ... (voir également le cas des pronoms interrogatifs, §1.1.3).

Au sein d'un ensemble de SSém équivalentes, on isole les **SSém réduites** (notion dont une première version est donnée dans Polguère 1990 : 93) : la SSém d'une phrase donnée est dite *réduite* si aucune réduction sémantique obligatoire n'est possible sur cette SSém. Seules les SSém réduites sont admises à l'entrée du mécanisme de passage à la représentation SyntP. Ceci permet de bloquer les phrases inacceptables du type (6b) et de leur préférer des phrases du type (6a) :

- (6) a. *Pierre a arraché une page de ce livre.*  
b. *Pierre, en tirant sur une page du livre, l'a détachée du livre, ce qui a porté atteinte au livre.*

Une SSém comme telle ne vise pas à refléter le sémantisme complet d'une phrase de façon non ambiguë. Sa cible n'est que l'invariant langagier de toutes les paraphrases jugées suffisamment équivalentes du point de vue de leur contenu propositionnel. Le reste du contenu sémantique de la phrase doit être pris en charge par la structure communicative (sémantique). En particulier, la SSém de la Fig. 1 est insuffisante pour assurer uniquement la production de la phrase (3) et des phrases vraiment synonymes. En effet, à partir de cette SSém, on obtient également des phrases comme (7), qui ne sont pas synonymes de (3) :

- (7) a. *Une page a été arrachée par Pierre du livre que je prends, et il sait laquelle.*  
b. *Je prends le livre dont une page a été arrachée par Pierre (il sait laquelle).*

Pour bloquer la production de phrases non synonymes à partir d'une même



SSém, cette dernière doit être complétée par une structure communicative sémantique.

### 1.1.2. La structure communicative sémantique

La **structure communicative sémantique** [= SComm-Sém] présente l'organisation communicative du sens d'une phrase, c'est-à-dire cette partie de son contenu sémantique qui reflète la façon dont le locuteur veut gérer la communication du sens propositionnel (par où il veut commencer, ce qu'il veut souligner ou, au contraire, atténuer, ce qu'il veut présenter comme connaissance commune avec son interlocuteur ...).

Formellement, une SComm-Sém pour une SSém donnée est constituée d'un système de **sous-réseaux communicatifs** sur cette SSém et de **marqueurs communicatifs** associés à ces sous-réseaux.

#### Sous-réseaux communicatifs

Nous appelons **sous-réseau communicatif** toute portion connexe d'une SSém que le locuteur souhaite considérer comme un tout. Il peut s'agir ou bien d'une SSém complète, c'est-à-dire d'un réseau Sém que le locuteur a l'intention de réaliser comme un énoncé indépendant (une phrase ou un syntagme autonome, comme par exemple un titre), ou bien d'une portion de la SSém que le locuteur veut affubler d'un marqueur communicatif. Un sous-réseau communicatif peut correspondre soit à une proposition (= sous-réseau propositionnel), soit à un syntagme non propositionnel (= sous-réseau non propositionnel).

Tout sous-réseau communicatif non propositionnel a un nœud particularisé, qu'on appelle le **nœud (communicativement) dominant** du sous-réseau. Le nœud dominant d'un (sous-)réseau S est le nœud auquel on peut réduire S de telle sorte que l'information véhiculée par S ne soit pas déformée, mais seulement appauvrie; en termes métaphoriques, le nœud dominant "résume" le contenu de son réseau (Polguère 1990: 46). On dit que le nœud dominant d'un réseau **domine** les autres nœuds de ce réseau, ce qui définit une relation de **dominance communicative**. Le nœud dominant d'un (sous-)réseau est indiqué par un soulignement de son étiquette. En raison du recoupement des sous-réseaux communicatifs, il s'avère nécessaire, dans certains cas, de préciser à quel sous-réseau renvoie le marquage d'un nœud comme nœud dominant: ce sera assuré par un indice rattaché au soulignement (voir Fig. 4 et sv).

L'indication du nœud dominant d'un sous-réseau communicatif permet de spécifier le sens voulu de façon précise: ainsi, 'brûler'  $-_1 \rightarrow$  'maison' donne (a) *l'incendie d'une maison*, alors que 'brûler'  $-_1 \rightarrow$  'maison' donne (b) *une maison en flammes* ou *une maison qui brûle*. On voit que les expressions (a) et (b) ne sont pas synonymes, même si elles portent le même sens propositionnel<sup>5</sup>. Notons que le même réseau Sém ('brûler'  $-_1 \rightarrow$  'maison') caractérisé comme sous-réseau propositionnel (et par conséquent muni d'un division thématique) donne la proposition (c) *Une maison brûle*.

5. Une mise en garde: nous n'avons pu éviter l'emploi de l'adjectif *propositionnel* dans deux sens différents: un sens "sémantique" — 'ce qui est exprimable par des propositions au sens logique' (*sens propositionnel*) —, et un sens "syntaxique" — 'ce qui est exprimable par une proposition au sens syntaxique' (*sous-réseau propositionnel*).

Les exemples précédents mettent en évidence un lien étroit entre le nœud dominant d'un sous-réseau communicatif et tout arbre SyntP correspondant à ce sous-réseau : le nœud dominant d'un sous-réseau communicatif donne forcément naissance au **sommet** de l'arbre SyntP correspondant à ce sous-réseau. Par contre, pour un sous-réseau propositionnel, qui n'a pas de nœud dominant, il y a plusieurs nœuds Sém qui peuvent chacun donner naissance au sommet de l'arbre SyntP correspondant. Un tel nœud s'appelle un **nœud d'entrée potentiel** (Polguère 1990 : 165). Le nœud d'entrée d'un (sous-)réseau Sém est soit son nœud dominant, soit, si ce réseau n'a pas de nœud dominant, un nœud déterminé par des règles à partir des nœuds dominants de sa subdivision thématique. (Des règles pour la détermination du nœud d'entrée dans une SSém ont été proposées dans Iordanskaja & Polguère 1988, Iordanskaja 1990 et Polguère 1990. Nous dirons quelques mots, dans la Section 3, sur la façon dont ces règles s'articulent avec les autres règles du sous-module sémantique de correspondance.)

### Marqueurs communicatifs : thémativité

Les marqueurs communicatifs sont les valeurs des différentes oppositions communicatives (Mel'čuk 1999). Comme nous ne caractériserons pas la SComm-Sém en toute généralité, nous nous limiterons à la seule opposition communicative vraiment pertinente dans le cas des phrases à extraction : la thémativité.

Dans chaque SSém que le locuteur veut réaliser comme une proposition complète, on distingue une portion qui constitue la communication à proprement parler (ce qui est DIT), à savoir le **rhème**, et la partie sur laquelle porte le rhème (CE DONT on parle), à savoir le **thème**, notés respectivement **R** et **T**. Bien que ce soient des notions très étudiées en linguistique, elles restent relativement vagues. Nous allons, pour cette raison, éclaircir les cinq points suivants.

1) La division de la SSém en des sous-réseaux communicatifs de rhème et thème n'est pas une véritable partition : les domaines du rhème et du thème peuvent avoir une partie commune (dans la SSém de la phrase *Pierre veut qu'on lui donne une pomme*, le sémantème 'Pierre' est le thème et fait partie du rhème) ; le thème peut être absent (*Il est venu beaucoup de monde*, où la phrase entière exprime un rhème) ; certains sémantèmes — les **spécifieurs** — peuvent ne faire partie ni du thème, ni du rhème (dans la SSém de la phrase *Hier, ma mère a reçu une lettre*, 'hier' est un spécifieur, mais pas dans celle de la phrase *Cette lettre est arrivée hier*, où il appartient au rhème).

2) La division de la SSém en rhème et thème est récursive : on peut avoir à l'intérieur d'un thème **T<sub>1</sub>** primaire, comme d'un rhème **R<sub>1</sub>** primaire, une nouvelle division en rhème-thème secondaires (**T<sub>2</sub>-R<sub>2</sub>**), et ainsi de suite. A chaque division thématique de la SSém, c'est-à-dire aux sous-réseaux communicatifs correspondants, est associé son **degré (communicatif)**, qui mesure son degré d'enclassement : la division thématique primaire a le degré 1, la division thématique secondaire le degré 2 ... Le degré des rhème et thème joue un rôle important dans la formulation des règles d'extraction (cf. §2.3.2).

3) Nous considérons chaque proposition comme la réponse à une question

## SYNTHÈSE DES PHRASES À EXTRACTION

implicite sous-jacente. Si cette question est spécifique, la partie du rhème constituant l'information requise par cette question est appelée le **focus rhématique** (ou **FR** en abrégé). Par exemple, la phrase *Pierre gagne 100 000 F par an* comme réponse à la question sous-jacente *Combien gagne Pierre par an ?* aura pour rhème 'gagner 100 000 F par an' et pour FR '100 000 F'. Le FR ne coïncide pas nécessairement avec le nœud dominant du rhème (dans l'exemple précédent le nœud dominant du rhème est 'gagner'). Nous reviendrons sur le FR dans le §1.2.3.

4) Deux divisions thématiques différentes d'une même SSém peuvent avoir la même réalisation syntaxique. Par exemple, dans la SSém 'Pierre sait quelle page il a arrachée', 'Pierre' peut être le thème, le reste du sens constituant le rhème (la question sous-jacente étant alors *Qu'y a-t-il à propos de Pierre ?*); le locuteur peut également prendre comme thème 'Pierre sait', le rhème étant 'quelle page il a arrachée' (la question sous-jacente étant *Qu'y a-t-il à propos des connaissances de Pierre ?*). Ces deux divisions thématiques différentes conduisent à la même phrase *Pierre sait quelle page il a arrachée* (à la prosodie près). Dans la Fig. 8, nous avons donné à cette proposition la dernière structure thématique, qui nous paraissait plus naturelle.

5) Un (sous-)réseau propositionnel répond aux deux propriétés suivantes : 1) il possède une division thématique — en rhème, thème et spécifieurs — et cette division recouvre nécessairement l'ensemble du réseau ; 2) il ne peut pas avoir de nœud dominant, car le sens d'une proposition ne peut pas, en règle générale, être résumé par un seul sémantème (ce sont ses rhème et thème qui ont des nœuds dominants).

Montrons maintenant, en utilisant deux exemples, comment on peut distinguer à l'aide de l'opposition de thémativité certaines phrases non strictement synonymes ayant la même SSém.

### Exemple 1

- (8) **a.** *Les linguistes qui étaient fatigués se sont arrêtés.*  
**b.** *Les linguistes qui se sont arrêtés étaient fatigués.*

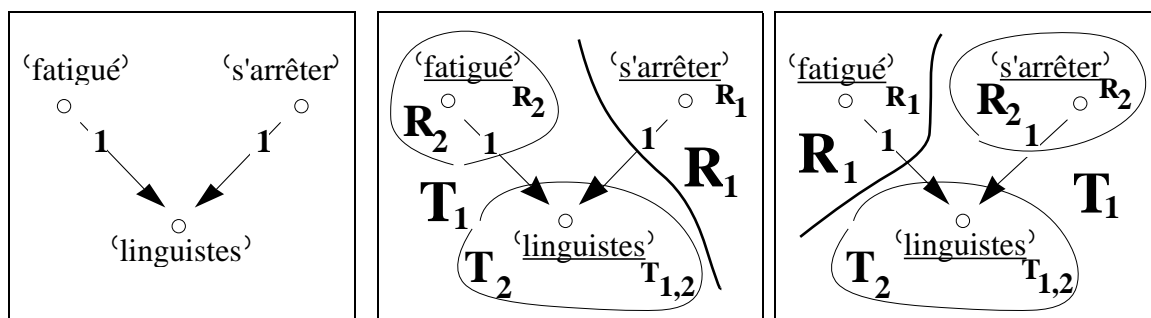


Figure 3. SSém de (8) Figure 4. RSém de (8a) Figure 5. RSém de (8b)

Ces deux phrases possèdent la même SSém (= le même sens propositionnel)<sup>6</sup> (Fig. 3), mais des SComm-Sém différentes. Remarquons que, dans la Fig. 4, bien que le thème  $T_1$  ait le même domaine que le sous-réseau propositionnel  $T_2 \sim R_2$ ,  $T_1$  n'est pas lui-même un sous-réseau propositionnel : il est exprimé par un syntagme non propositionnel *les linguistes qui étaient fatigués* et possède donc un nœud dominant ('linguistes'). Même chose dans la Fig. 5.

### Exemple 2

- (9) a. *Je me demande qui tu as invité que je connais.*  
 b. *Je me demande qui je connais que tu as invité.*

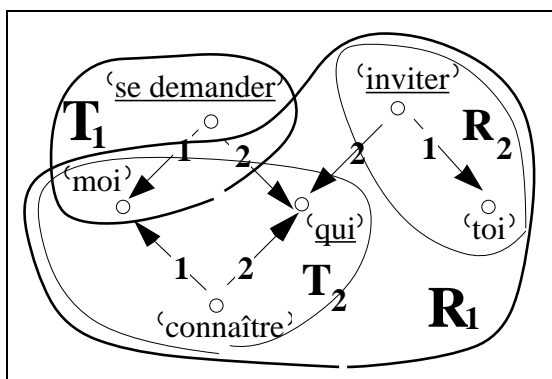


Figure 6. La RSém de (9a)

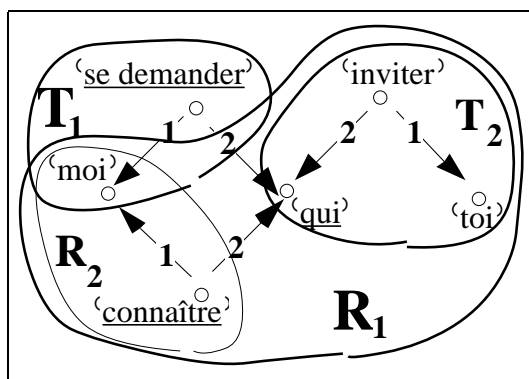


Figure 7. La RSém de (9b)

De façon analogue à (8), les deux phrases (9) ont la même SSém ; de plus, elles ont la même distribution de thème  $T_1$  et rhème  $R_1$  primaires. Cependant, elles ne sont pas synonymes, ce qui est capturé en spécifiant les thème  $T_2$  et rhème  $R_2$  secondaires à l'intérieur de  $R_1$  : dans (9a), le thème  $T_2$  sera 'qui [personne] que je connais' et le rhème  $R_2$  'tu as invité', alors que pour (9b), le thème  $T_2$  sera 'qui [personne] que tu as invité' et le rhème  $R_2$  'je connais' (cf. Fig. 6 et 7 ; la division thématique tertiaire — de la relative —, n'a pas été indiquée).

Nous pouvons maintenant donner la représentation sémantique de la phrase (3) avec tous les éléments thématiques nécessaires. La SComm-Sém de (3) possède trois divisions thématiques enchâssées correspondant aux trois propositions de la phrase (3) : la principale, l'interrogative indirecte et la relative.

a) La division thématique primaire ( $T_1 \sim R_1$ ) recouvre la SSém de la phrase entière. Elle assure la production des phrases où le thème est 'moi' et le rhème englobe le reste, donc les phrases qui incluent *Je prends le livre ...* Avec une autre distribution de rhème et thème primaires, on pourrait obtenir, par exemple, *Le livre ... est pris par moi.*

b) La division thématique secondaire (du rhème  $R_1$  en  $T_2 \sim R_2$ ) concerne la proposition relative. Elle assure la production des phrases où le thème de

6. Les deux phrases de (8) diffèrent évidemment par leurs référents : (8a) réfère à un groupe de linguistes fatigués et (8b) à un groupe de linguistes qui se sont arrêtés. Ceci est encodé dans la Structure Référentielle, dont nous éviterons de parler ici.

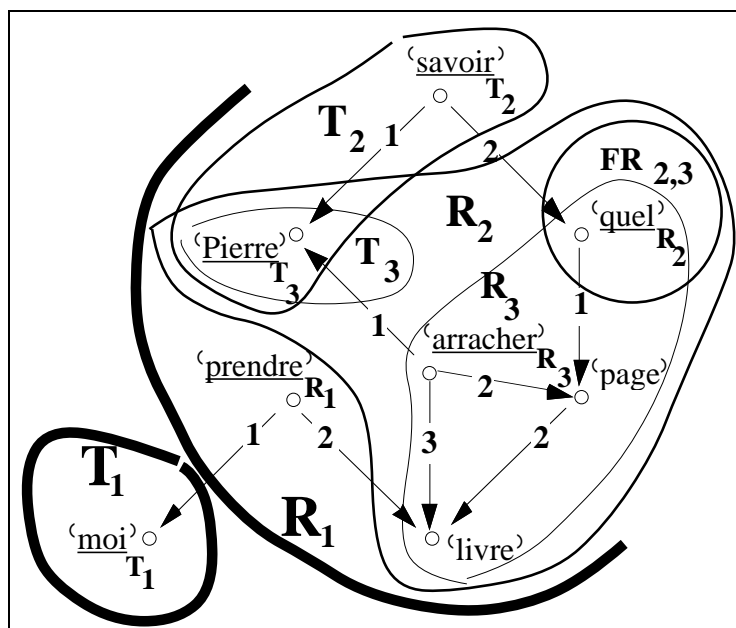


Figure 8. La SSém+SComm-Sémde la phrase de référence (3)

*Je prends le livre dont Pierre sait quelle page il a arrachée*

la relative est «Pierre sait» et son rhème inclut le contenu du savoir, donc les phrases comprenant ... *le livre dont Pierre sait ...* Avec une autre distribution de rhème et thème secondaires, on pourrait obtenir, par exemple, ... *le livre dont Pierre a arraché une page, et il sait laquelle*. De plus, à l'intérieur de  $R_2$ , nous indiquons que «quel» en est le focus rhématique  $FR_2$ .

Notons que la différence entre les propositions relatives et les tournures participiales qui leur sont assez synonymes (*le garçon qui lit un livre* vs *le garçon lisant un livre*; *le livre que Pierre a volé/le livre qui a été volé par Pierre* vs *le livre volé par Pierre*) est aussi encodée par le biais de la thémativité : la RSém d'une relative possède la division thématique prévue pour toute proposition, alors que la RSém d'une tournure participiale ne possède pas de division thématique.

c) La troisième division (du rhème  $R_2$  en  $T_3 \sim R_3$ ) concerne la proposition interrogative indirecte. Elle assure la production des phrases où le thème de l'interrogative indirecte est «Pierre» et son rhème est «a arraché ...», donc les phrases qui contiennent *quelle page il a arrachée*. Avec une autre distribution de thème et rhème tertiaires, on pourrait obtenir, par exemple, *quelle page a été arrachée par lui*.

### 1.1.3. Les pronoms interrogatifs dans la structure sémantique

Nous allons maintenant nous intéresser à la description, dans la structure sémantique, des pronoms interrogatifs, qui sont centraux dans toute étude des interrogatives indirectes. Indiquons, pour commencer, que les traitements sémantiques respectifs des pronoms interrogatifs et des pronoms relatifs sont très différents. Un pronom relatif n'a pas de sens en lui-même, mais remplace un nom dans des contextes particuliers : c'est un vrai PRO-nom ; par conséquent, il n'apparaît que dans la structure syntaxique de surface et les structures plus

surfaciques. Le pronom interrogatif, lui, a un sens : c'est un lexème plein, qui est représenté dans la SSém par son sémantème. Comme les pronoms interrogatifs (directs et indirects) ont des propriétés très spéciales, cause de nombreuses difficultés pour toutes les tentatives de description, l'étude de leur sémantisme s'impose.

Le sens d'un pronom interrogatif peut être représenté par la combinaison de deux éléments : l'un est le sémantème 'quel', qui donne au pronom interrogatif son caractère particulier ; l'autre est le sémantème caractérisant la classe d'entités sur laquelle porte 'quel'. Ainsi, 'qui' se décompose en 'quelle personne', 'quoi' en 'quelle chose', 'combien' en 'quelle quantité', 'quand' en 'à quel moment', 'où' en 'à quel endroit' ... Cette façon de faire est étayée par le fait que les deux expressions – le pronom interrogatif et sa décomposition sémantique – sont substituables dans le texte :

- (10) a. *Qui* (= *Quelle personne*) attends-tu ?  
 b. *Combien* (= *Quelle quantité*) d'argent as-tu dépensé ?  
 c. *Où* (= *A quel endroit*) penses-tu passer tes vacances ?

La Fig. 9 présente les décompositions proposées de façon formelle.

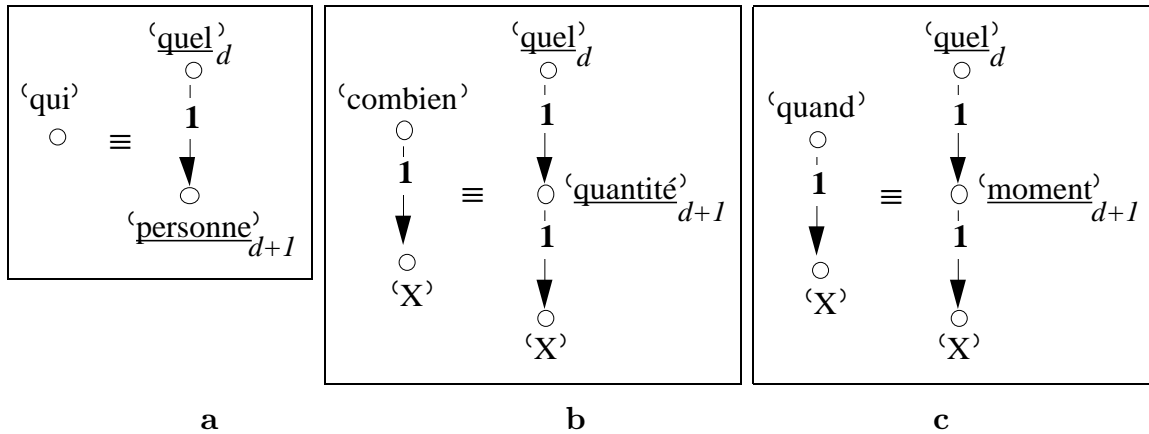


Fig. 9. Les décompositions sémantiques de 'qui', 'combien' et 'quand'

Les pronoms interrogatifs ont une propriété tout à fait exceptionnelle : la SComm-Sém de leur sens contient DEUX nœuds communicativement dominants. En effet, le choix du sémantème qui "résume" leur contenu dépend du point de vue. Prenons le cas de la décomposition de la Fig. 9a. Ici, le contenu du réseau 'quelle personne' se résume par 'quel' vis-à-vis du prédicat interrogatif régissant : 'je me demande quelle personne est venue'  $\approx$  'je me demande quel [= laquelle]'; par contre, le contenu du même réseau est résumé par 'personne' vis-à-vis du prédicat le régissant au sein de la proposition interrogative : 'quelle personne est venue'  $\approx$  '[une] personne est venue'. Autrement dit, dans le réseau 'quelle personne', 'quel' est le nœud dominant de degré  $d$ , tandis que 'personne' est le nœud dominant de degré  $d + 1$ , où  $d + 1$  est le degré de la division thématique de la proposition interrogative, et  $d$  est le degré de la division thématique de la proposition dans laquelle est enchâssée l'interrogative (voir Fig. 10 ; pour les

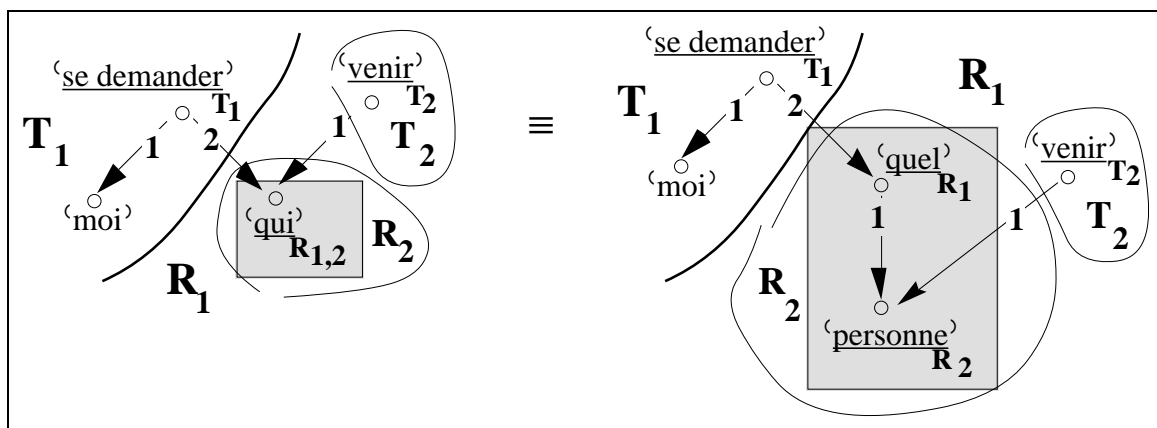


Figure 10. Application de la règle de décomposition sémantique Fig. 9a  
*Je me demande qui/quelle personne est venu(e).*

métarègles qui gèrent l'application des règles de décomposition, voir Annexe A).

Considérons maintenant de plus près le sémantème 'quel' commun à tous les pronoms interrogatifs. On serait tenté de dire qu'il porte l'interrogation; cependant, ce qu'on appelle le pronom interrogatif n'est pas en lui-même vraiment interrogatif, comme l'indiquent des emplois de QUI tels qu'en (11) :

- (11) a. *Pierre m'explique avec qui tu pars.*  
 b. *Pierre me montre avec qui tu pars.*  
 c. *Pierre se fout d'avec qui tu pars.*

La valeur interrogative des expressions avec un pronom interrogatif est en fait portée soit par le verbe régissant (*Je te **demande** avec qui tu pars.*), soit par la prosodie et/ou la construction syntaxique (*Avec qui pars-tu?*). En conséquence, le sémantème 'quel X' équivaut plutôt à un sens neutre: 'information sur X'. Mais de quel type d'information s'agit-il? A la question *Qui attends-tu?*, il y a deux types de réponses possibles: une réponse identifiant la ou les personnes attendues (*J'attends Marie et Jean*), ou bien une réponse identifiant les propriétés de la personne attendue plutôt que la personne elle-même (*J'attends une personne qui doit m'aider*). Comme les deux types de réponse peuvent être coordonnées (*J'attends Marie et quelqu'un qui doit m'aider*), on en déduit que le sémantème 'quel' est disjonctif; c'est donc, d'après le critère de Green-Appresjan<sup>7</sup>, un seul sens comportant une disjonction. Nous proposons alors pour 'quel' — en première approximation — la décomposition sémantique suivante :

7. Le **critère de Green-Appresjan** est un critère lexicographique permettant de déterminer la nécessité de réunir deux sens 's<sub>1</sub>' et 's<sub>2</sub>' sous une seule lexie (dont le signifié sera 's<sub>1</sub> ou s<sub>2</sub>'). Le critère prescrit la réunion de deux sens 's<sub>1</sub>' et 's<sub>2</sub>' dans une seule définition lexicographique d'une lexie L si L peut être combinée simultanément avec deux lexies L<sub>1</sub> et L<sub>2</sub> telles que L<sub>1</sub> porte sémantiquement sur 's<sub>1</sub>' et L<sub>2</sub> sur 's<sub>2</sub>'. Exemple: SE REFROIDIR 'devenir plus froid [= 's<sub>1</sub>'] ou froid [= 's<sub>2</sub>']; on peut dire *Le liquide se refroidit d'abord un peu* [= L<sub>1</sub>, porte sur 'plus froid'] et puis *complètement* [= L<sub>2</sub>, porte sur 'froid']. Voir Mel'čuk, Clas & Polguère 1995: 64-66.

(12) «quel X»  $\approx$  «information sur l'identité de X ou sur les propriétés caractéristiques de X»

Cette décomposition peut être illustrée par les exemples suivants :

- (13) a. «Tu me diras quel livre tu as lu»  $\equiv$  «Tu me diras l'information sur l'identité du livre que tu as lu»  
b. «Je me demande quel livre tu as lu»  $\equiv$  «Je me demande l'information sur l'identité du livre que tu as lu»  
c. «Je sais quel livre tu as lu»  $\equiv$  «Je sais l'information sur l'identité du livre que tu as lu»

(Attention : ces substitutions ne sont valables qu'au niveau sémantique. Dans le texte véritable, l'expression *je me demande l'information sur ...* viole les règles de syntaxe et de cooccurrence lexicale ; néanmoins, le sens «je me demande l'information sur ...» est bien formé et peut être exprimé en français standard par, par exemple, *j'essaye de me souvenir de l'information sur ...*)

Cependant, on trouve des contextes où la décomposition (12) ne passe pas :

- (14) a. «J'ai vu quel livre il a acheté»  $\neq$  «J'ai vu l'information sur l'identité du livre qu'il a acheté»  
b. «Je te montrerai quel livre j'ai lu»  $\neq$  «Je te montrerai l'information sur l'identité du livre que j'ai lu»  
c. «J'ai remarqué quel livre il préfère»  $\neq$  «J'ai remarqué l'information sur l'identité du livre qu'il préfère»

Pour de tels contextes, où on a un prédicat de perception, la décomposition suivante est meilleure : «Je te montrerai quel livre j'ai lu»  $\equiv$  «Je te montrerai le livre que j'ai lu et par conséquent tu auras l'information sur l'identité de ce livre». Comme on le voit, ici «quel» a un sens légèrement différent. Néanmoins, il ne s'agit pas d'une autre acception de «quel» : comme le montre, encore une fois, le critère de Green-Apresjan (*Je sais et je te montrerai quel livre il a acheté*), cet «ajout» sémantique doit constituer un élément disjonctif au sein d'une même définition. Finalement, nous proposons comme définition de «quel» la décomposition suivante :

(15) «quel X»  $\approx$  «(... et par conséquent le percepteur aura) information sur l'identité de X ou sur les propriétés caractéristiques de X»

Le sens de «quel» a, pour ainsi dire, deux facettes : l'une, tournée vers les verbes informationnels (*savoir, demander, s'expliquer, s'en foutre, ...*), est manifestée par la composante «information» ; l'autre, assurant l'interaction avec les verbes de perception («voir», «remarquer», «montrer», ...), sous la forme de l'ajout sémantique que l'on vient d'introduire. Cela tient au fait que les décompositions sémantiques des pronoms interrogatifs de la forme «quel X» ont toujours deux nœuds communicativement dominants.

Quant à l'interrogation directe, elle est reflétée dans la SSém par une expression du type «je veux savoir ...» ou «je te prie de me dire si ... est vrai», qui donnera lieu à la réalisation prosodique et/ou syntaxique typique de l'interrogative.



## 1.2. La représentation syntaxique profonde

La **représentation syntaxique profonde** des phrases est constituée de quatre structures : la structure syntaxique profonde, qui est la structure centrale, et trois structures périphériques, la structure communicative syntaxique profonde, la structure anaphorique syntaxique profonde et la structure prosodique syntaxique profonde (Mel'čuk 1988 : 60, 1997 : 21, 24-25). Dans cet article, nous ne présenterons que la structure syntaxique profonde (§1.2.1) et la structure communicative syntaxique profonde (§1.2.3).

### 1.2.1. La structure syntaxique profonde

La **structure syntaxique profonde** [= SSyntP] d'une phrase est un **arbre de dépendance** caractérisé par un étiquetage particulier.

- Chaque nœud de la SSyntP est étiqueté par une **lexie pleine** de la phrase, éventuellement accompagnée de **grammèmes** ; les grammèmes sont notés en indice de la lexie, et par conséquent n'occupent pas de nœud en propre. Les lexies vides, c'est-à-dire les servitudes grammaticales, sont absentes de la SSyntP (à l'exception des translatifs verbalisateurs comme la copule ou les verbes supports) ; elles sont introduites lors du passage à la structure syntaxique de surface [= SSyntS].

- Chaque branche de la SSyntP est étiquetée d'une relation SyntP. (Remarque : lorsqu'un graphe est un arbre, il est d'usage d'appeler les arcs des *branches*.) Ce sont des relations universelles, qui auront, en fonction de la langue considérée, des réalisations diverses en syntaxe de surface. Dans cet article, nous distinguons six relations actanciennes (I, II, ..., VI), deux relations attributives [= modificatives] — restrictive et qualificative — (ATTR et ATTR<sub>QUAL</sub>), une relation de coordination (COORD) et une relation appenditive (APPEND), qui rattache au sommet SyntP tous les éléments "extrastructuraux" — tels l'apostrophe, les parenthétiques, les interjections, etc. Toutes ces relations, à l'exception de ATTR<sub>QUAL</sub>, ont déjà été utilisées en syntaxe profonde Sens-Texte. La relation SyntP ATTR<sub>QUAL</sub>, que nous proposons d'ajouter, encode toutes les modifications qualificatives, c'est-à-dire non définitoires, entre autres, la dépendance entre le nom modifié et la relative qualificative qui le modifie. (Une relative restrictive n'est pas explicitement marquée dans la SSyntP comme telle : il s'agit du cas par défaut, c'est-à-dire le cas prototypique des relations modificatives.)

Dans la SSyntP de (3), il faut indiquer la relation de coréférentialité entre les trois nœuds étiquetés LIVRE et les deux nœuds étiquetés PIERRE (cette relation est indiquée par un trait discontinu). Cette indication constitue la structure anaphorique profonde (nous reviendrons brièvement sur ce sujet au §3.1 en discutant de la Métrarègle 5 de surconsommation).

La SSyntP de la phrase de référence (3) (Fig. 11) exige, tout comme sa SSém, une clarification sur la représentation des pronoms interrogatifs.

### 1.2.2. Les pronoms interrogatifs dans la structure syntaxique profonde

Le pronom interrogatif (d'une subordonnée interrogative indirecte) a dans la syntaxe de la phrase deux rôles bien distincts à remplir : d'une part, il joue le rôle syntaxique de dépendant au sein de la proposition interrogative, rôle que

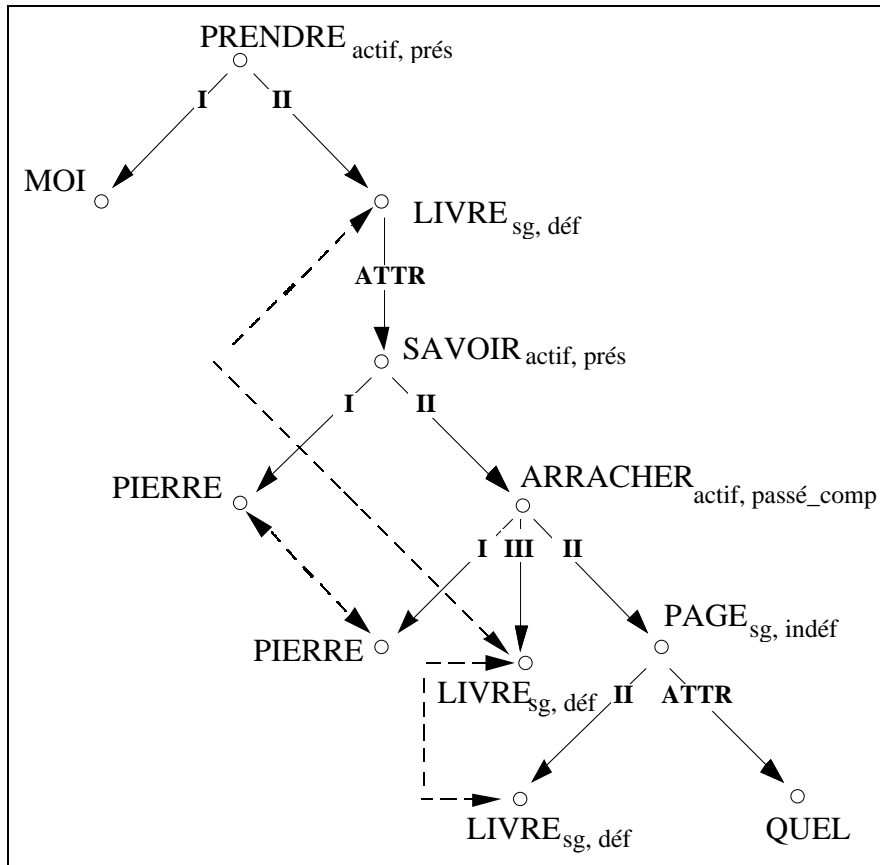


Figure 11. La SSyntP de la phrase de référence (3)  
*Je prends le livre dont Pierre sait quelle page il a arrachée*

peut jouer n'importe quelle autre lexie (pronominale ou non) ; d'autre part, il joue un rôle dit **subordinatif**, c'est-à-dire qu'il confère à la proposition qui le comprend la capacité d'être insérée dans des structures plus larges.

La dualité de rôle du pronom interrogatif a conduit à deux approches de sa représentation dans la structure syntaxique.

- La première approche attribue au pronom interrogatif dans la SSyntP uniquement son premier rôle — son rôle de dépendant (par exemple, Jespersen 1937 (1984) : §23.1, Mel'čuk & Pertsov 1987 : 395 et sv). Par conséquent, sous cette approche, les propositions du type *Pierre parle à Marie*, *A qui parle Pierre ?* et [*Je me demande*] à *qui Pierre parle* ont la même SSyntP (à l'étiquetage des nœuds près). L'autre rôle du pronom interrogatif — le rôle subordinatif — est pris en charge par d'autres moyens que la dépendance syntaxique (cf. notre proposition en §1.2.3).

- La seconde approche attribue au pronom interrogatif dans la SSyntP les deux rôles ensemble : son rôle de dépendant au sein de la proposition interrogative (comme dans la première approche) et son rôle subordinatif — celui de sommet syntaxique de cette proposition. Autrement dit, on obtient une double dépendance du pronom interrogatif (à moins d'effectuer un éclatement

du pronom interrogatif en deux nœuds) : il dépend de son gouverneur syntaxique dans la subordonnée et simultanément du gouverneur de la subordonnée (Tesnière 1959 : 560-61, Hudson 1990, Kahane 1996, Candito & Kahane 1998b). Ceci produit une structure syntaxique non purement arborescente. On sacrifie donc la simplicité logique de la structure pour rendre compte explicitement au niveau syntaxique du caractère particulier du pronom interrogatif. Remarquons que, dans la plupart des théories syntaxiques basées sur les constituants (GB, HPSG, LFG, ...), le pronom subordinatif est placé dans une position syntaxique particulière, par exemple COMP en GB (Chomsky 1981), position qui accueille également une conjonction de subordination comme QUE. De plus, cette position se trouve toujours en “liaison” directe avec le verbe régissant la proposition interrogative (comme sous-constituant immédiat d’un même constituant, qui est selon les cas, une projection de COMP ou du verbe), si bien que, même s’il peut être délicat de traduire les structures considérées par ces théories en structures de dépendance, on peut néanmoins considérer qu’elles sont beaucoup plus proches de la deuxième approche que de la première.

Ce qui vient d’être dit concerne en fait tous les types de **pronoms subordinatifs**, c’est-à-dire, en plus du pronom interrogatif, le pronom relatif et le pronom pseudo-relatif (cf. Note 1).

Dans cet article, nous adoptons la première approche. Nous pensons que les deux rôles joués par le pronom interrogatif sont de nature différente et doivent donc être représentés dans des structures différentes. C’est seulement son rôle purement syntaxique que nous représentons dans la SSyntP. Son autre rôle — le rôle subordinatif — est pris en charge dans la structure communicative SyntP.

### 1.2.3. La structure communicative syntaxique profonde

De façon analogue à la représentation sémantique, la représentation SyntP comprend sa propre structure communicative [= SComm-SyntP]. La SComm-SyntP est un objet formel de même nature que la SComm-Sém : il s’agit d’un marquage — par des marqueurs communicatifs — de portions connexes de la structure centrale, c’est-à-dire dans ce cas de la SSyntP. Cependant, alors que la SComm-Sém encode les caractéristiques communicatives des configurations de sémantèmes (= sous-réseaux), la SComm-SyntP encode les caractéristiques communicatives des configurations de lexies de la phrase (= sous-arbres). Étant donné la différence de nature de ces objets (sémantèmes vs lexies) et la différence de nature topologique de la structure centrale (réseau vs arbre), la SComm-Sém et la SComm-SyntP sont, de façon générale, différentes :

1. Le marquage des nœuds dominants dans la SComm-Sém, qui est utilisé dans le processus d’arborescence (en particulier la détermination des nœuds d’entrée), n’apparaît plus dans la SComm-SyntP, car l’information correspondante est complètement encodée par la structure d’arbre de la SSyntP.

2. Même si les divisions thématiques d’une RSyntP sont grosso modo héritées de la RSém correspondante, au niveau SyntP, un rhème et un thème associés sont nécessairement disjoints (ils ne peuvent posséder de nœuds communs).

Le passage d’un réseau à un arbre peut entraîner des changements importants de topologie — c’est-à-dire des discontinuités — qui provoquent des pertes

d'informations essentielles. Dans de tels cas, on ne peut récupérer dans l'arbre SyntP toutes les informations contenues dans le réseau Sém; c'est ici que la SSyntP doit être secondée par la SComm-SyntP. Voici un problème de ce type concernant les phrases à extraction.

Considérons les deux phrases suivantes :

(16) a. *Je sais que tu as dit à Pierre quel bateau tu avais pris.*

b. *Je sais quel bateau tu as dit à Pierre que tu avais pris.*

Leur sens propositionnel est différent et, par conséquent, leur SSém est aussi différente : pour (16a), l'arc 2 partant de 'savoir' pointe sur 'dire', pour (16b), sur 'quel'. Autrement dit, en (16a), le contenu de mon savoir est que tu as dit ce que tu as dit, et en (16b), c'est l'identité du bateau. Malgré ce fait, les deux phrases ont la même SSyntP. (Dans la seconde approche du traitement des pronoms subordinatifs, cette différence sémantique est directement encodée dans la structure syntaxique, par l'intermédiaire d'une dépendance supplémentaire entre QUEL et le verbe régissant l'interrogative.)

C'est par leur SComm-SyntP que ces deux phrases diffèrent : en (16b), QUEL est le focus rhématique  $\mathbf{FR}_1$  du rhème syntaxique primaire  $\mathbf{R}_1$  (*quel bateau tu as dit à Pierre que tu avais pris*) (Fig. 13), alors que ce n'est pas le cas en (16a) (Fig. 12). En effet, dans les deux cas, la division thématique primaire  $\mathbf{R}_1 \sim \mathbf{T}_1$  répond à la question sous-jacente *Que sais-je ?* : (16a) répond '... [que] tu as dit que ...' (SAVOIR porte sur DIRE), et (16b), '... quel bateau ...' (SAVOIR porte sur QUEL). Par contre, en (16b) comme en (16a), QUEL est le focus rhématique  $\mathbf{FR}_2$  du rhème syntaxique secondaire  $\mathbf{R}_2$  (*quel bateau ... tu avais pris*). Il est aussi dans les deux cas le  $\mathbf{FR}_3$  de  $\mathbf{R}_3$ .

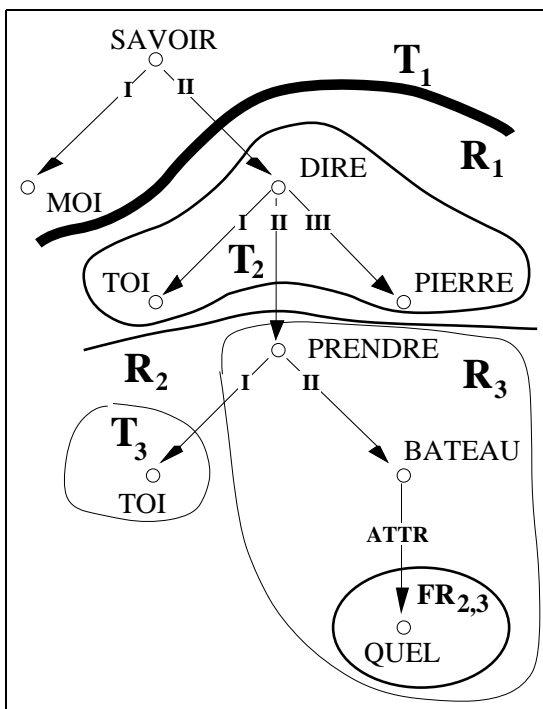


Figure 12. La RSyntP de (16a)

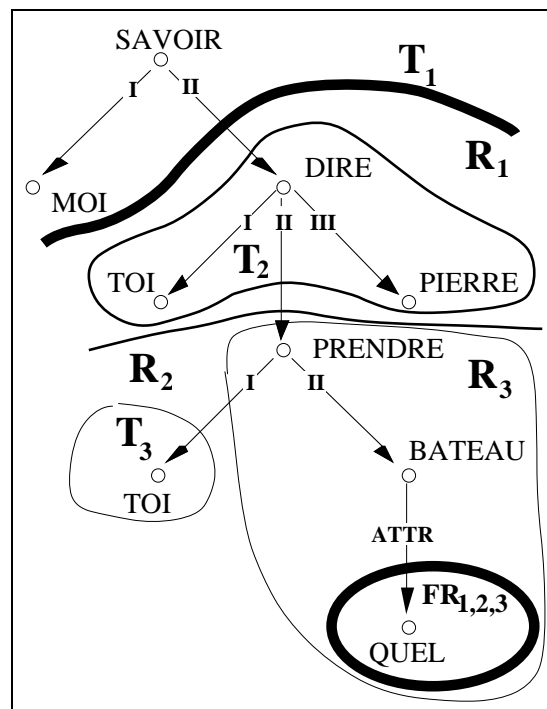


Figure 13. La RSyntP de (16b)

Dans l'étape suivante de la synthèse (le passage  $\text{RSyntP} \Rightarrow \text{RSyntS}$ ), le marquage du **FR** syntaxique sera utilisé par les règles linguistiques de correspondance pour assurer la production de deux  $\text{SSyntS}$  différentes : la  $\text{SSyntS}$  de (16a) aura QUE sous SAVOIR, mais pas sous DIRE, et inversement pour (16b), c'est-à-dire que la conjonction QUE sera ou non insérée, en fonction de la présence et de la nature du **FR** syntaxique.

## 2. LES RÈGLES SÉMANTIQUES DE CORRESPONDANCE

Dans cette section, nous nous intéresserons seulement aux **règles sémantiques de correspondance** (en laissant de côté les règles sémantiques d'équivalence, évoquées au §1.1.1). Les règles de correspondance sont des règles linguistiques (déclaratives) qui expriment les faits pertinents de la langue décrite ; elles doivent être mises en jeu par un ensemble de règles procédurales (impératives).

Nous commencerons par donner une ébauche générale du **sous-module sémantique de correspondance** d'un modèle Sens-Texte, c'est-à-dire des règles linguistiques de correspondance et des règles procédurales qui les mettent en jeu (§2.1). Dans un deuxième temps, nous examinerons les règles linguistiques de correspondance, en dédiant à chaque groupe de règles une sous-section à part (§§ 2.2 et 2.3).

### 2.1. Ebauche générale du sous-module sémantique de correspondance

Le sous-module sémantique de correspondance d'un modèle Sens-Texte doit traiter de toutes les structures composant la représentation sémantique ( $\text{SSém}$  et  $\text{SComm-Sém}$ ) et de toutes les structures composant la représentation syntaxique profonde (structure  $\text{SyntP}$ , structure communicative  $\text{SyntP}$ , structure anaphorique  $\text{SyntP}$  et structure prosodique  $\text{SyntP}$ ). L'interaction entre ces différentes structures rend le passage entre les deux représentations en question assez complexe. Dans le cadre de cet article, nous nous limiterons à la construction de la  $\text{SSyntP}$  à partir d'une  $\text{SSém}$  donnée, munie d'une  $\text{SComm-Sém}$ . Insistons sur le fait que le passage de la  $\text{SSém}$  à la  $\text{SSyntP}$  dépend de façon essentielle de la  $\text{SComm-Sém}$ .

Avant de proposer des règles sémantiques de correspondance, nous allons expliquer grossièrement comment celles-ci sont mises en jeu (par des règles procédurales).

#### 2.1.1. La mise en jeu des règles sémantiques de correspondance

On peut comparer le passage " $\text{SSém}+\text{SComm-Sém} \Rightarrow \text{SSyntP}$ " à la traduction mot à mot d'un texte d'une langue  $\mathcal{L}_1$  dans une langue  $\mathcal{L}_2$  à l'aide d'un dictionnaire bilingue  $\mathcal{L}_1 \Rightarrow \mathcal{L}_2$ . La procédure consiste alors à lire le texte en langue  $\mathcal{L}_1$  (sans le modifier), à consulter, à chaque mot du texte, le dictionnaire pour trouver la ou les traductions possibles de ce mot en langue  $\mathcal{L}_2$  et à écrire la traduction trouvée là où il faut. Dans le cas du passage " $\text{SSém}+\text{SComm-Sém} \Rightarrow \text{SSyntP}$ ", la procédure est rendue beaucoup plus complexe du fait qu'une  $\text{SSém}$ , à la différence d'un texte, n'est pas un objet linéaire avec un sens de parcours évident. Mais malgré sa plus grande complexité, la nature logique de

la procédure  $RSém \Rightarrow SSyntP$  reste la même que celle d'une traduction mot à mot.

Un mécanisme destiné à effectuer de tels passages s'appelle un **transducteur**. A notre connaissance, la théorie des transducteurs s'est essentiellement intéressée à la transduction chaîne-chaîne ou arbre-arbre (Aho & Ullman 1972). Quant à nous, il nous faut explorer la transduction graphe-arbre.

Le passage " $SSém+SComm-Sém \Rightarrow SSyntP$ " suppose trois tâches majeures :

- lire entièrement la  $SSém$  — un élément (un nœud ou un arc) après l'autre ;
- déclencher, pour chaque élément lu, une règle sémantique correspondante ;
- mettre l'élément fourni par la règle déclenchée — c'est-à-dire un nœud et sa lexie ou une branche et sa relation  $SyntP$  — à la bonne place dans la  $SSyntP$  qui est en train d'être construite.

En raison des discontinuités inhérentes entre la  $SSém$  et la  $SSyntP$  des phrases à extraction, il n'est pas possible d'assurer, dans le sens de la synthèse, à la fois le parcours continu de la  $SSém$  et le traçage continu de la  $SSyntP$  (et vice versa dans le sens de l'analyse).

|| La procédure que nous proposons privilégie le traçage CONTINU de l'arbre  
||  $SyntP$  du sommet vers les feuilles en permettant en contrepartie la lecture  
|| DISCONTINUE de la  $SSém$ .

Une procédure inverse — qui privilégie la lecture continue de la  $SSém$  en permettant le traçage discontinu de l'arbre  $SyntP$  — est aussi possible (cf. Candito & Kahane 1998). Soulignons que les règles sémantiques de correspondance considérées dans le présent article sont indépendantes de la procédure choisie.

Techniquement, la procédure de passage utilise un outil formel, que nous appellerons le **pointeur**<sup>8</sup>.

**Définition 1.** Un *pointeur* est un objet formel constitué de trois composants :

- une **tête de lecture**, qui doit être braquée sur un élément de la  $SSém$  (nœud ou arc  $Sém$ ) et le lire ;
- une **tête d'écriture**, qui doit tracer le (ou les) élément(s) correspondant(s) de la  $SSyntP$  (nœud ou branche  $SyntP$ ) ;
- un **bloc-note** (une sorte de mémoire), sur lequel peuvent être inscrites les conditions d'application des règles sémantiques (communiquées d'une règle à l'autre).

Comme le lecteur l'a déjà compris, nous modélisons la procédure de passage par une sorte de machine de Turing à deux rubans : l'un sur lequel est inscrite la  $SSém$  à lire et l'autre sur lequel sera tracée la  $SSyntP$  correspondante. La différence avec la machine de Turing classique est que, dans notre perspective, il s'agit de la lecture d'un réseau, et non d'un ruban linéaire, ce qui rend nécessaire

---

8. Notre pointeur est plus ou moins un pointeur au sens usuel. Il possède néanmoins comme caractéristique particulière d'être associé à une mémoire individuelle, que nous appelons son bloc-note, à la différence d'autres systèmes, où une seule mémoire est utilisée.

l'utilisation de plusieurs pointeurs ; ceux-ci sont créés au fur et à mesure de la lecture du réseau Sém par des règles procédurales.

Le maniement des pointeurs, qui est essentiellement d'ordre procédural, n'est pas mentionné dans les règles linguistiques de correspondance. Il est assuré par un ensemble de règles procédurales, que nous appelons les **métarègles de pointage**, présentées en détail dans la Section 3. A chaque type de règle de correspondance est associée une métarègle de pointage.

Pour fixer les idées, présentons dès maintenant — de façon informelle — la procédure de synthèse des phrases simples, où trois de nos métarègles de pointage sont impliquées. Rappelons qu'en l'absence de règles discontinues, la tête de lecture d'un pointeur passe d'un nœud à un arc adjacent, de cet arc à son autre extrémité, etc. Le fait qu'un nœud puisse posséder plusieurs arcs adjacents nécessite de multiplier les pointeurs à chaque "branchement" du réseau. La SSyntP, quant à elle, est tracée de façon continue du sommet vers les feuilles. La procédure déplace les têtes d'écriture des pointeurs parallèlement au déplacement des têtes de lecture correspondantes. Si les têtes de lecture assurent la lecture exhaustive de la SSém, les têtes d'écriture correspondantes assurent le bon positionnement — dans la SSyntP — des éléments SyntP fournis par les règles de correspondance.

La synthèse débute ainsi :

- Dans la SSém considérée, un nœud précis, appelé le **nœud de départ**, est choisi. Un premier pointeur est créé et sa tête de lecture est braquée sur le nœud de départ. Dans notre procédure, le nœud de départ est le nœud d'entrée (cf. §1.1.2)<sup>9</sup>. La tête d'écriture du premier pointeur est donc prête à tracer le nœud SyntP qui sera le sommet de l'arbre SyntP. Ces actions sont contrôlées par la Métarègle 1 d'initialisation. (Les différentes métarègles nécessaires à la synthèse sont présentées au §3.1.)

Ensuite, la synthèse se déroule suivant deux cas de figures, selon que l'on manie un pointeur braqué sur un nœud ou un arc Sém.

- Quand un pointeur arrive sur un nœud (étiqueté par un sémantème lexical  $\langle s \rangle$ ), il consulte le lexique du modèle pour y trouver (dans un article de dictionnaire) la règle faisant correspondre au sémantème  $\langle s \rangle$  sa lexie  $L(\langle s \rangle)$  et déclenche cette règle ; conformément à la métarègle de pointage associée à la règle en question, un nœud SyntP d'étiquette  $L(\langle s \rangle)$  est créé là où se trouve la tête d'écriture du pointeur. Ensuite, si le nœud considéré possède  $n$  arcs adjacents (sans compter l'arc par lequel le pointeur est arrivé sur ce nœud), le

---

9. Le choix du nœud d'entrée en qualité de nœud de départ présuppose que le nœud d'entrée a été calculé au préalable, c'est-à-dire qu'une partie de la SSém (en particulier les nœuds dominants des rhème et thème primaires) a déjà été considérée. Si on veut éviter cette phase, c'est-à-dire l'intégrer à la lecture proprement dite, on peut imaginer une procédure qui ait, par exemple, pour nœud de départ le nœud dominant du rhème et qui, tout en commençant la production de la SSyntP — en déclenchant une règle sémantique pour ce nœud —, décide si ce nœud est ou non le nœud d'entrée de la SSém, c'est-à-dire si le nœud SyntP correspondant est ou non le sommet de l'arbre SyntP. Une telle procédure pourrait être plus économique que la nôtre, puisque la détermination du nœud d'entrée de la SSém dépend des lexicalisations possibles des nœuds dominants des rhème et thème primaires — notamment du fait que ces nœuds sont ou non lexicalisés par un verbe.

pointeur se démultiplie en  $n$  pointeurs et la tête de lecture de chacun de ces  $n$  pointeurs est déplacée sur l'un des  $n$  arcs adjacents — afin de continuer la lecture de la SSém en parallèle ; en même temps, les  $n$  têtes d'écriture correspondantes sont prêtes à tracer une branche dépendante du nœud SyntP qui vient d'être créé. Dans le cas particulier où  $n = 0$ , c'est-à-dire quand le nœud considéré n'a pas d'autre arc adjacent que l'arc d'arrivée, le pointeur disparaît (= se démultiplie en 0). Tout ceci est assuré par la Métrarègle 4 de pointage pour une règle nodale standard.

- Quand un pointeur arrive sur un arc (étiqueté par une relation sémantique  $r$ ,  $r = 1, 2, \dots, 6$ ), ce pointeur consulte la composante sémantique de la grammaire du modèle pour y trouver une règle faisant correspondre à la relation sémantique  $r$  une relation syntaxique profonde  $R$  ; conformément à la métrarègle de pointage associée, la branche SyntP d'étiquette  $R$  est tracée là où se trouve la tête d'écriture du pointeur ; puis cette tête est déplacée à l'extrémité dépendante de la branche qui vient d'être tracée, tandis que la tête de lecture est braquée sur l'autre extrémité de l'arc Sém considéré. Ceci se déroule sous le contrôle de la Métrarègle 6 de pointage pour une règle sagittale standard.

Dans ce cas, la règle sémantique choisie dépend du **sens de parcours** de l'arc Sém considéré. Si le sens de parcours est **positif** (= sens prédicat → argument), l'arc Sém donne lieu à une branche SyntP étiquetée d'une relation syntaxique profonde  $R$  **actancielle** ( $R = I, II, \dots, VI$ ) et orientée dans le même sens ; la règle utilisée, qui ne change donc pas l'orientation des dépendances, est dite **positive**. Si le sens de parcours est **négatif**, l'arc Sém donne lieu à une branche SyntP étiquetée d'une relation syntaxique profonde  $R$  **attributive** ( $R = ATTR$  ou  $ATTR_{QUAL}$ ), orientée dans le sens inverse ; la règle utilisée, qui change l'orientation des dépendances, est dite **négative**. (Comme nous l'avons déjà dit, nous laissons de côté la synthèse des relations COORD et APPEND.) Une règle de correspondance où figurent des arcs Sém doit donc comprendre l'indication du sens de parcours qui autorise son application.

S'il n'y a pas d'échec dans la procédure de synthèse, la lecture de la SSém s'arrête quand tous les pointeurs (tous issus du pointeur initial) ont disparus ; les métrarègles précédentes, complétées d'autres métrarègles appropriées, permettent d'assurer que le réseau entier a été lu.

Insistons sur un dernier point : la principale difficulté dans l'élaboration de la procédure de synthèse est d'assurer une consommation correcte de la SSém. Un élément de la SSém (nœud ou arc) est dit **consommé** lorsqu'il reçoit une traduction SyntP. Pour que la consommation de la SSém soit correcte, la condition suivante doit être remplie :

|| Chaque arc de la SSém doit être consommé une et une seule fois.

Les nœuds Sém, quant à eux, peuvent être consommés autant de fois que nécessaire. Expliquons-nous. Alors que la SSém peut posséder des cycles, la SSyntP, qui est un arbre de dépendance, est acyclique. Ainsi, lors de l'établissement de la correspondance entre un réseau Sém et un arbre SyntP, chaque cycle de la SSém doit être coupé en un point ; ce point, unique, est



appelé le **point de coupure** du cycle. Le point de coupure d'un cycle sera consommé deux fois (ou  $n + 1$  fois s'il est le nœud de coupure de  $n$  cycles). A chaque fois qu'un nœud Sém est atteint, on obtient un nouveau nœud SyntP ; tous les nœuds SyntP issus d'un même nœud Sém doivent être étiquetés par la même lexie et sont reliés entre eux par une relation de **coréférence**. (Les relations de coréférence constituent, comme nous l'avons dit, la structure SyntP anaphorique.) Ces tâches sont également gérées par des métarègles de pointage : la Métarègle de surconsommation assure la consommation d'un nœud qui a déjà été consommé une ou plusieurs fois, tandis que la Métarègle de surpointage assure qu'un arc ne soit consommé qu'une seule fois (cf. §3.1). En l'absence de discontinuités et de cycles, un seul parcours de la SSém est possible, une fois déterminé le nœud d'entrée.

### 2.1.2. Les règles sémantiques de correspondance

Les règles linguistiques de correspondance en général et les règles sémantiques de correspondance en particulier représentent la composante purement linguistique — et donc spécifique à chaque langue — d'un modèle Sens-Texte ; elles sont censées ne refléter que les faits de la langue, c'est-à-dire les connaissances linguistiques du locuteur. Ces règles sont présentées de façon déclarative : elles sont statiques et non ordonnées entre elles.

Les règles sémantiques de correspondance peuvent être classées selon deux points de vue :

- le point de vue sémantique (quel est le type d'élément traité dans la partie gauche, nœud ou arc Sém ?) ;
- le point de vue syntaxique (quel est le type d'élément traité dans la partie droite, nœud ou branche SyntP ?).

Le premier point de vue donne la division en **règles nodales** (traitant des nœuds Sém) et **règles sagittales** (traitant des arcs Sém)<sup>10</sup> ; la deuxième en **règles de lexicalisation** (traitant des nœuds SyntP, c'est-à-dire des lexies et/ou des grammèmes) et **règles d'arborisation** (traitant des branches SyntP). Ces deux classifications se croisent en donnant quatre classes possibles de règles sémantiques. Les règles nodales standard sont des règles de lexicalisation (à un nœud Sém correspond un nœud SyntP) et les règles sagittales standard sont des règles d'arborisation (à un arc Sém correspond une branche SyntP). Cependant, il existe aussi des règles nodales qui sont des règles d'arborisation (à un nœud Sém correspond, entre autres, une branche SyntP ; par exemple la lexicalisation de 'lutter' par MENER  $\rightarrow$  LUTTE) et des règles sagittales qui sont des règles de lexicalisation (par ex., à un arc Sém correspond un grammème attaché à un nœud SyntP ; voir la règle de réflexivisation  $R_{+2/3}^{\text{réf}}$ , Fig. 21).

Dans la suite, nous présenterons essentiellement des règles sémantiques standard, c'est-à-dire des règles nodales de lexicalisation et des règles sagittales

---

10. La classification en règles nodales et sagittales, introduite explicitement dans cet article, peut être étendue à tous les modules d'un modèle Sens-Texte, puisqu'à tous les niveaux la structure centrale est un graphe étiqueté.

d'arborisation<sup>11</sup>.

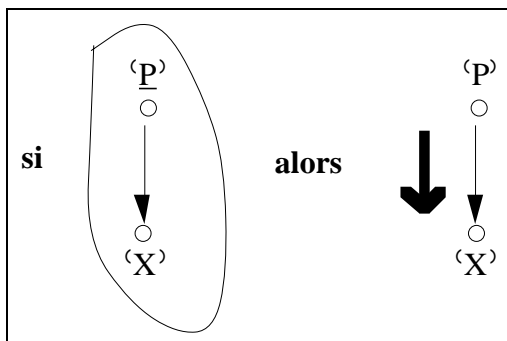
### Conventions d'écriture des règles sémantiques

1) Dans une règle, nous marquons en gras les éléments manipulés par la règle ; pour la synthèse, ce seront les éléments Sém consommés et les éléments SyntP créés et vice versa pour l'analyse. Les éléments présentés en maigre constituent le contexte.

2) La grosse flèche à côté d'un arc Sém (dans la partie gauche d'une règle) marque le sens de parcours de cet arc. Le sens de parcours des branches SyntP (dans la partie droite) est supposé être toujours le sens gouverneur → dépendant et n'est donc pas indiqué. Ce que nous appelons, dans cet article, le sens de parcours d'un arc Sém est en fait un sens de parcours bien particulier : c'est le sens de parcours de l'arc Sém lorsque la branche SyntP correspondante est tracée dans le sens positif. (La procédure proposée dans Candito & Kahane 1998 parcourt, dans certains cas, un arc Sém à contre-sens du sens de parcours adopté ici : dans un tel cas, la branche SyntP correspondante est tracée dans le sens négatif, c'est-à-dire du dépendant vers le gouverneur.)

La flèche de sens de parcours dans une règle sémantique est en fait un renvoi à l'une des deux conditions qui expriment les contraintes communicatives sur la correspondance  $SSém \Leftrightarrow RSyntP$  (c'est ici que se manifeste le rôle crucial de la SComm-Sém dans la synthèse sémantique)<sup>12</sup>.

### Condition pour l'application d'une règle positive



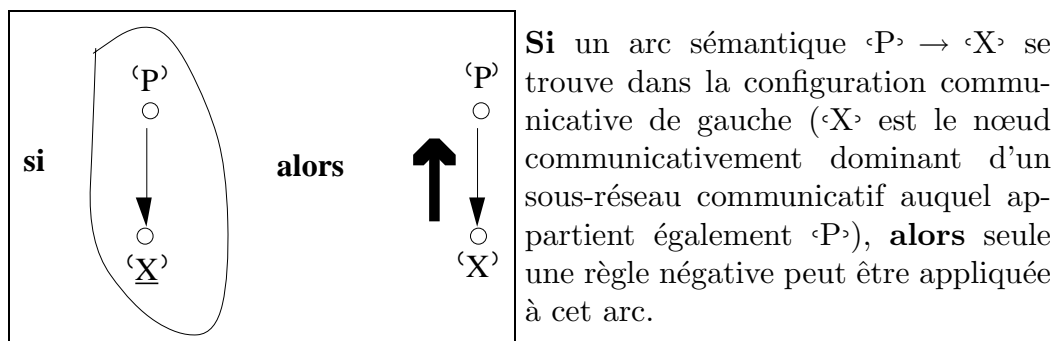
Si un arc sémantique  $\langle P \rangle \rightarrow \langle X \rangle$  se trouve dans la configuration communicative de gauche ( $\langle P \rangle$  est le nœud communicativement dominant d'un sous-réseau communicatif auquel appartient également  $\langle X \rangle$ ), **alors** seule une règle positive peut être appliquée à cet arc.

**Exemple :** *Jean lit* [L( $\langle P \rangle$ )] *un livre* [L( $\langle X \rangle$ )].

11. Il faut noter, concernant l'architecture générale du système de règles sémantiques, que la première tentative de séparer règles de lexicalisation et règles d'arborisation est due à Polguère 1990. Les précédentes études dans le cadre du modèle Sens-Texte proposaient un seul type de règle sémantique assurant conjointement la lexicalisation et l'arborisation (cf. une formalisation avec implémentation dans Boyer & Lapalme 1985) ; considérées du point de vue de la synthèse, de telles règles produisaient conjointement une lexie et les branches conduisant à ses actants SyntP. Notre étude fait un pas en avant dans le sens de la décomposition logique des règles ; plusieurs règles proposées par Polguère 1990 : 142-46, notamment la (très partielle) règle de relativisation (Règle II, p. 142), apparaissent dans notre cadre comme la composée d'au moins deux règles.

12. Lorsqu'on ne se trouve dans aucune des deux situations considérées par ces conditions, l'application d'une règle de correspondance n'est pas sujette à des contraintes de nature communicative. Cela ne signifie cependant pas que n'importe quel type de règle puisse être alors utilisé : le choix d'une règle sémantique applicable est contraint par les règles déjà choisies et appliquées sur les portions adjacentes du réseau. Cela découle de la nécessité d'obtenir à la sortie du processus de passage une structure strictement arborescente, c'est-à-dire de pouvoir coller ensemble de façon satisfaisante les différents éléments SyntP fournis par les différentes règles sémantiques utilisées.

## Condition pour l'application d'une règle négative



**Exemple :** *Le livre* [L( $\langle X \rangle$ )] *lu* [L( $\langle P \rangle$ )] *par Jean est sur la table.*

Il nous faut faire ici une remarque importante concernant la relation entre la structure communicative et le sens de parcours et plus particulièrement le concept même de dominance communicative. Pour assurer le bon parcours du réseau Sém, il a été proposé par A. Polguère (1990) d'indiquer par un soulignement dans chaque règle sémantique un nœud "communicativement dominant" dans la partie gauche. Nous n'avons pas retenu ce procédé et nos raisons doivent maintenant être claires. Si le nœud souligné dans une règle Sém doit être interprété comme le nœud communicativement dominant au sens strict — c'est-à-dire comme le nœud qui "résume" le contenu d'un sous-réseau communicatif —, alors certaines configurations de sémantèmes dans la SSém ne seront pas couvertes par les règles sémantiques : par exemple, si le réseau n'a pas de nœud dominant (comme c'est le cas des réseaux propositionnels du type *Jean pleure*). Si nous voulons que les règles couvrent tous les cas de figure possibles, il faut donner au soulignement proposé dans Polguère 1990 une autre interprétation, à savoir : le nœud qui est souligné est le nœud qui domine ses partenaires du point de vue d'une hiérarchie INDUITE par la dominance communicative au sens strict. Le terme *dominance communicative* dans Polguère 1990 recouvre en fait deux notions apparentées, mais bien distinctes : la dominance communicative au sens strict et la *précédence communicative* induite par la dominance communicative. La dominance communicative exprime une décision du locuteur et est indiquée dans la représentation Sém de départ, avant l'application des règles sémantiques de correspondance. Par contre, la **précédence communicative** induite, qui n'est pas en général unique pour une représentation Sém donnée, ne préexiste pas à l'application des règles Sém et est calculée au fur et à mesure du passage vers la SyntP. C'est la relation de précédence communicative induite que nos flèches de sens de parcours dans les règles sémantiques sont appelées à capturer.

## 2.2. Les règles sémantiques (nodales) de lexicalisation

Les règles sémantiques de lexicalisation incluent les règles lexicales et les règles grammaticales (= flexionnelles et dérivationnelles). Nous ne présenterons que les premières.

Une **règle sémantique lexicale** fait correspondre à un nœud Sém étiqueté d'un sémantème un nœud SyntP étiqueté d'une lexie du français dont le signifié

est ce sémantème. La lexie introduite par une règle lexicale est donc une lexie pleine. Il y a autant de règles lexicales que de lexies (quelques centaines de milliers). Néanmoins, le format d'une règle sémantique lexicale est toujours le même<sup>13</sup> (Fig. 14).

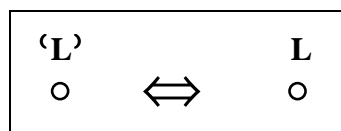


Figure 14. La règle sémantique lexicale pour la lexie L

Les règles lexicales ne sont pas aussi triviales qu'elles en ont l'air dans la Fig. 14. En effet, l'étiquette L, dans la partie droite de la règle, n'est pas autre chose qu'un renvoi à un article de dictionnaire (d'un DEC) contenant toutes les informations pertinentes pour le bon usage de la lexie L. Ainsi, par exemple, pour L = ARRACHER, l'article de dictionnaire correspondant fournit les informations suivantes :

- 1) la partie du discours (ARRACHER est un verbe, et, par conséquent, exige la présence de grammèmes de mode, de temps, de personne et de nombre) ;
- 2) la diathèse (c'est-à-dire la correspondance entre ses trois actants Sém et ses actants SyntP, notamment le caractère obligatoire ou non et les moyens de leur expression) ;
- 3) le groupe conjugationnel (groupe I) ;
- 4) la cooccurrence lexicale restreinte (*arracher d'un coup sec, sans ménagement ...*) ;
- 5) les capacités de dérivation (*arrachage, arrachable ...*) ; etc.

Pour le lexème QUEL, l'article de dictionnaire précise que c'est un pronom interrogatif adjectival. Par conséquent, en tant qu'adjectif, il doit dépendre, dans l'arbre SyntP, d'un nom par une relation de type ATTR. (Les informations dictionnaires sont également utilisées aux niveaux plus superficiels : le fait qu'il s'agit d'un adjectif impose au niveau SyntS l'introduction de grammèmes de genre et de nombre, induits par l'accord avec le nom modifié ; le fait qu'il s'agit d'un pronom subordonatif impose un placement adéquat lors de la linéarisation de la structure SyntS.)

Il est courant, dans la tradition Sens-Texte, d'indiquer dans les règles lexicales la diathèse (c'est-à-dire les actants Sém dans la partie gauche et les actants SyntP dans la partie droite) (voir, par exemple, Fig. 15). Cette façon de présenter les règles est équivalente à la précédente. Cependant, si, pour fins pédagogiques, on décide de représenter les actants, ils doivent être écrits en maigre, puisqu'ils ne sont pas consommés par la règle.

---

13. Une exception importante : les règles lexico-fonctionnelles. Une telle règle produit, au lieu d'une vraie lexie possédant un article de dictionnaire, le symbole d'une fonction lexicale, qui sera réalisé par une vraie lexie à l'étape suivante de la synthèse (Polguère 1998).

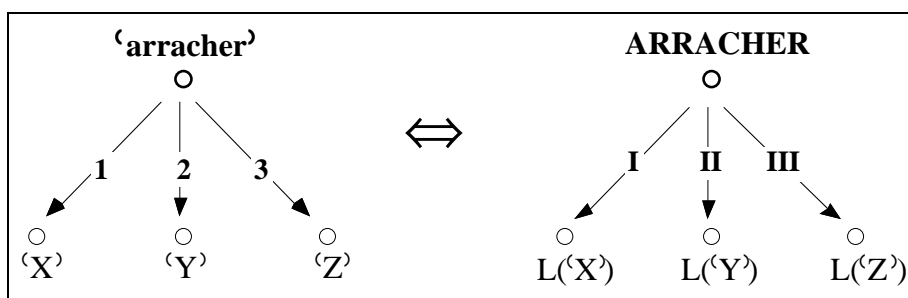


Figure 15. La règle sémantique lexicale pour ARRACHER (variante)

### 2.3. Les règles sémantiques (sagittales) d'arborisation

Il existe trois types majeurs de règles d'arborisation : les règles continues (toujours locales), les règles discontinues locales et les règles dicontinues non locales. Nous en présenterons deux types dans ce paragraphe : d'abord, des règles qui servent à la synthèse des phrases simples et qui sont des règles continues (§2.3.1), ensuite nos règles d'extraction, qui sont des règles discontinues non locales (§2.3.2).

#### 2.3.1. Les règles d'arborisation continues

Dans le cas prototypique, une règle d'arborisation continue fait correspondre à un arc de la SSém reliant deux sémantèmes une branche de l'arbre SyntP reliant les deux lexies correspondantes. Dans cette correspondance, l'orientation du lien peut être soit préservée, soit inversée. La relation sémantique  $r$  qui étiquette l'arc est toujours une relation actancielle (= prédicat-argument). La nature de la relation SyntP correspondante  $R [= R(r)]$  dépend du sens du parcours de l'arc sémantique d'étiquette  $r$  : comme on l'a dit, le parcours d'une relation Sém dans le sens de la dépendance donne une relation SyntP actancielle, alors que le parcours dans le sens inverse donne une relation SyntP attributive.

Il faut noter à ce propos le trait typique suivant des langues naturelles :

|| Seule une relation sémantique 1 est susceptible d'être parcourue négativement par une règle d'arborisation continue, c'est-à-dire de donner une relation SyntP attributive.

Strictement parlant, cette affirmation n'est pas tout à fait correcte ; nous y reviendrons dans le commentaire (5) après les règles.

Les règles d'arborisation sont en petit nombre (de l'ordre de la dizaine) ; elles sont universelles en ce sens qu'elles mentionnent aussi peu que possible les traits et les propriétés d'une langue particulière et constituent un répertoire dont chaque langue sélectionne un sous-ensemble approprié.

Nous proposons ici quelques règles d'arborisation continues, dont celles nécessaires pour la synthèse de la phrase (3).

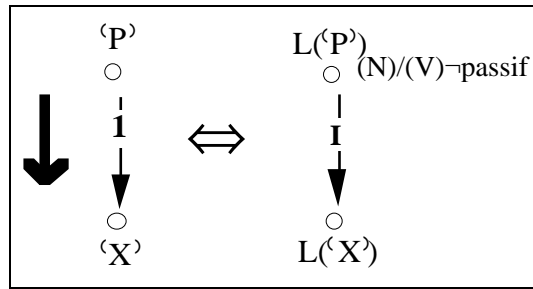


Figure 16. La règle  $R_{+1}$   
(parcours positif de la relation sémantique 1)

Exemple : *Je* ←<sub>I</sub> *prends le livre* ; *Il* ←<sub>I</sub> *a arraché une page* ; *Pierre* ←<sub>I</sub> *se rase* ; *les connaissances* →<sub>I</sub> *de Pierre* ; *l'action* →<sub>I</sub> *de Pierre*.

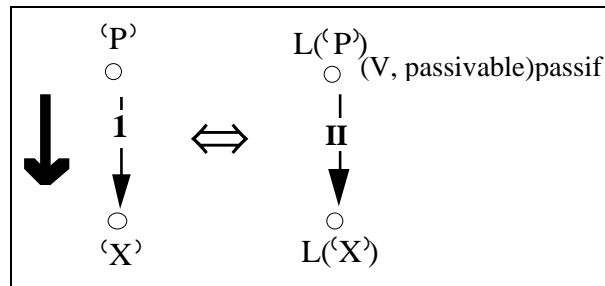


Figure 17. La règle  $R_{+1}^{\text{pass}}$   
(parcours positif de la relation sémantique 1 avec passivisation)

Exemple : *Le livre* ←<sub>I</sub> *est pris* ; *La page* ←<sub>I</sub> *a été arrachée*.

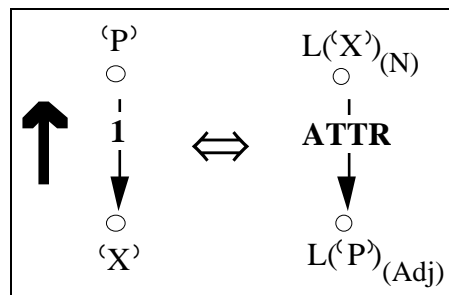


Figure 18. La règle  $R_{-1(\text{Adj})}$   
(parcours négatif de la relation sémantique 1  
avec production d'un modifieur adjectival)

Exemple : *un beau* ←<sub>ATTR</sub> *livre* ; *quel* ←<sub>ATTR</sub> *livre*.

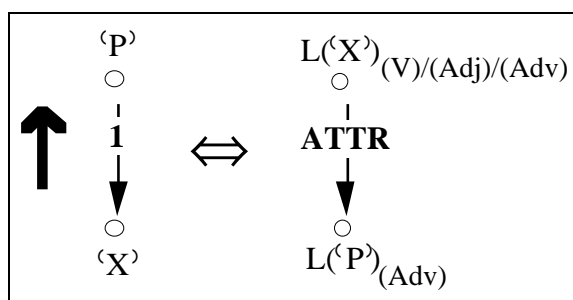


Figure 19. La règle  $R_{-1(Adv)}$   
(parcours négatif de la relation sémantique 1  
avec production d'un modifieur adverbial)

**Exemple :** *Pierre arrache* –ATTR→ *souvent des pages* ; *un très* ←ATTR– *beau livre*.

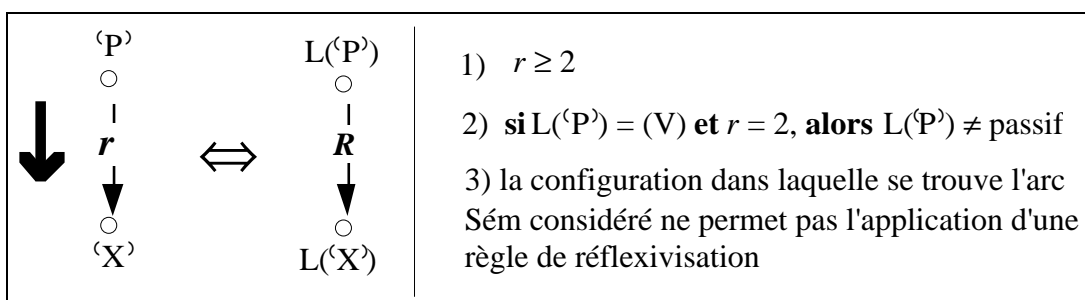


Figure 20. La règle  $R_{+≥2}$   
(parcours positif de la relation sémantique  $r, r ≥ 2$ )

**Exemple :** *Je prends* –II → *le livre* ; *Il a arraché* –II → *une page* ; *Pierre offre* –III → *à Marie un cadeau* ; *le cadeau offert* –III → *à Marie* ; *l'arrachage* –II → *d'une page*.

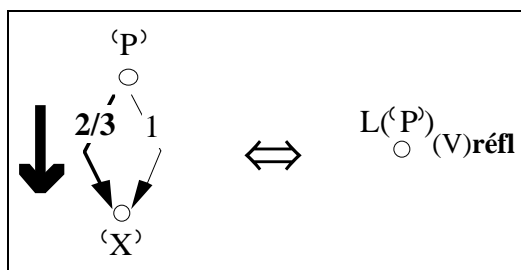


Figure 21. La règle  $R_{+2/3}^{réfl}$   
(parcours positif de la relation sémantique 2 ou 3 avec réflexivisation)

**Exemple :** *Pierre se parle* (2 ⇒ réfl) ; *Pierre se livre à la police* (2 ⇒ réfl) ; *Pierre s'envoie une lettre* (3 ⇒ réfl).

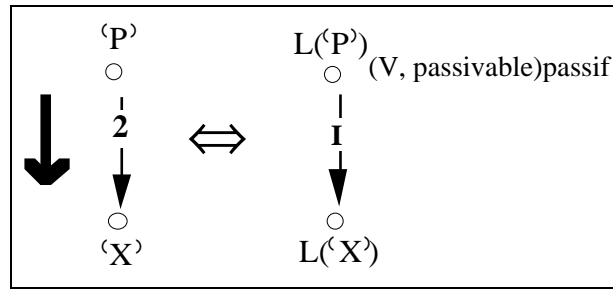


Figure 22. La règle  $R_{+2}^{\text{pass}}$

(parcours positif de la relation sémantique 2 avec passivisation)

Exemple : *Le livre a été pris* –II→ *par Pierre* ; *Pierre est aimé* –II→ *de tout le monde*.

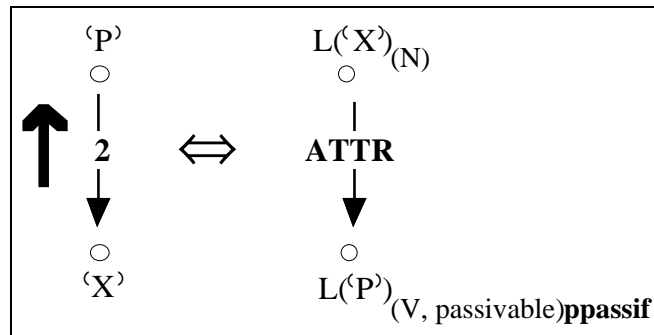


Figure 23. La règle  $R_{-2}^{\text{pass}}$

(parcours négatif de la relation sémantique 2 avec passivisation)

Exemple : *la page* –ATTR→ *arrachée par Pierre*.

### Commentaires sur les règles d'arborisation

1) Le parcours négatif de la relation sémantique 1 (règles  $R_{-1(\text{Adj})}$  et  $R_{-1(\text{Adv})}$ ) nécessite d'exprimer le gouverneur de cette relation, le prédicat sémantique  $\langle P \rangle$ , par un adjectif (Adj) ou un adverbe (Adv), qui devient le dépendant dans la relation SyntP ATTR, c'est-à-dire un modifieur. Ces règles sont des cas particuliers de règles plus générales : le dépendant  $L(\langle P \rangle)$  peut en fait être de n'importe quelle partie du discours, mais s'il n'est pas de la partie du discours requise (Adj ou Adv), il doit être translaté en celle-ci — par un translatif morphologique ou analytique (Tesnière 1959). Par exemple, un verbe peut occuper une position SyntP où un adjectif est attendu — à condition d'être translaté en participe actif (= présent) ou passif (= passé), cf. règle  $R_{-2}^{\text{pass}}$ .

Seul un nom peut être modifié par un adjectif ; les lexies des trois autres parties du discours sont modifiées par des adverbes. Par contre, le parcours positif d'une relation sémantique 1 (règle  $R_{+1}$ ) exclut d'exprimer le gouverneur de cette relation, toujours le prédicat sémantique  $\langle P \rangle$ , par un Adj ou un Adv et nécessite donc d'exprimer  $\langle P \rangle$  par un verbe (V) ou un nom (N), qui est alors le gouverneur de la relation actancielle SyntP I. En conséquence, seul un Adj ou un Adv peut être un **modifieur direct**, c'est-à-dire un modifieur résultant



d'une règle continue. De plus, un Adj ou un Adv n'a jamais d'actant SyntP I, puisqu'il n'est pas possible de parcourir positivement la relation sémantique 1 d'un sémantème exprimé par un Adj ou un Adv. En fait, cette propriété définit les Adj et Adv en tant que partie du discours.

2) La règle  $\mathbf{R}_{+\geq 2}$  est la règle standard pour une relation  $r \geq 2$ . La valeur de  $R$  correspond directement à celle de  $r$  : pour  $r = 2$ ,  $R = \text{II}$  ; pour  $r = 3$ ,  $R = \text{III}$  ; etc.

3) Dans la règle  $\mathbf{R}_{+2/3}^{\text{réfl}}$ , le parcours positif de la relation sémantique 2 ou 3 (dont le dépendant est identique au dépendant de la relation sémantique 1) produit le grammème de voix réfléchie. Nous avons présenté la règle  $\mathbf{R}_{+2/3}^{\text{réfl}}$  ici, car il s'agit d'une règle sagittale (elle consomme un arc sémantique comme les règles sagittales standard), mais ce n'est pas une règle d'arborisation — c'est une règle (de lexicalisation) flexionnelle. De plus, cette règle doit être prioritaire sur la règle  $\mathbf{R}_{+\geq 2}$  afin d'éviter la production de phrase comme \**Pierre rase lui-même*<sup>14</sup>.

4) Les règles  $\mathbf{R}_{+1}$ ,  $\mathbf{R}_{+1}^{\text{pass}}$ ,  $\mathbf{R}_{+\geq 2}$  et  $\mathbf{R}_{+2}^{\text{pass}}$  contrôlent la diathèse de  $L(\langle P \rangle)$  de la façon suivante. Par exemple, si le parcours positif d'une relation sémantique 1 produit une relation SyntP I (Fig. 16), alors  $L(\langle P \rangle)$  ne doit pas être à la voix passive. Par contre, si ce parcours produit une relation SyntP II (Fig. 17), le verbe doit être au passif (ce qui n'est bien sûr possible que pour des verbes transitifs passivables). Néanmoins, telles que nous les avons présentées, ce ne sont pas ces règles qui introduisent les grammèmes de l'actif et du passif. Pour dire vrai, il y a ici un point que nous n'avons pas entièrement résolu. Ou bien nous devons écrire des règles séparées qui produisent ces grammèmes de voix (ce que nous n'avons pas fait), ou bien nous devons attacher la production de ces grammèmes à l'une des règles, probablement à celle qui consomme l'arc sémantique 2 (étant donné que le passif est avant tout une opération morphologique servant à promouvoir des compléments en position sujet).

5) L'affirmation que seule une relation sémantique 1 peut donner une relation SyntP ATTR doit être modulée. La situation réelle est plus complexe : certaines langues possèdent un outil pour exprimer par une relation SyntP ATTR a priori n'importe quelle relation sémantique. Cet outil est la voix grammaticale : en changeant la diathèse de la lexie  $L(\langle s \rangle)$ , la voix peut en principe faire correspondre à la relation SyntP I n'importe quelle relation Sém  $r$  issue de  $\langle s \rangle$  — en effectuant une conversion des actants SyntP. En français, seule la relation sémantique 2 peut être concernée par une telle conversion (la passivisation). Cette conversion est réalisée directement par les règles d'arborisation : par exemple, en français, un parcours positif d'une relation Sém

14. Notons que notre règle de réflexivisation est assez grossière. D'une part, elle ne couvre pas tous les cas où la voix réfléchie est produite : dans *Pierre se construit une maison*, le réfléchi correspond au bénéficiaire de l'action qui n'est pas un actant sémantique de  $\langle \text{construire} \rangle$  ; dans *Pierre s'est fait raser*,  $\langle \text{Pierre} \rangle$  est l'actant de deux sémantèmes distincts,  $\langle \text{causer} \rangle$  (= le causatif) et  $\langle \text{raser} \rangle$ . D'autre part, le cas où trois actants d'un sémantème réalisé par un verbe référent à une même entité, comme dans *Pierre se vend à lui-même*, n'est pas clairement pris en compte ; dans ce cas, la règle ne doit s'appliquer qu'une seule fois.

2 donnera une relation SyntP I si le verbe en cause est à la voix passive (Fig. 22). Les mêmes changements de diathèse de  $L(\langle s \rangle)$  permettent un parcours négatif d'une relation sémantique  $r$  (en français,  $r = 2$ ) et la production d'une relation SyntP ATTR, dont  $L(\langle s \rangle)$  est alors le dépendant. D'un point de vue formel, tout se passe comme si la conversion avait lieu dans la SSém : la conversion attribue l'étiquette sémantique 1 à l'arc considéré, puis selon son sens de parcours, cet arc donne une relation SyntP I ou ATTR (par la règle  $\mathbf{R}_{+1}$  ou par la règle  $\mathbf{R}_{-1(\text{Adj})}$ ). Néanmoins, d'un point de vue linguistique, le changement de diathèse doit être envisagé en SSyntP : il concerne des lexies et non pas des sémantèmes. On aura donc une règle combinée effectuant simultanément le changement de diathèse et le parcours négatif de l'arc en question : c'est la règle  $\mathbf{R}_{-2(\text{Adj})}^{\text{pass}}$  pour le français, assurant le parcours négatif d'un arc Sém 2 à condition que le nœud  $\langle P \rangle$  gouvernant cet arc soit lexicalisé par un verbe transitif à la voix passive (Fig. 23).

### 2.3.2. Règles d'arborisation discontinues non locales : règles d'extraction

Comme nous l'avons dit dans l'Introduction, les phrases à extraction présentent des discontinuités non bornées : deux nœuds directement liés dans la SSém donnent deux nœuds de la SSyntP qui sont reliés par une chaîne non bornée de dépendances (et vice versa). De telles phrases doivent donc être traitées par des règles discontinues non locales. Nous allons présenter maintenant deux **règles d'extraction**, qui sont donc des règles discontinues non locales : la **règle de relativisation**  $\mathbf{R}^{\text{rel}}$  (Fig. 24) et la **règle d'interrogation indirecte**  $\mathbf{R}^{\text{interrog}}$  (Fig. 25). Nous discuterons ensuite les trois contraintes apparaissant dans l'énoncé des règles, puis nous expliquerons plus précisément le fonctionnement des règles.

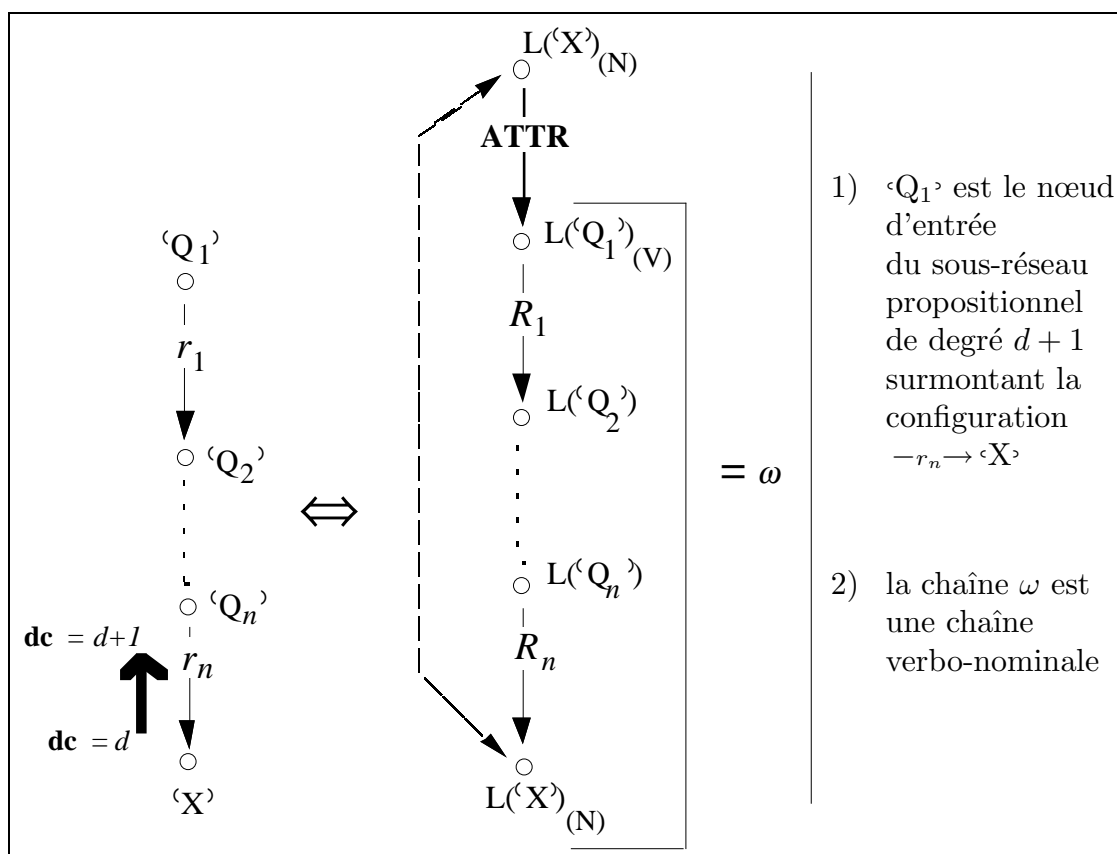
#### Contrainte 1 : $\langle X \rangle \supseteq \langle \text{quel} \rangle$

Cette première contrainte concerne uniquement la règle d'interrogation  $\mathbf{R}^{\text{interrog}}$ . La formule " $\langle X \rangle \supseteq \langle \text{quel} \rangle$ " signifie soit que  $\langle X \rangle = \langle \text{quel} \rangle$ , soit que la décomposition sémantique de  $\langle X \rangle$  contient le sémantème  $\langle \text{quel} \rangle$  comme nœud dominant. Autrement dit,  $\langle X \rangle$  est le signifié d'un pronom interrogatif.

#### Contrainte 2 : le degré communicatif

Le degré (communicatif) d'un sous-réseau propositionnel a été défini au §1.1.2 : il s'agit du degré d'enchâssement de la division thématique de ce sous-réseau ; la division thématique primaire, associée à la RSém complète, a le degré 1.

Lorsque le passage d'une RSém à une SSyntP débute, c'est la division thématique primaire et elle seule qui "pilote" la synthèse ; en particulier, le nœud d'entrée de la RSém complète est calculé en fonction de la division thématique primaire. A chaque fois que débute la synthèse d'une proposition enchâssée, c'est la division thématique du sous-réseau propositionnel concerné qui prend (momentanément) le relais : en particulier, le nœud d'entrée de ce sous-réseau propositionnel est calculé en fonction de sa division thématique. Nous appelons **degré courant** le degré de la division thématique qui, au moment considéré du processus de synthèse, "pilote" la synthèse (nous verrons au §3.1 de quelle façon est réalisé ce "pilotage"). Le degré courant est noté **dc**.


 Figure 24. La règle de relativisation  $R^{\text{rel}}$ 
**Exemples d'application de la règle  $R^{\text{rel}}$** 

- a. le type  $[L(\langle X \rangle)]$  qui [= type =  $L(\langle X \rangle)$ ] veut  $[L(\langle Q_1 \rangle)]$  acheter la maison d'à côté;  $n = 1$ .
- b. la fille  $[L(\langle X \rangle)]$  à laquelle [= fille =  $L(\langle X \rangle)$ ] mon ami parle  $[L(\langle Q_1 \rangle)]$ ;  $n = 1$ .
- c. la fille  $[L(\langle X \rangle)]$  au père  $[L(\langle Q_2 \rangle)]$  de laquelle [= fille =  $L(\langle X \rangle)$ ] mon ami parle  $[L(\langle Q_1 \rangle)]$ ;  $n = 2$ .
- d. le type  $[L(\langle X \rangle)]$  que [= type =  $L(\langle X \rangle)$ ] mon ami cherche  $[L(\langle Q_1 \rangle)]$  à rencontrer  $[L(\langle Q_2 \rangle)]$ ;  $n = 2$ .
- e. la fille  $[L(\langle X \rangle)]$  que [= fille =  $L(\langle X \rangle)$ ] Jean pense  $[L(\langle Q_1 \rangle)]$  que Pierre veut  $[L(\langle Q_2 \rangle)]$  épouser  $[L(\langle Q_3 \rangle)]$ ;  $n = 3$ .
- f. le livre  $[L(\langle X \rangle)]$  dont [= livre =  $L(\langle X \rangle)$ ] Pierre sait  $[L(\langle Q_1 \rangle)]$  quelle page il a arraché  $[L(\langle Q_2 \rangle)]$ ;  $n = 2$ .

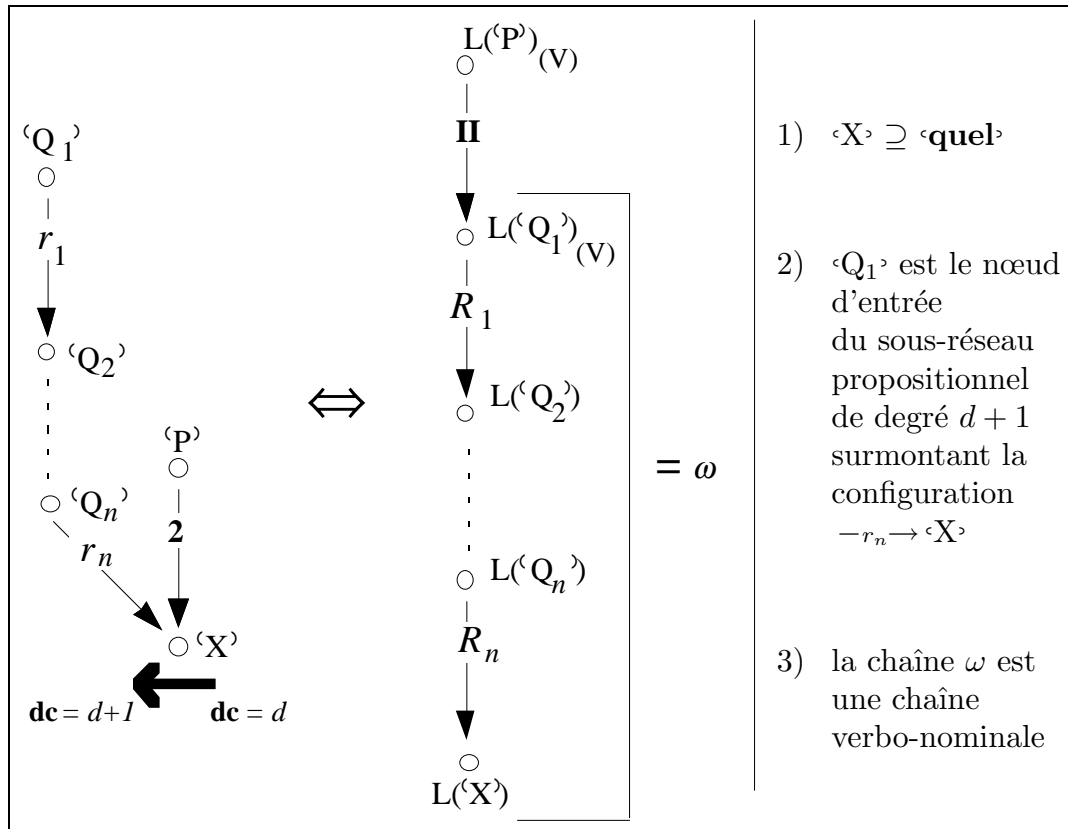


Figure 25. La règle d'interrogation indirecte  $R^{\text{interrog}}$

**Attention :** 1) Contrairement aux apparences, la règle  $R^{\text{interrog}}$  est une règle nodale, bien que particulière : elle ne consomme pas le nœud  $\langle X \rangle$  qui la déclenche. La grosse flèche représentant le sens de parcours indique que cette règle est déclenchée lorsque  $\langle X \rangle$  est atteint pour la première fois — après la lecture de l'arc Sém 2 allant de  $\langle P \rangle$  à  $\langle X \rangle$  et l'application de la règle  $R_{+2}$  qui produit la branche SyntP II issue de  $L(\langle P \rangle)$ . Ceci est confirmé par le fait que le degré courant  $dc$  est égal à  $d$  au moment du déclenchement de la règle.

2) La relation sémantique  $r_n$  (figurant dans la partie gauche de la règle  $R^{\text{interrog}}$ ) peut être orientée dans n'importe quel sens, en particulier dans le sens inverse de celui choisi pour l'illustration ci-dessus. C'est le cas lorsque  $\langle X \rangle$  est un prédicat (comme  $\langle \text{quel} \rangle$ ,  $\langle \text{combien} \rangle$  ou  $\langle \text{où} \rangle$ ).

#### Exemples d'application de la règle $R^{\text{interrog}}$

- a. Je me demande  $[L(\langle P \rangle)]$  à qui  $[L(\langle X \rangle)]$  mon ami parle  $[L(\langle Q_1 \rangle)]$  ;  $n = 1$ .
- b. Je me demande  $[L(\langle P \rangle)]$  à quelle  $[L(\langle X \rangle)]$  fille  $[L(\langle Q_2 \rangle)]$  mon ami parle  $[L(\langle Q_1 \rangle)]$  ;  $n = 2$ .
- c. Je me demande  $[L(\langle P \rangle)]$  au père  $[L(\langle Q_2 \rangle)]$  de quelle  $[L(\langle X \rangle)]$  fille  $[L(\langle Q_3 \rangle)]$  mon ami parle  $[L(\langle Q_1 \rangle)]$  ;  $n = 3$ .
- d. Je me demande  $[L(\langle P \rangle)]$  combien  $[L(\langle X \rangle)]$  de personnes  $[L(\langle Q_3 \rangle)]$  mon ami doit  $[L(\langle Q_1 \rangle)]$  rencontrer  $[L(\langle Q_2 \rangle)]$  ;  $n = 3$ .

- e. *Je me demande* [L(⟨P⟩)] *quelle* [L(⟨X⟩)] *fille* [L(⟨Q<sub>4</sub>⟩)] *Jean pense* [L(⟨Q<sub>1</sub>⟩)] *que Pierre veut* [L(⟨Q<sub>2</sub>⟩)] *épouser* [L(⟨Q<sub>3</sub>⟩)];  $n = 4$ .
- f. *Je prends le livre dont Pierre sait* [L(⟨P⟩)] *quelle* [L(⟨X⟩)] *page* [L(⟨Q<sub>2</sub>⟩)] *il a arraché* [L(⟨Q<sub>1</sub>⟩)];  $n = 2$ .

Lorsqu'on déclenche une règle **R** discontinue, en l'occurrence la règle de relativisation ou d'interrogation indirecte, on s'apprête à débiter la synthèse d'une nouvelle proposition (relative ou interrogative indirecte) enchâssée dans la proposition en cours de synthétisation. Pour cette raison, le degré courant doit être incrémenté de 1 au cours de l'application de la règle **R**. C'est ce qu'indiquent les valeurs  $\mathbf{dc} = d$  et  $\mathbf{dc} = d + 1$  aux deux extrémités de la grosse flèche dans la partie gauche de **R**. La valeur finale de  $\mathbf{dc}$  égale à  $d + 1$  est le degré d'enchâssement du sous-réseau propositionnel *S* qui donnera la proposition (relative ou interrogative) à synthétiser.

Le sous-réseau *S* est caractérisé, dans nos règles d'extraction, comme étant le seul sous-réseau propositionnel de degré  $d + 1$  surmontant la configuration  $-r \rightarrow \langle X \rangle$ , où  $\langle X \rangle$  est le nœud Sém donnant l'antécédent de la relative et  $-r \rightarrow$  est l'arc Sém qu'on est en train de considérer quand la règle est déclenchée. On dit qu'un sous-réseau propositionnel surmonte une configuration  $-r \rightarrow \langle X \rangle$  si ce sous-réseau contient le nœud  $\langle X \rangle$  et l'arc  $-r \rightarrow$  qui y entre sans contenir l'arc par lequel on arrive sur  $\langle X \rangle$ . On notera que, dans la SSém, il existe nécessairement un unique réseau de ce type qui soit de degré  $d + 1$ . Illustrons ce point par la RSém de la phrase (3) (Fig. 8) : lorsque  $\langle X \rangle = \langle \text{livre} \rangle$  et l'arc Sém considéré est l'arc 3 (de  $\langle \text{arracher} \rangle$  à  $\langle \text{livre} \rangle$ ), on peut déclencher la règle **R<sup>rel</sup>** ; dans ce cas, il existe deux sous-réseaux propositionnels surmontant la configuration  $-3 \rightarrow \langle \text{livre} \rangle$ , à savoir 1) la somme de **R<sub>2</sub>** et **T<sub>2</sub>** et 2) la somme de **R<sub>3</sub>** et **T<sub>3</sub>** ; lorsque la règle est déclenchée, le degré courant vaut 2 ( $= d$ ) et c'est donc le sous-réseau propositionnel **R<sub>3</sub>+T<sub>3</sub>** de degré 3 ( $= d + 1$ ) qui sera pris en compte pour le choix du nœud d'entrée  $\langle Q_1 \rangle$  (qui sera  $\langle \text{savoir} \rangle$ ).

### Contrainte 3 : la chaîne verbo-nominale

La formulation des contraintes des règles d'extraction appelle un concept syntaxique complexe — la **chaîne verbo-nominale** —, définie à partir de deux autres concepts, le **nucléus verbal** et le **nucléus nominal** (introduits dans Kahane 1996, 1997)<sup>15</sup>.

L'intérêt des trois concepts n'est pas limité aux règles d'extraction ; ils sont nécessaires pour la description de plusieurs phénomènes syntaxiques des langues naturelles. En effet, les concepts de nucléus verbal et de nucléus nominal représentent une généralisation des notions de verbe et de nom. Regardons, par exemple, le cas du verbe : dans de nombreuses constructions, une position syntaxique où un seul verbe apparaît peut également accueillir toute une suite de verbes ou de tournures verbales équivalente à ce seul verbe du point de vue

15. Le terme *nucléus* apparaît pour la première fois dans Tesnière (1959 : 44), où il est introduit pour résoudre des problèmes de non correspondance entre structure sémantique et structure syntaxique.

des manipulations formelles qu'elle subit ou déclenche dans la phrase (cf. *la fille que Pierre aide* et *la fille que Pierre a l'intention de faire l'effort de continuer à aider*). Une telle suite est formalisée par le concept de nucléus verbal<sup>16</sup>.

Citons quelques constructions syntaxiques où le nucléus verbal joue un rôle crucial :

(17) a. la coordination (sans ellipse)

*Pierre écrit et parle à Jean.* vs  
*Pierre écrit et a envie de parler à Jean.*

[Le verbe PARLER est syntaxiquement équivalent au nucléus verbal AVOIR ENVIE DE PARLER.]

b. la coordination (avec ellipse)

*Pierre parle à Jean et Igor [parle] à Sylvain.* vs  
*Pierre a envie de parler à Jean et Igor [a envie de parler] à Sylvain.*

c. l'accord de l'adjectif attributif

*Marie reste belle.* vs  
*Marie a envie de rester belle.*

d. la négation (*ne* + forclusif)

*Pierre ne parle à personne.* vs  
*Pierre n'a envie de parler à personne.*

e. la comparaison (avec ellipse)

*Pierre a acheté le même livre que Jean [a acheté].* vs  
*Pierre a envie d'acheter le même livre que Jean [a envie d'acheter].*

f. la cooccurrence lexicale restreinte (décrite en termes de **fonctions lexicales**)

*Un incendie se déclare ici.* vs  
*Un incendie a toutes les chances de se déclarer ici.*  
[*se déclarer* = **IncepFunc**<sub>0</sub>(*incendie*)].

Comme on le voit, le concept de nucléus verbal est un outil nécessaire pour toute théorie syntaxique générale, indépendamment du cadre de référence choisi.

**Définition 2.** La chaîne SyntP  $L(\langle Q_1 \rangle) -_{R_1} \rightarrow \dots -_{R_{p-1}} \rightarrow L(\langle Q_p \rangle)$  est appelée un *nucléus verbal* si et seulement si elle satisfait les trois conditions suivantes :

- 1)  $L(\langle Q_1 \rangle)$ , la tête de la chaîne, est un verbe ;
- 2) chaque  $L(\langle Q_k \rangle)$  ( $k = 2, \dots, p$ ) est

---

16. Mel'čuk & Pertsov (1987 : 489) introduisent une notion similaire au nucléus verbal, qu'ils nomment un arbre standard  $\Delta VP$ . "With respect to its environment  $\Delta VP$  behaves like a single verb [...] E.g., *may seem likely to contribute* is syntactically equivalent, as regards the combinability with possible grammatical subjects, objects and complements, to the verb *contribute* [...]" (*ibid.*). Cette notion est néanmoins plus restrictive que le nucléus verbal et, bien que fort utilisée dans la description de la syntaxe de l'anglais proposée par Mel'čuk et Pertsov, elle n'est pas mise en valeur.

- soit un verbe appartenant à un ensemble désigné de verbes appelés **verbes ponts**<sup>17</sup> ;
  - soit un nom, et alors le gouverneur  $L(\langle Q_{k-1} \rangle)$  du nom  $L(\langle Q_k \rangle)$  est obligatoirement un verbe support<sup>18</sup> ;
  - soit un adjectif, et alors le gouverneur  $L(\langle Q_{k-1} \rangle)$  de l'adjectif  $L(\langle Q_k \rangle)$  est obligatoirement un verbe copulatif (*être, devenir, trouver, avoir l'air, ...*) ;
- 3) les relations SyntP  $R_k$  ( $k = 1, \dots, p - 1$ ) sont toutes actancielles.

La Définition 2, de même que les Définitions 3 et 4, s'applique à des chaînes SyntP qui ne contiennent pas de chaînes coordinatives. Dans le cas général, il est toujours possible qu'un nœud d'une chaîne syntaxique quelconque soit remplacé par une chaîne coordinative : ainsi, dans le syntagme *le livre que Marie pense que Pierre a acheté ou que Jean lui a prêté*, la chaîne coordinative *a acheté ou a prêté* joue le même rôle syntaxique qu'un seul verbe (*a acheté*). Il est possible d'intégrer la coordination dans ces définitions, mais nous ne le ferons pas ici<sup>19</sup>.

---

17. On appelle traditionnellement verbes ponts les verbes ou tournures verbales qui autorisent l'extraction hors de la complétive qu'ils sous-catégorisent (cf., par exemple, Godard 1988 : 37). Remarquons que la propriété "être un verbe pont" est graduelle (cf. l'exemple (i), emprunté à Godard, *op. cit.*) :

- (i) *La SNCF, que ce journaliste disait/? confiait/\* hurlait que les grèves allaient tuer, a dû effectivement être renflouée.*

Dans un nucléus verbal, on peut accepter des verbes non ponts, à la condition toutefois d'observer des contraintes additionnelles au niveau SyntS (voir Annexe B).

18. Les verbes supports associés à un nom  $L$  sont indiqués dans l'article de dictionnaire de  $L$ . Ceux-ci sont les valeurs de certaines fonctions lexicales spécifiques comme **Oper**<sub>1</sub>, **Oper**<sub>2</sub>, ... (l'indice  $i$  de **Oper** <sub>$i$</sub>  indique que c'est l'actant Sém  $i$  de  $\langle L \rangle$  qui devient le sujet syntaxique de **Oper** <sub>$i$</sub> ( $L$ )). Exemples : **Oper**<sub>1</sub>(INTENTION) = AVOIR : *un livre que Pierre a l'intention de lire* ; **Oper**<sub>1</sub>(EFFORT) = FAIRE : *un livre que Pierre a fait l'effort de lire* ; **Oper**<sub>2</sub>(DIFFICULTÉ) = AVOIR, ÉPROUVER : *un livre que Pierre a/éprouve beaucoup de difficultés à lire*. Il est important de souligner que la distinction entre verbes libres et verbes phraséologisés s'avère cruciale pour le traitement des extractions.

19. Comme on le sait, la coordination est "perpendiculaire" à la subordination en général. Pour cette raison, les difficultés liées à la description des constructions coordinatives ne sont pas spécifiques au cas particulier de l'extraction, de sorte que nous pouvons en faire abstraction dans cet article. Voici cependant quelques remarques sur les points où l'extraction et la coordination se touchent. On peut "extraire" le complément commun à des verbes coordonnés (i-a,b,c), mais pas le complément d'un des verbes coordonnés sans que ce complément soit partagé avec les autres verbes de la coordination (i-d,e) :

- (i) **a.** *les carottes que Jean a achetées et pelées*  
**b.** *les carottes que Jean a achetées, a voulu peler et pensait faire bouillir*  
**c.** *les carottes que je crois que Jean a achetées, que Paul a voulu peler et que Marie pensait faire bouillir*  
**d.\*** *les carottes que Jean a achetées et pelé les pommes*  
**e.\*** *les carottes que Jean a achetées et leur a ajouté les pommes*

Le complément est dit **partagé** par une chaîne de verbes coordonnés si ce complément apparaît dans la SSyntP seulement sous un seul de ces verbes tout en étant prévu dans la valence de tous ces verbes ; à noter qu'un complément partagé par des verbes coordonnés doit correspondre au régime de chacun des verbes. De plus, ce qui complique largement la chose, c'est que chaque élément d'une chaîne coordinative peut être un nucléus verbal, comme c'est le cas dans les exemples (i-b,c), où sont coordonnés les nucléus verbaux ACHETER, VOULOIR PELER et PENSER FAIRE BOUILLIR. Une solution à ce problème est proposée dans Kahane 1997. L'idée est d'enrichir la structure SyntP en en faisant un **arbre à bulles** : une chaîne coordinative est alors représentée par une bulle occupant un nœud de l'arbre et les compléments partagés par les éléments coordonnés

**Définition 3.** La chaîne SyntP  $L(\langle Q_{p+1} \rangle) - R_{p+1} \rightarrow \dots - R_{n-1} \rightarrow L(\langle Q_n \rangle)$  est appelée un *nucléus nominal* si et seulement si les maillons de la chaîne sont de la forme  $N_1 - \Pi \rightarrow N_2$ , où  $N_1$  et  $N_2$  sont des noms, avec deux cas particuliers : le premier maillon peut être de la forme  $\text{PRÉP} - \Pi \rightarrow N$ , où  $\text{PRÉP}$  est une préposition pleine et  $N$  est un nom, et le dernier maillon peut être de la forme  $N_1 - \text{ATTR} \rightarrow \text{DE} - \Pi \rightarrow N_2$ , où  $\text{DE}$  est la préposition pleine signifiant ‘appartenir’.

Exemples :

- (18) a. *J'aimerais savoir **à l'attaque contre qui** il a réagi.*  
 b. *Mon ami est l'homme **à la fille de qui** tu as parlé.*  
 c. *J'aimerais savoir **avec la fille de qui** tu es parti.*  
 d. *Je connais l'homme **dont** Marie a visité **la maison**.*  
 e. *Je connais l'homme **dans la maison duquel** tu as dormi.*

La réalisation de surface des nucléus nominaux est en gras. Les prépositions À et CONTRE en (18a) et À et DE en (18b) sont des régimes qui marquent les relations SyntP actancielles et n'apparaissent qu'en surface. Les prépositions pleines AVEC et DANS en (18c) et (18e) constituent le premier élément du nucléus nominal. En (18d) et (18e), le dernier maillon du nucléus est de la forme  $\text{MAISON} - \text{ATTR} \rightarrow \text{DE} - \Pi \rightarrow \text{HOMME}$ , où  $\text{DE}$  est la préposition pleine signifiant ‘appartenir’ ; le segment  $\text{DE} - \Pi \rightarrow \text{HOMME}$  a donné en surface DONT en (18d) et DUQUEL en (18e).

Deux commentaires sur la Définition 3 :

1) La présence d'un maillon  $N_1 - \text{ATTR} \rightarrow \text{DE} - \Pi \rightarrow N_2$  dans une chaîne nominale réduit l'acceptabilité d'une phrase à extraction. Lorsque ce maillon est en dernière position de la chaîne nominale, l'acceptabilité est néanmoins meilleure :

- (19) a. *Je me demande **de quel homme** tu attends l'arrivée de l'épouse.*  
 [ARRIVÉE -  $\Pi$  → EPOUSE -  $\Pi$  → HOMME]  
 b. *Je me demande **de quel homme** tu as pris la couverture du livre.*  
 [COUVERTURE -  $\Pi$  → LIVRE -  $\text{ATTR}$  → DE -  $\Pi$  → HOMME]  
 c. \* *Je me demande **de quel homme** tu as pris le livre de l'épouse.*  
 [LIVRE -  $\text{ATTR}$  → DE -  $\Pi$  → ÉPOUSE -  $\Pi$  → HOMME]

D'après la Définition 3, les chaînes nominales en (19a) et (19b) sont des nucléus nominaux, mais pas celle en (19c), et la phrase à extraction (19c) est donc rejetée<sup>20</sup>.

---

sont attribués à la bulle ; la contrainte sur un nucléus est alors exprimée sur la chaîne de ses nœuds, indépendamment du fait qu'il s'agisse d'un nœud simple ou d'une bulle coordinative. Les nucléus sont également représentés par des bulles (d'un autre type que les bulles coordinatives) et ces bulles peuvent appartenir à une chaîne coordinative au même titre que des nœuds simples. En ajoutant à la structure syntaxique une dimension — la bulle —, on peut rendre compte du caractère ‘orthogonal’ de la coordination et de la subordination. Une telle solution ne nécessite pas l'addition de contraintes complexes, comme c'est le cas si l'on se limite structurellement à un arbre de dépendance.

20. Les contraintes de linéarisation ne sont pas en cause, puisqu'aucune des linéarisations envisageables pour la SSyntP de (19a) n'est acceptable :



## SYNTHÈSE DES PHRASES À EXTRACTION

2) La présence d'un maillon  $N_1 - \text{ATTR} \rightarrow \text{PRÉP} - \text{II} \rightarrow N_2$  avec une préposition PRÉP autre que DE réduit également l'acceptabilité<sup>21</sup> :

(20) a. *Je me demande la robe de qui tu as abimée.*

b. <sup>??</sup> *Je me demande la robe pour qui tu as abimée.*

Cela reflète le caractère très grammaticalisé de la préposition DE : même dans ses emplois "sémantiques", elle se rapproche d'un marqueur syntaxique. Pour cette raison, parmi les prépositions pleines, seule DE est acceptable dans le dernier maillon du nucléus nominal<sup>22</sup>.

**Définition 4.** Une chaîne SyntP  $\omega = L(\langle Q_1 \rangle) - R_1 \rightarrow L(\langle Q_2 \rangle) - R_2 \rightarrow \dots - R_{n-1} \rightarrow L(\langle Q_n \rangle) - R_n \rightarrow L(\langle X \rangle)$  est appelée une *chaîne verbo-nominale* si et seulement si elle peut être partitionnée en deux sous-chaînes  $\omega_1$  et  $\omega_2$  telles que :

a. la chaîne  $\omega_1 = L(\langle Q_1 \rangle) - R_1 \rightarrow \dots - R_{p-1} \rightarrow L(\langle Q_p \rangle)$  est un nucléus verbal ( $1 \leq p \leq n$ ) ;

b. la chaîne  $\omega_2 = L(\langle Q_{p+1} \rangle) - R_{p+1} \rightarrow \dots - R_{n-1} \rightarrow L(\langle Q_n \rangle) - R_n \rightarrow L(\langle X \rangle)$  est :  
 - un nucléus nominal, si  $L(\langle X \rangle)$  est un nom ;  
 - un nucléus nominal à l'exclusion de son dernier maillon, si  $r_n = \text{ATTR}$ , c'est-à-dire si  $L(\langle X \rangle)$  est un élément pronominal (comme QUEL ou COMBIEN) qui dépend de son gouverneur  $L(\langle Q_n \rangle)$  par une relation SyntP ATTR.

Exemple :

(21) *Je me demande avec le mari de quelle amie Marie a l'intention de parler.*

(i) a. \* *Je me demande de quel homme tu as pris le livre de l'épouse.*

b. <sup>??</sup> *Je me demande de l'épouse de quel homme tu as pris le livre.*

c. \* *Je me demande le livre de l'épouse de quel homme tu as pris.*

21. En fait, la Définition 3 doit être raffinée pour prendre en charge des cas tels que (i) :

(i) a. *Je me demande la robe pour qui tu as achetée.*

b. <sup>??</sup> *Je me demande la robe pour qui tu as abimée.*

Selon cette définition, le groupe nominal *la robe pour qui* n'est pas un nucléus nominal — à cause de la préposition pleine POUR ; les deux phrases de (i) devraient donc être mauvaises au même degré. Cependant, (i-a) est bonne, alors que (i-b) est à peu près inacceptable. Nous croyons que cela s'explique par le phénomène syntaxique dit d'**attraction**. Dans (i-a), le complément en POUR, bien que régi par le nom ROBE (tout comme en (i-b)) est "attiré" par le verbe ACHETER, qui a dans sa valence un complément optionnel en POUR (*acheter qqch pour qq'un*) ; cette attraction par le verbe principal améliore l'acceptabilité.

22. Avec des prépositions vides (qui marquent une relation actancielle SyntP et n'apparaissent qu'en SSyntS), on observe également un contraste entre les prépositions très grammaticalisées DE et À et les autres prépositions : ainsi, (i-a) est perçue comme meilleure que (i-b).

(i) a. *Je me demande l'amour de qui l'a poussé au suicide.*

b. *Je me demande l'amour pour qui l'a poussé au suicide.*

Néanmoins, cette différence d'acceptabilité étant minimale, nous en faisons abstraction et nous admettons dans la définition du nucléus nominal toutes les relations actancielle quelle que soit la préposition qui les exprime. On aura noté que la distinction entre prépositions vides et prépositions pleines, que peu de modèles linguistique font systématiquement, joue un rôle important dans la description de nombreux phénomènes linguistiques, dont les extractions.

Dans cet exemple,  $p = 3$  et  $n = 5$  :  $a_1$  *l'intention*<sub>2</sub> [*de*] *parler*<sub>3</sub> est un nucléus verbal,  $a_4$  *le mari*<sub>4</sub> [*de quelle*] *amie*<sub>5</sub> est un nucléus nominal et  $a$  *l'intention* [*de parler*] [*avec*] *le mari* [*de quelle*] *amie* une chaîne verbo-nominale.

Trois commentaires sur la Définition 4 :

1) Une chaîne verbo-nominale est la combinaison d'un nucléus verbal, qui "équivaute" à un verbe, et d'un nucléus nominal, qui, lui, "équivaute" à un nom. Notre définition prévoit le cas général, qui, dans un sens métaphorique, se ramène au cas simple d'un verbe gouvernant un nom. Dans ce cas, formellement, on pourrait utiliser des règles (discontinues) locales pour la relativisation et l'interrogation indirecte. C'est la capacité de ces deux éléments de base d'être développés de façon itérative et non bornée qui entraîne la non-localité de nos règles.

2) La relation  $R_p$  entre les deux nucléus  $\omega_1$  et  $\omega_2$  —  $L(\langle Q_p \rangle) - R_p \rightarrow L(\langle Q_{p+1} \rangle)$  — n'est pas contrainte : elle peut être actancielle, comme en (21), ou attributive. Dans ce dernier cas, le nucléus nominal débute par une préposition pleine  $L(\langle Q_{p+1} \rangle)$ , comme en (18c,e).

3) La complication dans le cas **b** de la Définition 4 est due au fait que, à la différence des pronoms subordonatifs nominaux, les pronoms du type QUEL ou COMBIEN, qui sont adjectivaux<sup>23</sup> (au niveau SyntP), ne font pas partie du nucléus nominal. Autrement dit, dans notre présentation, un nucléus nominal inclut un pronom subordonatif, si ce pronom est nominal, mais gouverne un pronom subordonatif extérieur, si ce pronom est adjectival. Il est clair que nous aurions pu faire un choix différent : inclure le pronom subordonatif adjectival dans le nucléus nominal (sans y inclure pour autant les autres dépendants adjectivaux). Le premier choix nous semble plus naturel, notamment parce qu'il préserve le parallélisme avec le nucléus verbal.

L'application d'une règle d'extraction contraint que la chaîne SyntP produite après soit une chaîne verbo-nominale ( $\mathbf{R}^{\text{rel}}$ , Condition 2 et  $\mathbf{R}^{\text{interrog}}$ , Condition 3). De façon informelle, cette condition peut se formuler comme suit :

|| L'extraction d'un élément n'est possible qu'à partir d'un nucléus nominal dépendant directement du nucléus verbal principal de la phrase.

On peut illustrer cela par deux exemples (quasi) synonymes :

(22) **a.** *Je me demande sans l'ami de qui il veut partir.*

**b.\*** *Je me demande sans emmener l'ami de qui il veut partir.*

Dans (22a), l'élément extrait QUI appartient un nucléus nominal *sans l'ami de qui* dépendant directement du nucléus verbal principal *veut partir* ; dans (22b), l'élément extrait QUI dépend du verbe EMMENER qui n'appartient pas à nucléus verbal (à cause de la préposition pleine SANS, la chaîne *veut partir sans emmener* n'est pas un nucléus verbal).

Attention : les contraintes que la définition de la chaîne verbo-nominale impose aux règles d'extraction ne sont suffisantes qu'au niveau syntaxique

23. On n'a pas de pronoms relatifs adjectivaux en français. Mais, on aurait le même problème avec le pronom relatif adjectival (au niveau SyntP) WHOSE 'de qui' de l'anglais ou CUYO 'de qui' de l'espagnol.

profond. Pour assurer les transitions correctes aux niveaux syntaxique de surface et morphologique profond, d'autres contraintes, reléguées aux modules concernés, sont nécessaires (voir Annexe B).

### Commentaires sur le fonctionnement des règles d'extraction

Remarquons, pour commencer, que dans les deux règles d'extraction énoncées —  $\mathbf{R}^{\text{rel}}$  (Fig. 24) et  $\mathbf{R}^{\text{interrog}}$  (Fig. 25) —, on voit apparaître des lexies  $L(\langle Q_k \rangle)$  ( $k = 1, \dots, n$ ) et des relations SyntP  $R_k$  ( $k = 1, \dots, n$ ) qui ne sont pas créées par la règle. La règle se contente de prévoir leur existence; entre les deux parties de la SSyntP réellement contrôlées par la règle, il doit être créé un nœud verbal  $L(\langle Q_1 \rangle)$  et éventuellement d'autres nœuds  $L(\langle Q_2 \rangle)$ ,  $\dots$ ,  $L(\langle Q_n \rangle)$ . On voit également que la règle contraint la nature de la chaîne  $L(\langle Q_1 \rangle) - R_1 \rightarrow \dots - R_{n-1} \rightarrow L(\langle Q_n \rangle) - R_n \rightarrow L(\langle X \rangle)$ , même si cette chaîne est une chaîne "fantôme" — un moule qui devra être rempli par l'application d'autres règles de correspondance. Nous donnerons une description procédurale du traitement de cette chaîne "fantôme" au §3.1. A noter que la chaîne Sém  $\langle Q_1 \rangle - r_1 \rightarrow \dots - r_{n-1} \rightarrow \langle Q_n \rangle - r_n \rightarrow \langle X \rangle$  n'est une chaîne continue que si aucune règle discontinue n'est appliquée lors de la synthèse de la chaîne SyntP  $L(\langle Q_1 \rangle) - R_1 \rightarrow \dots - R_{n-1} \rightarrow L(\langle Q_n \rangle) - R_n \rightarrow L(\langle X \rangle)$ ; par exemple, dans le cas de la synthèse de la phrase (3), du fait qu'une interrogative est enchâssée dans la relative, la chaîne sémantique menant du nœud d'entrée  $\langle Q_1 \rangle = \langle \text{savoir} \rangle$  de la relative au nœud  $\langle X \rangle = \langle \text{livre} \rangle$  n'est pas continue.

La règle de relativisation  $\mathbf{R}^{\text{rel}}$  se caractérise par le fait qu'elle ne consomme pas la relation sémantique  $r_n$  qui la déclenche. (Cette relation sera consommée ultérieurement par une règle positive —  $\mathbf{R}_{+1}$  si  $r_n = 1$  ou  $\mathbf{R}_{+\geq 2}$  si  $r_n \geq 2$ , ou une autre règle en cas de changement de diathèse —, ce qui donnera une relation SyntP  $R_n$ .) Bien qu'elle ne consomme aucun élément Sém, la règle  $\mathbf{R}^{\text{rel}}$  produit une relation SyntP ATTR, qui subordonne l'ensemble du sous-arbre relativisé et relie son sommet au nom modifié (= l'antécédent du pronom relatif). La relation  $r_n$  sera donc parcourue deux fois et donnera lieu à deux relations SyntP : la relation ATTR produite par la règle  $\mathbf{R}^{\text{rel}}$  et la relation  $R_n$  produite par la règle positive qui consomme  $r_n$  pendant la synthèse de la relative. C'est ce "dédoublement" d'un lien sémantique en deux liens SyntP qui induit le caractère non local de la règle  $\mathbf{R}^{\text{rel}}$ ; a priori, un tel dédoublement n'entraîne pas nécessairement la non-localité, mais il la permet.

La règle d'interrogation indirecte  $\mathbf{R}^{\text{interrog}}$ , quant à elle, ne consomme rien et ne produit rien ! Elle se limite, et c'est là sa spécificité, à indiquer que le dépendant de la relation SyntP II issue du verbe régissant l'interrogative indirecte n'est pas la lexie correspondant au sémantème dépendant de la relation Sém 2, c'est-à-dire que ce n'est pas le pronom interrogatif, mais le verbe principal  $L(\langle Q_1 \rangle)$  de la proposition interrogative (= la lexie correspondant au nœud d'entrée  $\langle Q_1 \rangle$  du sous-réseau propositionnel de l'interrogative). La règle  $\mathbf{R}^{\text{interrog}}$  est donc, à la différence de  $\mathbf{R}^{\text{rel}}$ , fondamentalement discontinue. C'est la discontinuité qui ouvre la porte à la non-localité.

Bien que la relativisation et l'interrogation soient de nature linguistique très différente, elles ont en commun la propriété d'induire la non-localité.

Cette non-localité apparaît comme relativement contrainte, et les contraintes correspondantes — c'est-à-dire les contraintes définissant la chaîne verbo-nominale — ne sont pas, comme on l'a vu, directement liées à la relativisation ou l'interrogation.

Du point de vue linguistique, la non-localité des règles d'extraction provient du caractère a priori non borné des nucléus, cf. le cas du nucléus verbal dans *Il marche* ~ *Il a l'air de marcher* ~ *Il commence à avoir l'air de marcher* ... A son tour, la non-localité donne lieu à une autre caractéristique des extractions : la récursivité. Supposons qu'on vient d'appliquer une règle d'arborisation non locale **R** ; elle envisage une chaîne "fantôme" de nœuds et de relations SyntP qui ne sont pas spécifiés par **R**. Lors du calcul de ces éléments par d'autres règles, la SSém de départ peut fort bien déclencher à nouveau une règle non locale et ainsi enchâsser une interrogative indirecte dans une autre subordonnée "extraite" (relative ou interrogative indirecte) ; c'est le cas de notre phrase de référence (3). (On ne peut enchâsser une relative car celle-ci contient une relation SyntP ATTR entre un nom et un verbe — l'antécédent et le verbe principal de la relative —, ce qui est exclu par la définition de la chaîne verbo-nominale.) Bien que la récursivité de cet enchâssement soit théoriquement illimitée, un seul niveau d'enchâssement est en réalité admis. Un double enchâssement donne des phrases que les locuteurs ne peuvent pas produire et qui sont absolument incompréhensibles dans des conditions d'élocution normales ; cf. (23) :

(23) \**Je crois connaître le problème dont Pierre sait à qui Jean se demande quand parler.*

où l'on observe un double enchâssement (*Jean se demande quand parler* [*de ce problème à quelqu'un*]  $\Rightarrow$  *Pierre sait à qui Jean se demande quand parler* [*de ce problème*]  $\Rightarrow$  *le problème dont Pierre sait à qui Jean se demande quand parler*).

### 3. PROCÉDURE DE SYNTHÈSE

Les règles sémantiques de correspondance que nous avons présentées dans la Section 2 sont délibérément écrites comme bidirectionnelles, bonnes autant pour la synthèse que pour l'analyse. Cependant, la procédure qui doit les mettre en jeu n'est pas — de façon évidente — la même pour la synthèse et pour l'analyse : en synthèse, la procédure doit lire un réseau et produire un arbre ; en analyse, elle doit lire un arbre et produire un réseau. Dans ce qui suit, nous ne considérons que la PROCÉDURE DE SYNTHÈSE.

Nous présenterons d'abord les règles procédurales nécessaires à la définition du passage SSém  $\Rightarrow$  SSyntP — les métarègles de pointage — (§3.1), puis nous illustrerons ce passage par un exemple (§3.2).

#### 3.1. Les métarègles de pointage

Les règles sémantiques de correspondance sont insuffisantes pour assurer à elles seules la production d'une SSyntP à partir d'une RSém. En effet, les règles de lexicalisation et d'arborisation ne font que fournir les pièces — les lexies et les relations SyntP — qui doivent constituer une SSyntP. Il nous

manque les instructions de montage de ces pièces, c'est-à-dire les indications du type "quoi rattacher à quoi et où exactement". Comme nous l'avons ébauché au §2.1, de telles indications sont fournies par le biais des pointeurs : plus précisément, pour chaque type de règle sémantique de correspondance, nous devons décrire le mouvement des pointeurs concernés. Ainsi à chaque type de règle de correspondance associons-nous ce que nous appelons une **métarègle de pointage spécifique** ; nous donnerons ci-dessous cinq exemples de métarègles de ce type. Il est également nécessaire de disposer de trois **métarègles de pointage générales** — qui ne sont pas associées à des règles de correspondance particulière. Les métarègles de pointage font partie des règles procédurales qui assurent le passage  $RSém \Rightarrow SSyntP$ .

Avant de présenter les métarègles de pointage, nous allons caractériser la procédure que nous avons choisie (qui n'est pas, comme nous l'avons déjà dit, la seule possible ; cf. Candito & Kahane 1998).

La procédure envisagée construit l'arbre de dépendance SyntP du sommet vers les feuilles de façon continue, tout en admettant un parcours discontinu de la SSém. Autrement dit, l'ordre de synthèse des nœuds SyntP est exactement l'ordre hiérarchique induit par la structure arborescente de la SSyntP : un nœud doit être synthétisé avant ses fils. Par contre, il n'y a pas d'ordre imposé entre les différents fils d'un même nœud (= entre frères), ni, plus généralement, entre les descendants de deux lignées différentes (= entre cousins).

Le fait que l'ordre de synthèse des nœuds frères de la SSyntP ne soit pas contraint constitue un avantage considérable. Telle que nous l'avons présentée jusqu'à maintenant, on peut penser que nous envisageons la synthèse d'une phrase niveau par niveau, c'est-à-dire en réalisant complètement le passage du niveau Sém au niveau SyntP, puis du niveau SyntP au niveau SyntS et ainsi de suite. Tel n'est pas le cas. Nous pensons, au contraire, que les différents modules interagissent constamment et que la synthèse d'une phrase suit en grande partie l'ordre d'émission des mots de la phrase (et ceci d'autant plus qu'on s'approche des niveaux de surface). Si l'on s'en tient au passage  $RSém \Rightarrow SSyntP$ , cela signifie qu'on doit assurer, autant que faire se peut, la synthèse des nœuds de la SSyntP dans l'ordre où les mots correspondants seront émis à la surface. Autrement dit, l'arbre SyntP doit être produit, autant que la langue le permet, syntagme par syntagme, et les syntagmes doivent être produits dans l'ordre où ils sont émis. La procédure  $RSém \Rightarrow SSyntP$  n'est évidemment pas concernée par l'ordonnement des syntagmes — celui-ci est assuré par le module syntaxique de surface, qui gère, entre autres, la linéarisation —, mais elle doit éviter d'imposer, dans la production des nœuds SyntP, un ordre qui serait contraire à l'ordre d'émission. La procédure  $RSém \Rightarrow SSyntP$  proposée remplit le cahier des charges que nous venons de fixer, en contraignant au minimum l'ordre de production des nœuds de l'arbre SyntP.

Nous allons maintenant énoncer les métarègles de pointage. Rappelons qu'une règle de correspondance **R** est toujours déclenchée par la présence d'un pointeur en attente et qu'un pointeur est composé de trois éléments : une tête de lecture braquée sur un élément de la SSém (nœud ou arc), une tête d'écriture

prête à tracer un élément de la SSyntP (nœud ou branche) et un bloc-note contenant des conditions que la règle **R** devra vérifier pour être déclenchée (Définition 1, §2.1.1).

## Métarègles générales

### 1. Métarègle d'initialisation

La synthèse débute. Des règles linguistiques spéciales, que nous n'avons pas présentées (cf. §1.1.2), déterminent le nœud d'entrée de la RSém (plusieurs choix sont a priori possibles) et imposent des propriétés syntaxiques au sommet de l'arbre SyntP (en français, par exemple, ce sommet doit, dans le cas standard, être un verbe fini). Dans cet état de choses, la Métarègle 1 peut fonctionner :

- || Un premier pointeur doit être créé :
- || - sa tête de lecture doit être braquée sur le nœud d'entrée choisi pour la SSém ;
- || - sa tête d'écriture doit être prête à tracer le sommet de l'arbre SyntP ;
- || - son bloc-note doit contenir les propriétés syntaxiques que doit satisfaire le nœud SyntP à tracer.

La Métarègle 1 n'est utilisée qu'une seule fois — au début de la synthèse d'une phrase.

### 2. Métarègle d'appariement (d'une règle sagittale avec la règle nodale qui suit)

- || Lorsqu'une règle sagittale **R** vient d'être effectuée en consommant un arc Sém  $a$  et en déplaçant la tête de lecture d'un pointeur sur l'une des extrémités  $x$  de  $a$ , une règle nodale concernant  $x$  doit être déclenchée immédiatement.

Autrement dit, une règle nodale ne reste jamais en attente (ce qui est par contre permis pour les règles sagittales). Une telle condition sera assurée par la partie de la procédure qui gère les pointeurs en attente (et dont nous ne souhaitons pas parler davantage ici).

La Métarègle 2 permet d'éviter qu'une branche SyntP reste sans nœud dépendant : grâce à cette métarègle, dès qu'une branche SyntP est créée, un nœud SyntP est créé à la suite. Elle permet également d'éviter que deux pointeurs se trouvent simultanément braqués sur un nœud, puisqu'un pointeur ne peut jamais être laissé en attente sur un nœud : dès qu'un pointeur est braqué sur un nœud Sém, une règle nodale est déclenchée conformément à la Métarègle 2 et ce pointeur est déplacé.

A l'inverse, rien n'empêche que deux pointeurs se trouvent simultanément braqués sur un même arc Sém. La Métarègle 3 gère cette situation.

### 3. Métarègle de surpointage

- || Des deux pointeurs braqués sur un même arc Sém  $a$ , celui qui présuppose la lecture NÉGATIVE de  $a$  doit être supprimé.

Attention : la SUPPRESSION d'un pointeur par la Métarègle 3 de surpointage est une opération en soi, différente de la DISPARITION automatique d'un pointeur

après l'application d'une règle sémantique qui n'est pas suivie d'une autre règle sémantique (c'est-à-dire, lorsque le pointeur arrive dans un cul-de-sac).

Lors du passage  $RSém \Rightarrow SSyntP$ , la Métrarègle 3 est utilisée autant de fois que la  $SSém$  de départ possède de cycles. En effet, pour chaque cycle de la  $SSém$ , l'un des nœuds du cycle, le nœud de coupure effectif, est consommé deux fois en étant atteint par deux arcs différents appartenant au cycle (§2.1). L'un de ces deux arcs est nécessairement pointé deux fois (il s'agit de celui qui sera consommé en deuxième), et la Métrarègle 3 est donc nécessaire. À l'inverse, il s'agit de la seule situation où un arc peut être pointé deux fois simultanément. Dans le passage de la  $RSém$  à la  $SSyntP$  de notre phrase de référence (3), la Métrarègle 3 est utilisée deux fois (cf. Pas 14 et 19, §3.2), puisque la  $SSém$  de (3) possède deux cycles ( $\leftarrow$ -Pierre-savoir-quel-page-arracher-Pierre- $\rightarrow$  et  $\leftarrow$ -arracher-page-livre-arracher- $\rightarrow$ ).

On notera que lorsque deux pointeurs sont en concurrence, c'est celui qui présuppose la lecture négative de l'arc  $Sém$  pointé qui doit être supprimé. Ceci permet de rendre compte du fait qu'au plus un des arcs sortant d'un nœud  $Sém$  donné (c'est-à-dire allant de ce nœud à l'un de ses actants  $Sém$ ) peut être parcouru négativement. En effet, d'après la Métrarègle 4, lorsqu'un nœud  $Sém$   $x$  est consommé pour la première fois, tous les arcs sortants (et entrants) de  $x$  n'ayant pas été consommé sont pointés. En raison de la Métrarègle 3, un seul arc adjacent à  $x$  a déjà pu être consommé<sup>24</sup>, à savoir l'arc par lequel  $x$  vient d'être atteint. De par la Métrarègle 3, aucun des pointeurs braqués sur un arc sortant de  $x$  ne pourra être supprimé et, par conséquent, aucun arc sortant de  $x$  (et présupposant par conséquent une lecture positive) autre que l'arc par lequel  $x$  a été atteint la première fois ne pourra être parcouru négativement.

Les Métrarègles 2 et 3 sont complémentaires : la première permet d'éviter le double pointage des nœuds, la deuxième résout le conflit de double pointage des arcs.

### Métrarègles spécifiques : métrarègles pour les règles locales

#### 4. Métrarègle de pointage pour une règle nodale standard

Les règles nodales standard (qui font correspondre à un nœud  $Sém$  un nœud  $SyntP$ ) sont les règles de lexicalisation (Fig. 14, §2.2).

Une règle nodale standard  $\mathbf{R}$  est nécessairement déclenchée après une règle précédente sagittale  $\mathbf{R}^<$ , qui a créé une branche  $SyntP$   $\tilde{a}$  et laissé la tête de lecture d'un pointeur braquée sur un nœud sémantique  $x$  non encore consommé ; la tête d'écriture de ce pointeur est prête à tracer un nœud  $\tilde{x}$  à l'extrémité dépendante de la branche  $SyntP$   $\tilde{a}$ . La règle  $\mathbf{R}$  lit le nœud  $Sém$   $x$  et lui fait

---

24. Dans le cas standard, lorsqu'un nœud  $Sém$   $x$  est atteint pour la première fois et consommé, seul l'arc par lequel  $x$  a été atteint a été consommé (ceci est assuré par la Métrarègle 2). Il existe néanmoins un cas (le seul à notre connaissance) où un nœud  $Sém$  n'est consommé que lorsqu'il est atteint pour la deuxième fois : il s'agit des sémantèmes de pronoms interrogatifs. La première fois où un sémantème interrogatif  $\langle s \rangle$  est atteint, la règle  $\mathbf{R}^{interrog}$  est déclenchée, mais  $\langle s \rangle$  n'est pas consommé (Fig. 25) ; il n'est consommé qu'au cours de la synthèse de la proposition interrogative. Dans ce cas, lorsque le nœud  $\langle s \rangle$  est consommé, deux des arcs adjacents ont déjà été consommés (cf. Pas 23, §3.1).

correspondre le nœud SyntP  $\tilde{x}$ . Ici la démultiplication des pointeurs est de mise.

Quand la règle nodale **R** a été déclenchée et que le nœud SyntP  $\tilde{x}$  a été produit, le pointeur braqué sur  $x$  doit être démultiplié en  $n$  pointeurs, où  $n$  est le nombre d'arcs sémantiques  $a_i$  (entrants et sortants) adjacents à  $x$  qui n'ont pas encore été consommés (ce qui exclut l'arc  $a$ , par lequel le pointeur est arrivé sur  $x$ ). La tête de lecture de chacun des pointeurs issus de cette démultiplication est braquée sur un des arcs sémantiques  $a_i$  ; la tête d'écriture correspondante est prête à tracer une branche SyntP  $\tilde{a}_i$ .

Autrement dit, il y aura  $n$  règles sémantiques sagittales déclenchées à la suite de **R**.

L'application de la règle **R** peut être assujettie à certaines conditions, vérifiées par le biais du bloc-note du pointeur déclenchant la règle. Attention : ceci est vrai pour n'importe quelle métarègle spécifique. Inversement, lorsqu'une métarègle crée un nouveau pointeur, elle place dans son bloc-note toutes les contraintes et conditions linguistiques attachées à la règle sémantique que cette métarègle manipule (ces contraintes seront utilisées par les métarègles intervenant par la suite). Voir l'exemple des règles d'extraction (Pas 7 et 12, §3.2).

La Métarègle 4 est déclenchée lorsqu'un pointeur est braqué sur un nœud qui n'a pas encore été consommé. Or il est possible qu'un pointeur soit braqué sur un nœud qui a déjà été consommé. Dans ce cas, on déclenche une métarègle particulière — de nature essentiellement procédurale : la Métarègle 5.

### 5. Métarègle de surconsommation (expression de la coréférence)

Si un pointeur est braqué sur un nœud Sém  $x$  d'étiquette «s» QUI A DÉJÀ ÉTÉ CONSOMMÉ en produisant un nœud SyntP  $\tilde{x}$  d'étiquette L(«s»), il faut produire une copie  $\tilde{x}'$  du nœud SyntP  $\tilde{x}$  avec la même étiquette L(«s») et suspendre cette copie à l'extrémité dépendante de la branche SyntP sur laquelle se trouve braquée la tête d'écriture de ce pointeur ; de plus, le pointeur doit établir entre la copie  $\tilde{x}'$  et sa source  $\tilde{x}$  un lien de coréférence (marqué dans nos figures par des flèches bidirectionnelles en pointillé) ; ensuite, le pointeur doit disparaître.

Autrement dit, lorsqu'un nœud Sém est atteint pour la deuxième fois, il n'est pas nécessaire de déclencher à nouveau une règle de lexicalisation pour  $x$ <sup>25</sup>.

### 6. Métarègle de pointage pour une règle sagittale standard

Comme exemples de règles sagittales standard (qui font correspondre à un arc Sém une branche), nous pouvons indiquer la règle sagittale positive  $\mathbf{R}_{+\geq 2}$  ou la règle sagittale négative  $\mathbf{R}_{-1(\text{Adj})}$  (Fig. 20 et 18, §2.3.1).

25. Cette règle prévoit l'expression de la coréférence par la pronominalisation "pure" ; pour l'utilisation de quasi-synonymes de toutes sortes (comme dans *Après avoir épluché les carottes et les navets, placez les légumes dans l'eau bouillante*), une autre métarègle de surconsommation est nécessaire.



Une règle sagittale standard  $\mathbf{R}$  est nécessairement déclenchée après une règle précédente nodale  $\mathbf{R}^<$ , qui a consommé un nœud Sém  $x$ , créé un nœud SyntP  $\tilde{x}$  et laissé la tête de lecture d'un pointeur braquée sur un arc sémantique  $a$  adjacent à  $x$ ; la tête d'écriture de ce pointeur est prête à tracer une branche SyntP sortant de  $\tilde{x}$ . La règle  $\mathbf{R}$  lit l'arc sémantique  $a$  et lui fait correspondre une branche SyntP  $\tilde{a}$  sortant du nœud  $\tilde{x}$ . La Métrarègle 6 peut alors fonctionner :

|| Quand la branche SyntP  $\tilde{a}$  a été tracée, la tête de lecture doit être déplacée et braquée sur l'autre extrémité de l'arc sémantique  $a$  et la tête d'écriture doit être déplacée et braquée sur l'autre extrémité de la branche SyntP  $\tilde{a}$  (= la position qu'occupera le nœud dépendant de  $\tilde{a}$ ).

Autrement dit, il y aura d'une règle sémantique nodale déclenchée à la suite de  $\mathbf{R}$ ; en raison de la Métrarègle générale 2, cette règle nodale sera déclenchée immédiatement.

### Métrarègles spécifiques : métrarègles pour les règles non locales

#### 7. Métrarègle de pointage pour la règle (sagittale) de relativisation

La règle de relativisation  $\mathbf{R}$  (=  $\mathbf{R}^{\text{rel}}$ ) est déclenchée après une règle précédente nodale  $\mathbf{R}^<$ , qui a créé un nœud SyntP  $\tilde{x}$  (étiqueté par un nom, qui sera l'antécédent de la relative à venir) et laissé la tête de lecture d'un pointeur braquée sur un arc sémantique  $a$  ENTRANT en  $x$ ; la tête d'écriture de ce pointeur est prête à tracer une branche SyntP sortant de  $\tilde{x}$ .

La règle  $\mathbf{R}$  lit (négativement) l'arc sémantique  $a$  et trace une branche SyntP ATTR (ou ATTR<sub>QUAL</sub>)  $\tilde{b}$  sortant de  $\tilde{x}$ .

Avant de formuler la Métrarègle 7, précisons la situation dans laquelle la règle  $\mathbf{R}$  est déclenchée. Pour qu'une portion du réseau Sém puisse donner naissance à une proposition relative, il faut que cette portion soit un sous-réseau propositionnel  $S$  de degré  $d + 1$ , où  $d$  est le degré courant (voir définitions aux §§ 1.1.2 et 2.3.2). Le degré courant est toujours indiqué dans le bloc-note du pointeur (il est égal à 1 au départ de la synthèse). Rappelons que, pour un sous-réseau comme  $S$ , on calcule un nœud d'entrée, de la même façon qu'on calcule le nœud d'entrée de la SSém d'une phrase (§1.1.2).

|| Quand la branche SyntP  $\tilde{b}$  a été tracée, la tête de lecture, SANS CONSOMMER l'arc  $a$ , doit être déplacée pour être braquée sur le nœud d'entrée  $y$  du sous-réseau  $S$  (le nœud  $y$  a l'étiquette ' $\mathbf{Q}_1$ ' dans la Fig. 24). La tête d'écriture doit être braquée sur l'extrémité dépendante de la branche  $\tilde{b}$ . Quant au bloc-note, il subit deux manipulations :

- après vérification du fait que le degré de  $S$  est égal à  $d + 1$ , où  $d$  est le degré courant, le degré courant est incrémenté de 1 (et prend donc la valeur  $d + 1$ ) ;

- y sont inscrites des contraintes assurant que, dans la SSyntP produite, 1) il y aura une chaîne de dépendance SyntP allant du sommet de la proposition  $\tilde{y}$  (d'étiquette L('Q<sub>1</sub>')) à un nœud correspondant à  $x$  (qui sera le pronom relatif en SSyntS) et 2) cette chaîne sera une chaîne verbo-nominale (grâce à ces contraintes, les règles déclenchées à la suite de  $\mathbf{R}$  permettront de revenir au nœud  $x$  en produisant une chaîne verbo-nominale).

Le pointeur créé par la Métrarègle 7 est maintenant prêt au déclenchement d'une règle nodale (qui produira le sommet SyntP de la relative).

Comme on le voit, la règle de relativisation  $\mathbf{R}^{\text{rel}}$  ne consomme aucun élément de la SSém — ni arc, ni nœud. C'est quand même une règle sagittale, puisque le pointeur responsable de son déclenchement a sa tête de lecture braquée sur un arc  $a$  et doit donc être mentionné dans les conditions initiales de déclenchement. Mais, à la différence des autres règles sagittales,  $\mathbf{R}^{\text{rel}}$  ne consomme pas l'arc Sém  $a$  sur lequel est braquée la tête de lecture de son pointeur. L'arc  $a$  est pour ainsi dire dédoublé. En effet, il est parcouru (et donc pointé) deux fois : la première fois, lors de la relativisation proprement dite,  $a$  donne une branche SyntP ATTR, par laquelle la relative sera attachée à son antécédent ; la deuxième fois, lors de la synthèse de la relative,  $a$  donne une branche SyntP par laquelle le pronom relatif sera attaché à son gouverneur, et c'est à ce moment-là que  $a$  est consommé.

### 8. Métrarègle de pointage pour la règle (nodale) d'interrogation indirecte

La règle d'interrogation indirecte  $\mathbf{R}$  (=  $\mathbf{R}^{\text{interrog}}$ ) est déclenchée après une règle précédente sagittale  $\mathbf{R}^<$ , qui a consommé un arc sémantique  $a$  d'étiquette 2 sortant d'un nœud prédicatif “informationnel” (comme ‘se demander’, ‘savoir’, ‘dire’ ...), créé une branche SyntP  $\tilde{a}$  d'étiquette II, et laissé la tête de lecture d'un pointeur braquée sur un nœud sémantique  $x$  étiqueté par un sémantème interrogatif (= incluant le sémantème ‘quel’), la source du futur pronom interrogatif ; la tête d'écriture de ce pointeur est prête à tracer un nœud SyntP à l'extrémité dépendante de la branche SyntP  $\tilde{a}$ .

La règle  $\mathbf{R}$  lit le nœud sémantique  $x$ , mais ne le consomme pas et ne trace rien.

Comme pour la règle de relativisation  $\mathbf{R}^{\text{rel}}$ , la règle d'interrogation indirecte  $\mathbf{R}^{\text{interrog}}$  nécessite une situation initiale particulière : le nœud  $x$  doit être surmonté d'un sous-réseau propositionnel  $S$  de degré  $d + 1$ , où  $d$  est le degré courant.

Quand la règle  $\mathbf{R}$  est déclenchée, la tête de lecture, SANS CONSOMMER le nœud Sém  $x$ , doit vérifier que ce nœud est étiqueté par un sémantème interrogatif, puis être déplacée pour être braquée sur le nœud dominant  $y$  du sous-réseau  $S$  (d'étiquette ‘ $Q_1$ ’ dans la Fig. 25). La tête d'écriture ne bouge pas et reste donc braquée sur l'extrémité dépendante de la branche SyntP  $\tilde{a}$ . Quant au bloc-note, il subit les deux mêmes manipulations que dans la Métrarègle 7 (pour la relativisation).

Le pointeur créé par la Métrarègle 8 est ainsi prêt au déclenchement d'une règle nodale (qui produira le sommet SyntP de l'interrogative indirecte).

La règle  $\mathbf{R}^{\text{interrog}}$  ne consomme pas le nœud Sém  $x$  sur lequel est braquée la tête de lecture de son pointeur ; elle a pour seul effet de déplacer la tête de lecture sur un autre nœud — à savoir, le nœud communicativement dominant du sous-réseau surmontant  $x$ .  $\mathbf{R}^{\text{interrog}}$  est néanmoins une règle nodale, puisqu'elle est déclenchée par un pointeur braqué sur un nœud (étiqueté par un sémantème interrogatif). Soulignons l'importance du fait que  $\mathbf{R}^{\text{interrog}}$  est déclenchée lors de la première visite d'un pointeur au nœud  $x$  : à la deuxième visite, c'est-à-dire

au retour d'un pointeur au cours de la synthèse de la proposition interrogative, le nœud  $x$  doit être consommé — par la règle de lexicalisation du pronom.

Les règles de relativisation  $\mathbf{R}^{\text{rel}}$  et d'interrogation  $\mathbf{R}^{\text{interrog}}$  diffèrent donc en un point notable : la première est sagittale, la deuxième nodale. Cependant, les deux ont une particularité importante en commun : elles ne consomment rien, mais préparent le terrain pour les “vraies” règles sémantiques de correspondance.

### 3.2. Le passage pas à pas d'une RSém à une SSyntP

Nous pensons utile de décrire la construction pas à pas de la SSyntP de notre phrase de référence (3) (Fig. 11) à partir de sa RSém (Fig. 8). Cette tâche est effectuée en 25 pas.

Rappelons que notre procédure produit un arbre SyntP dans l'ordre hiérarchique (= du sommet vers les feuilles), mais que, par contre, aucun ordre n'est imposé dans la production de nœuds frères. Dans cette illustration, les pas sont ordonnées linéairement selon deux principes :

1) Nous adoptons une synthèse SYNTAGME PAR SYNTAGME, c'est-à-dire que nous terminons la synthèse d'un syntagme (= la réalisation d'un sous-arbre SyntP) avant d'attaquer le suivant. Cela signifie que :

|| dès qu'un pointeur est choisi, il nous faut le suivre jusqu'au bout, c'est-à-dire jusqu'à ce que tous les pointeurs qui en sont issus (par démultiplication) aient disparu.

(Notons que ce principe subsume la Métrarègle générale 2.)

2) Lorsque plusieurs pointeurs sont en attente, nous commençons par produire le syntagme qui précède les autres dans l'ordre linéaire (= l'ordre d'émission) de la phrase. Dans un processus de synthèse automatisé, ce principe ne peut être observé qu'avec le concours du module syntaxique de surface, qui assure, entre autres, la linéarisation.

**Pas 1.** (Initialisation) Le seul nœud pouvant être choisi comme nœud d'entrée de notre SSém est le nœud dominant du rhème primaire  $\mathbf{R}_1$ , à savoir ‘prendre’. Conformément à la Métrarègle 1 (d'initialisation), un premier pointeur est créé dont la tête de lecture est braquée sur ‘prendre’. La tête d'écriture est prête à tracer le sommet de l'arbre SyntP. Il est inscrit dans le bloc-note du pointeur que ce sera un verbe fini (pour le français). Le degré courant est 1.

**Pas 2.** (PRENDRE) Le pointeur braqué sur le nœud Sém ‘prendre’ cherche dans le lexique une règle nodale de lexicalisation pour ‘prendre’ et la déclenche ; le nœud ‘prendre’ est ainsi consommé, alors que le lexème PRENDRE, fourni par la règle, est écrit dans la position initiale pour la construction d'une SSyntP (après qu'il ait été vérifié qu'il remplissait les conditions inscrites dans le bloc-note, à savoir être un verbe)<sup>26</sup>. Le caractère hiérarchique de la procédure

26. Dans le bloc-note, il est également précisé que ce verbe est fini, ce qui force la procédure à rechercher dans la SSém une configuration de sémantèmes déterminant le mode et le temps du verbe. Par la suite, nous ne traiterons pas des règles de ce type — les règles flexionnelles.

assure que ce nœud SyntP sera le sommet de l'arbre SyntP. Le nœud Sém 'prendre' ayant deux arcs adjacents, le pointeur de départ se démultiplie en deux pointeurs, chacun braqué sur l'un de ces arcs.

Comme nous l'avons dit, l'ordre dans lequel les pointeurs seront manipulés n'est pas pertinent et suit ici l'ordre linéaire des syntagmes.

**Pas 3.** (branche I) Le pointeur braqué sur l'arc sémantique 1 sortant de 'prendre' déclenche la règle sagittale  $\mathbf{R}_{+1}$ , qui crée la branche SyntP I rattachée au nœud PRENDRE ; l'arc 1 est ainsi consommé. Le pointeur est alors déplacé : la tête de lecture est braquée sur le nœud 'moi' et la tête d'écriture est placée à l'extrémité dépendante de la branche I nouvellement créée.

**Pas 4.** (MOI) Le pointeur braqué sur 'moi' déclenche la règle nodale de lexicalisation correspondante ; 'moi' est consommé et le lexème MOI est fourni et placé sur le nœud dépendant de la branche SyntP I créée au pas précédent ; après quoi, le pointeur disparaît, car il n'y a aucun arc à lire à partir du nœud 'moi'. A la fin de ce pas, nous disposons déjà de la portion de l'arbre SyntP correspondant à *je prends* : MOI  $\leftarrow_I$  PRENDRE.

**Pas 5.** (branche II) Le pointeur braqué sur l'arc sémantique 2 sortant de 'prendre' déclenche la règle sagittale  $\mathbf{R}_{+\geq 2}$ , qui consomme cet arc, crée la branche SyntP II sortant du nœud PRENDRE et braque le pointeur sur 'livre'. Ce pas est similaire au Pas 3.

**Pas 6.** (LIVRE) Le pointeur braqué sur 'livre' déclenche la règle nodale de lexicalisation correspondante et tout se passe comme pour le cas de 'prendre' (Pas 2). Le pointeur braqué sur 'livre' se démultiplie en deux pointeurs braqués sur les deux arcs entrants non consommés.

**Pas 7.** ( $\mathbf{R}^{\text{rel}}$ ) Considérons le pointeur issu de cette démultiplication braqué sur l'arc d'étiquette 2 allant de 'page' à 'livre' (rappelons encore que l'ordre dans lequel sont manipulés les deux pointeurs en attente n'est pas pertinent). Comme ce pointeur vient du nœud 'livre', il doit parcourir cet arc dans le sens négatif. La règle de relativisation  $\mathbf{R}^{\text{rel}}$  peut donc être déclenchée, d'autant que LIVRE est un nom (la condition " $L(\langle X \rangle) = N$ " imposée dans la règle  $\mathbf{R}^{\text{rel}}$  au sommet  $L(\langle X \rangle)$  du sous-arbre SyntP en train d'être construit est satisfaite). La Métarègle 7 associée à  $\mathbf{R}^{\text{rel}}$  fait les choses suivantes :

1) elle déclenche le calcul d'un nœud d'entrée pour le sous-réseau propositionnel de degré 2 qui surmonte la configuration  $-2 \rightarrow$  'livre', puis déplace le pointeur sur le nœud d'entrée choisi (le seul choix possible ici est le nœud dominant du thème secondaire  $\mathbf{T}_2$  — 'savoir') ;

2) elle trace une branche SyntP ATTR à partir de LIVRE et braque la tête d'écriture à l'extrémité dépendante de cette branche ;

3) elle inscrit dans le bloc-note du pointeur en attente des contraintes sur la chaîne verbo-nominale qui seront vérifiées au fur et à mesure de la construction de la relative ;

4) elle incrémente le degré courant (cf. Pas 1) de 1 : le degré courant est maintenant égal à 2.

**Pas 8.** (SAVOIR) Le pointeur braqué sur ‘savoir’ déclenche une règle nodale de lexicalisation. Tout se passe comme dans le Pas 2, à la différence suivante près : il y a maintenant les contraintes concernant la chaîne verbo-nominale à vérifier. Ces contraintes sont satisfaites par SAVOIR, qui est bien un verbe et est ainsi admis comme le premier élément du nucléus verbal exigé par les contraintes dans le bloc-note du pointeur. Le pointeur se démultiplie en deux pointeurs — pour les deux actants sémantiques de ‘savoir’.

**Pas 9.** (branche I) Le pointeur braqué sur l’arc 1 sortant de ‘savoir’ produit une branche SyntP I rattachée à SAVOIR (voir Pas 3), puis est braqué sur ‘Pierre’.

**Pas 10.** (PIERRE) Le nœud ‘Pierre’ donne naissance à PIERRE (voir Pas 4), et un pointeur est braqué sur l’autre arc 1 entrant en ‘Pierre’.

Après le Pas 10, on a une partie de la SSyntP qui correspond au début de la phrase *Je prends le livre*, plus un syntagme inachevé [*dont*] *Pierre sait [...]*. Jusqu’ici, la synthèse s’était déroulée sans complications. A ce point, nous devons faire face à un problème procédural : le pointeur braqué sur l’arc 1 entrant en ‘Pierre’ devrait parcourir cet arc négativement et de ce fait produire une construction modificative — une relative ou une tournure participiale — modifiant PIERRE. Le résultat en serait quelque chose comme :

(24) *Je prends le livre dont Pierre, qui en a arraché une page (en ayant arraché une page), sait laquelle.*

Cependant, (24) n’a pas la même SComm-Sém que (3), que nous voulons synthétiser ! En effet, pour (24), le sous-réseau exprimé par la construction modificative aurait dû se trouver dans le thème secondaire  $\mathbf{T}_2$ , auquel appartient l’élément modifié ‘Pierre’. Mais, comme ce n’est pas le cas dans la RSém que nous considérons (Fig. 8), le déclenchement d’une règle produisant une telle construction modificative est ici impossible (la SComm-Sém de la structure ne vérifiera pas les conditions d’application de la règle de correspondance en question). En conséquence, ce pointeur doit être mis en attente en sorte que l’arc en question (‘Pierre’  $\leftarrow_1$  ‘arracher’) n’est pas pour l’instant consommé. Il le sera plus tard — par un pointeur le parcourant dans le sens positif —, alors que le pointeur en attente sera supprimé par la Métrarègle 3 (de surpointage) : Pas 14 et 15. (Dans un vrai processus de synthèse, la mise en attente des pointeurs doit être gérée par des métrarègles adaptées ; en particulier, un pointeur qui présuppose une lecture positive d’un arc ne doit pas être mis en attente, car il ne pourra pas être supprimé par la suite ; cf. les conditions de la Métrarègle 3 de surpointage.)

**Pas 11.** (branche II) Le pointeur braqué sur l’arc 2 sortant de ‘savoir’ (c’est le second pointeur issu de la démultiplication du pointeur ayant consommé ‘savoir’ : Pas 8) produit une branche SyntP II rattachée à SAVOIR, après quoi il est braqué sur ‘quel’. Son bloc-note contient toujours les contraintes sur la chaîne verbo-nominale de la relative qui est en train d’être synthétisée.

**Pas 12.** ( $R^{\text{interrog}}$ ) Le pointeur braqué sur ‘quel’ déclenche la règle d’interrogation indirecte  $R^{\text{interrog}}$ . La Métrarègle 8, associée à la règle  $R^{\text{interrog}}$ , déclenche le calcul du nœud d’entrée du sous-réseau propositionnel de degré 3 surmontant le nœud ‘quel’ et l’arc qui en sort. Le nœud dominant du rhème tertiaire  $R_3$ , à savoir ‘arracher’, est choisi comme nœud d’entrée et la tête de lecture est braquée dessus. La tête d’écriture reste en place. De nouvelles contraintes — concernant la chaîne verbo-nominale de l’interrogative — viennent se superposer aux contraintes concernant la chaîne verbo-nominale de la relative, déjà inscrites dans le bloc-note de ce pointeur. Le degré courant est incrémenté de 1 et vaut donc 3.

**Pas 13.** (ARRACHER) Le pointeur braqué sur ‘arracher’ suite au pas précédent fait trois choses :

- il produit (en déclenchant la règle de lexicalisation correspondante) la lexie ARRACHER, rattachée à la branche SyntP II sortant de SAVOIR ;
- il vérifie, après consultation de son bloc-note, que ARRACHER est un élément légitime des deux nucléus verbaux emboîtés, l’un pour la relative, l’autre pour l’interrogative ;
- il se démultiplie en trois pointeurs braqués sur les arcs adjacents au nœud ‘arracher’ ; nous allons les considérer à tour de rôle.

**Pas 14 et 15.** (Surpointage ; branche I) Le pointeur braqué sur l’arc 1 sortant de ‘arracher’ se trouve en présence d’un pointeur en attente braqué sur le même arc. Ce dernier pointeur (créé au Pas 10) est supprimé par application de la Métrarègle 3 (de surpointage) ; l’autre pointeur peut maintenant produire une branche SyntP I rattachée à ARRACHER ; il est ensuite braqué sur ‘Pierre’.

**Pas 16.** (Surconsommation) Le nœud ‘Pierre’ a déjà été consommé. Par conséquent, en conformité avec la Métrarègle 5, le pointeur braqué sur ce nœud produit une copie de  $L(\text{‘Pierre’})$ , qu’il suspend à la branche SyntP I sortant d’ARRACHER ; en même temps, il établit le lien de coréférence entre ces deux occurrences de  $L(\text{‘Pierre’})$ .

**Pas 17.** (branche II) Le pointeur braqué sur l’arc 2 sortant de ‘arracher’ produit une branche SyntP II rattachée à ARRACHER, puis est braqué sur ‘page’.

**Pas 18.** (PAGE) Le nœud PAGE est créé (de la même façon que le nœud LIVRE : Pas 6) et deux pointeurs sont braqués sur les arcs adjacents à ‘page’.

**Pas 19 et 20.** (Surpointage ; branche II) Le pointeur braqué sur l’arc 2 qui sort de ‘page’ se trouve en concurrence avec un autre pointeur déjà braqué sur cet arc : il s’agit du pointeur issu de la lecture du nœud ‘livre’ (Pas 6). Comme ce dernier pointeur présuppose la lecture négative de l’arc, il est supprimé par la Métrarègle 3. Ensuite, le pointeur issu de ‘page’ produit une branche SyntP II rattachée à PAGE, puis est braqué sur ‘livre’.

**Pas 21.** (Surconsommation) Le pointeur braqué sur ‘livre’ se trouve, comme au Pas 16, braqué sur un nœud déjà consommé. En conformité avec la Métrarègle 5, il produit une copie de  $L(\text{‘livre’})$  et établit un lien de coréférence avec le nœud SyntP LIVRE créé au Pas 6.

**Pas 22.** (branche ATTR) Le deuxième pointeur de ‘page’ braqué sur l’arc 1 sortant de ‘quel’ parcourt cet arc négativement en le consommant et produit donc, conformément à la règle  $R_{-1}$ , une branche SyntP ATTR. Le pointeur est ensuite braqué sur ‘quel’.

**Pas 23.** (QUEL) Le nœud QUEL est créé par lexicalisation de ‘quel’ ; le pointeur de ‘quel’ disparaît, puisque les deux arcs adjacents à ‘quel’ ont déjà été consommés (l’arc venant de ‘savoir’ au Pas 11, l’arc allant de ‘quel’ vers ‘page’ au pas précédent).

**Pas 24.** (branche III) Le dernier pointeur encore en vie est braqué sur l’arc 3 sortant d’arracher’. (Cet arc a déjà été lu par un pointeur venu de ‘livre’ qui a déclenché la règle de relativisation — voir Pas 7 —, mais n’a pas été consommé par cette règle.) Le pointeur crée une branche SyntP III et se déplace sur ‘livre’.

**Pas 25.** (Surconsommation) Comme le nœud ‘livre’ a déjà été consommé deux fois, le pointeur produit une nouvelle copie de la même lexie, la suspend à la branche SyntP III précédemment construite et indique le lien de coréférence entre elle et la précédente copie. Après quoi, le dernier pointeur disparaît, ce qui couronne notre entreprise de synthèse de la SSyntP de la phrase (3).

## CONCLUSION

Nous pouvons résumer en trois points les aspects de notre recherche qui nous semblent constituer un apport dans les domaines de la sémantique linguistique et du traitement automatique du langage : introduction de nouveaux concepts, formalisation de règles sémantiques et formulation rigoureuse d’une procédure de synthèse (des énoncés).

### • Introduction de nouveaux concepts : nucléus et pointeur

Nous croyons avoir montré que les concepts linguistiques de NUCLÉUS verbal et nominal sont nécessaires pour la sémantique et la syntaxe. Ces concepts, introduits dans Kahane 1996 et 1997, formalisent certains types de syntagmes qui jouent un rôle marqué dans la construction des phrases ; il s’agit donc de concepts syntaxiques (plus précisément, syntaxiques profonds). Les concepts de nucléus sont également cruciaux pour la sémantique : les règles de correspondance entre les SSém et les SSyntP des énoncés — règles qui constituent le noyau du module sémantique de tout modèle linguistique, quel que soit son cadre de référence — ne peuvent pas être écrites de façon élégante et générale sans avoir recours à ces concepts. Par conséquent, les contraintes sur les règles SSém  $\Leftrightarrow$  SSyntP qui utilisent de façon essentielle les nucléus contraignent également la nature des SSém réalisables par une RSyntP, c’est-à-dire les sens que peut envisager d’exprimer aisément un locuteur.

D’autre part, la description de l’aspect dynamique de la synthèse des énoncés (en d’autres mots, de la procédure de synthèse) a nécessité l’introduction du concept de POINTEUR. Ce concept algorithmique très général n’est pas nouveau en soi (cf. , par ex., Aho & Ullman 1972), mais nous en proposons un nouvel usage dans le passage d’un réseau sémantique à un arbre SyntP.

En particulier, l'introduction de ce concept nous a permis de formuler les métarègles qui gèrent l'application des règles linguistiques de correspondance et qui assurent le parcours exhaustif de réseaux. Soulignons une fois de plus que la séparation rigoureuse des règles linguistiques à proprement parler non orientées (du point de vue de l'opposition "synthèse ~ analyse") et des métarègles procédurales orientées est pour nous une propriété fondamentale de l'approche proposée (cf. la comparaison entre TAG et le modèle Sens-Texte dans Candito & Kahane 1998).

#### • Formalisation des règles sémantiques

Malgré la position centrale qu'occupe le module sémantique dans un modèle Sens-Texte, le formalisme de ses règles n'avait pas encore été bien élaboré et sa structure interne restait peu claire. Le présent article apporte quatre modifications majeures à cet état de choses :

- nous avons proposé et illustré un formalisme d'écriture pour les différents types de règles sémantiques envisageables (développant et clarifiant les propositions de Polguère 1990) ;

- en introduisant des règles Sém discontinues (non locales), nous avons considérablement changé la vision du passage SSém  $\Rightarrow$  SSyntP, et notamment du parcours d'une SSém ;

- nous avons mis au point le concept de MÉTARÈGLE de pointage et formulé toutes les métarègles de ce type nécessaires dans la synthèse sémantique des phrases à extraction ;

- nous avons élaboré une taxinomie des règles sémantiques en établissant ainsi la structure du module sémantique.

#### • Utilisation de réseaux comme structures de départ de la synthèse

L'une des caractéristiques notables de notre approche est d'envisager la synthèse à partir de RÉSEAUX. Ceci entraîne deux particularités de l'approche proposée : une attention spéciale au parcours exhaustif du réseau et la prise en charge des discontinuités (notamment non locales) entre la structure de départ et celle d'arrivée. Les autres travaux en synthèse que nous connaissons considèrent pour la plupart une représentation de départ arborescente (Danlos 1985, 1998, McKeown 1985, Hovy 1987). Dans les travaux qui envisagent comme représentation de départ un vrai réseau (Nogier 1991, où les structures de départ sont des réseaux conceptuels à la Sowa 1984), il ne semble pas que le parcours complet d'un réseau quelque peu complexe soit assuré et que les phénomènes de dépendances non bornées soient pris en compte. Notre travail (à la suite de Iordanskaja & Polguère 1988, Polguère 1990) est donc, à notre connaissance, l'une des premières études poussées du passage des réseaux sémantiques aux arbres syntaxiques et notamment de la procédure de parcours de réseaux et de déclenchement de règles linguistiques de correspondance. Elle est certainement la première étude linguistique formelle prenant en compte l'application de règles discontinues (non locales) à partir d'un réseau. La description procédurale du passage d'un réseau à un arbre proposé est suffisamment générale pour être applicable dans tout modèle où il est question du passage entre un réseau et un arbre.



Nous nous demandons, chère lectrice, cher lecteur, ce que vous pensez que vous voulez pouvoir avoir l'intention d'essayer de proposer que nous tentions d'envisager de faire à la suite de cela ?

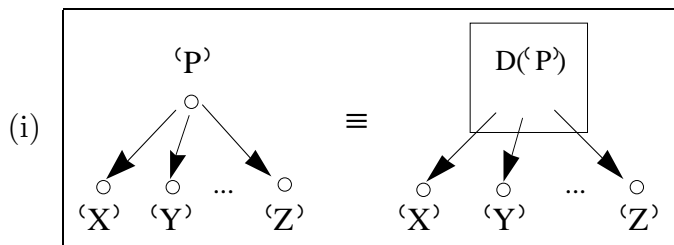
### Remerciements

Les auteurs ont pu écrire le présent article grâce au séjour d'I. Mel'čuk à Paris comme professeur au Collège de France ; ils tiennent donc à exprimer leur reconnaissance à cette institution et à son administration, qui leur a créé d'excellentes conditions de travail. I. Mel'čuk a, de plus, bénéficié, pour la préparation de cet article, de deux subventions du FCAR : 96-ER-0618 et 97-ER-2741. D'un autre côté, S. Kahane a bénéficié d'une subvention du laboratoire TALaNa pour un voyage à Montréal. Nous remercions pour leur remarques, critiques, questions et suggestions, lors de la lecture des précédentes versions du manuscrit ou lors de la présentation orale de ce travail, M.-H. Candito, L. Iordanskaja, H. Kamp, G. Lacombe, J. Milićević, N. Pertsov et A. Polguère, ainsi que les deux relecteurs de la revue. Les erreurs et les défauts qui ont survécu à leur examen ne sont imputables qu'à nous-mêmes.

### ANNEXE A : Métarègles d'équivalence sémantique

Les règles de décomposition sémantique des pronoms interrogatifs (Fig. 9) ont une particularité remarquable. Dans le cas général, lorsqu'on applique une règle de décomposition à un sémantème  $\langle P \rangle$  dans le sens de l'expansion, les arcs sémantiques entrant sur  $\langle P \rangle$  sont hérités par le nœud communicativement dominant de sa décomposition. Par contre, comme le sémantème  $\langle Q \rangle$  d'un pronom interrogatif a deux nœuds communicativement dominants dans sa décomposition, les arcs sémantiques entrants sur  $\langle Q \rangle$  doivent être distribués entre ces deux nœuds (voir Figure 10, où l'arc 2 venant de  $\langle \text{se demander} \rangle$  pointe sur  $\langle \text{quel} \rangle$ , alors que l'arc 1 venant de  $\langle \text{venir} \rangle$  pointe sur  $\langle \text{personne} \rangle$ ). Cette distribution ne peut pas bien sûr être arbitraire ; on a donc besoin de métarègles qui puissent contrôler l'application des règles de décomposition.

Ce fait mérite d'être présenté de façon formelle. Soit une règle de décomposition sémantique (i) :



où  $\langle P \rangle$  est un prédicat d'arguments  $\langle X \rangle$ ,  $\langle Y \rangle$ , ...,  $\langle Z \rangle$  et  $D(\langle P \rangle)$  est sa décomposition sémantique (c'est-à-dire un réseau non unitaire). L'application d'une telle règle de décomposition sémantique consiste à effectuer une des deux opérations suivantes :

- soit on remplace, dans un réseau Sém  $A$ , un nœud d'étiquette  $\langle P \rangle$  par sa décomposition  $D(\langle P \rangle)$  : expansion ;

- soit on remplace, dans un réseau Sém  $A'$ , un sous-réseau non unitaire qui représente la décomposition de  $\langle P \rangle$  par  $\langle P \rangle$  : réduction.

Le nœud  $\langle P \rangle$ , de même que  $D(\langle P \rangle)$ , est rattaché au réseau auquel il appartient par des arcs sortants et entrants. Les arcs sortant de  $\langle P \rangle$  (resp.  $D(\langle P \rangle)$ ) pointent vers les actants sémantiques de  $\langle P \rangle$  (resp.  $D(\langle P \rangle)$ ). Ces arcs et leurs extrémités sont explicitement indiqués dans la règle (i) ; leur gestion, lors de l'application de la règle, ne pose donc pas de difficultés. Les arcs entrant en  $\langle P \rangle$  (resp.  $D(\langle P \rangle)$ ), par contre, ne figurent pas dans la règle (i) : il est donc nécessaire d'indiquer, par une métarègle, de quelle manière ils doivent être gérés. Voilà la métarègle qui assure cette gestion dans le cas standard, c'est-à-dire le cas où  $D(\langle P \rangle)$  n'a qu'un seul nœud dominant.

### Métarègle d'équivalence sémantique 1 (cas standard)

- Condition :  $D(\langle P \rangle)$  n'a qu'un seul nœud communicativement dominant.
1. Expansion : Si un nœud  $\langle P \rangle$  est remplacé par le sous-réseau non unitaire  $D(\langle P \rangle)$ , les arcs entrant sur  $\langle P \rangle$  doivent être pointés sur le nœud communicativement dominant de  $D(\langle P \rangle)$ .
  2. Réduction : Si un sous-réseau non unitaire  $D(\langle P \rangle)$  est remplacé par le nœud  $\langle P \rangle$ , les arcs entrants sur le nœud communicativement dominant de  $D(\langle P \rangle)$  doivent être pointés sur  $\langle P \rangle$ .

Cette métarègle ne couvre pas le cas de réduction où il y a des arcs entrants en  $D(\langle P \rangle)$  qui pointent sur un nœud autre que son nœud communicativement dominant. Pour de tels cas, on a besoin de métarègles particulières, mais nous ne développerons pas ce point, qui est hors sujet.

La Métarègle d'équivalence sémantique 1 ne peut s'appliquer dans le cas des pronoms interrogatifs, dont la décomposition a toujours deux nœuds communicativement dominants ; on a besoin d'une deuxième métarègle (que nous n'énoncerons que pour le cas de l'expansion).

### Métarègle d'équivalence sémantique 2 (cas des pronoms interrogatifs)

- Condition :  $D(\langle P \rangle)$  a deux nœuds communicativement dominants de degré  $d$  et  $d + 1$ .
- Expansion : Si un nœud  $\langle P \rangle$  est remplacé par le sous-réseau non unitaire  $D(\langle P \rangle)$ , alors :
- les arcs entrant sur  $\langle P \rangle$  qui viennent du sous-réseau propositionnel de degré  $d + 1$  [ce sous-réseau constitue la proposition interrogative] doivent être pointés sur le nœud communicativement dominant de  $D(\langle P \rangle)$  de degré  $d + 1$  ;
  - l'arc 2 venant de l'extérieur du sous-réseau propositionnel de degré  $d + 1$  [c'est-à-dire du prédicat informationnel régissant l'interrogative] doit être pointé sur le nœud communicativement dominant de  $D(\langle P \rangle)$  de degré  $d$ .

L'application de cette métarègle est illustrée par la Figure 10, p. 38.

**ANNEXE B : Trois contraintes syntaxiques sur les extractions**

Il nous semble utile de citer trois contraintes surfaciques sur les extractions du français.

**Contrainte SyntP sur la relation I.** Si un nucléus verbal impliqué dans une extraction contient une relation SyntP  $R_k = I$  dont la tête  $L(\langle Q_{k+1} \rangle)$  est un verbe, alors  $R_k$  doit être exprimée dans la structure syntaxique de surface par la relation SyntS quasi-subjectale (c'est-à-dire avec un IL impersonnel) et non par la relation subjectale :

(25) a. *C'est la fille qu'il déplaît à Paul que Jean épouse.*

b. \* *C'est la fille que que Jean épouse déplaît à Paul.*

Dans la SSyntP de la phrase (25a),  $R_k$  est la relation actancielle I entre  $L(\langle Q_k \rangle) = \text{DÉPLAIRE}$  et  $L(\langle Q_{k+1} \rangle) = \text{ÉPOUSER}$ .

**Contrainte SyntP sur le régime des verbes non ponts.** Si un nucléus verbal contient un verbe non pont, il est non souhaitable que ce verbe impose le même régime que le dernier verbe du nucléus verbal, c'est-à-dire le verbe qui régit le nucléus nominal :

(26) a. *? un livre auquel je tiens à ce que Paul s'intéresse*

b. *un livre que je tiens à ce que Paul lise*

c. *un livre sur lequel je tiens à ce que Paul se prononce*

En (26-a), le verbe TENIR régit un complément d'objet introduit par la préposition À et, comme on le voit, l'acceptabilité est moindre lorsque le pronom relatif est lui-même un complément d'objet introduit par préposition À.

**Contrainte SyntS de linéarisation** Un syntagme qui comprend un pronom subordonnatif et qui est complément de nom d'un nom N ne peut pas être antéposé à N si N n'est pas sujet ou objet direct :

(27) a. \* *l'homme dont tu parles à la fille<sub>N</sub>*

b. *l'homme dont la fille<sub>N</sub> dort*

c. *l'homme dont tu fréquentes la fille<sub>N</sub>*

**Envoyé en juin 1998**

**Accepté en septembre 1999**

**RÉFÉRENCES**

- AHO Alfred & ULLMAN Jeffrey (1972) : *The Theory of Parsing, Translation and Compiling, Vol. I: Parsing*, London : Prentice-Hall.
- BOYER Michel & LAPALME Guy (1985) : "Generating paraphrases from meaning-text semantic networks", *Computational Intelligence, 1*, pp. 103-117.
- CANDITO Marie-Hélène & KAHANE Sylvain (1998a) : "Une grammaire TAG vue comme une grammaire Sens-Texte précompilée", *Actes TALN'98*, Paris, pp. 40-49.

- CANDITO Marie-Hélène & KAHANE Sylvain (1998b) : “Defining DTG derivations to get semantic graphs”, *Proc. 4th workshop TAG*, Philadelphie, pp. 25-28.
- CHOMSKY Noam (1965) : *Aspects of the Theory of Syntax*, Cambridge, Mass. : MIT Press (trad. franç. Paris : Le Seuil, 1971).
- CHOMSKY Noam (1973) : “Conditions on Transformations”, in Anderson & Kiparsky (eds), *A Festschrift for Morris Halle*, New York : Holt, Rinehart & Wilson, 232-286, (trad. franç. in Chomsky, *Essais sur la forme et le sens*, 1980, Paris : Le Seuil).
- CHOMSKY Noam (1977) : “On Wh-movement”, in P. Culicover & al. (eds), *Formal Syntax*, Dordrecht : Reidel.
- CHOMSKY Noam (1981) : *Lectures on Government and Binding*, Dordrecht : Foris.
- DANLOS Laurence (1985) : *Génération automatique de textes en langues naturelles*, Paris : Masson.
- DANLOS Laurence (1998) : “G-TAG : un formalisme lexicalisé pour la génération de textes inspiré de TAG”, *T.A.L.*, 39 :2, pp. 7-34.
- GAZDAR Gerald, KLEIN Ewan, PULLUM Geoffrey K. & SAG Ivan A. (1985) : *Generalized Phrase Structure Grammar*, Cambridge, MA : Harvard Univ. Press.
- GODARD Danièle (1988) : *La syntaxe des relatives en français*, Paris : CNRS.
- HOVY Eduard H. (1987) : “Generating natural language under pragmatics constraints”, *Journal of Pragmatics*, 11 :6, pp. 689-719.
- HUDSON Richard A. (1990) : *English Word Grammar*, Oxford : Blackwell.
- IORDANSKAJA Lidija (1990) : “Ot semantičeskoj seti k glubinno-sintaksičeskomu derevu : pravila naxoždenija veršiny dereva [Du réseau sémantique à l'arbre syntaxique profond : les règles pour la détermination du sommet de l'arbre]”, in Z. Saloni (ed.), *Metody formalne w opisie języków słowiańskich*, Białystok : Wydawnictwo Filii UW w Białymstoku, pp. 33-46.
- IORDANSKAJA Lidija & POLGUÈRE Alain (1988) : “Semantic processing for text generation”, in *Proc. First International Computer Science Conf. - 88*, Hong Kong, pp. 310-18.
- JESPERSEN Otto (1937) : *Analytic Syntax*, Chicago : Chicago U. Press.
- KAHANE Sylvain (1996) : “If HPSG were a dependency grammar ...”, *Actes TALN'96*, Marseille, pp. 45-49.
- KAHANE Sylvain (1997) : “Bubble trees and syntactic representations”, in Becker & Krieger (eds), *Proc. 5th Meeting of the Mathematics of Language (MOL5)*, Saarbrücken : DFKI, pp. 70-76.
- KAPLAN Ronald & ZAENEN Annie (1989) : “Long distance dependencies, constituent structure and functional uncertainty”, in A. Baltin & A. Kroch (éds), *Alternative conceptions of phrase structure*, Univ. of Chicago Press, pp. 17-42.
- KROCH Anthony (1987) : “Subjacency in a tree adjoining grammar”, in Manaster-Ramer (éd.), *Mathematics of Language*, Amsterdam : Benjamins, pp. 143-72.
- McKEOWN K. (1985) : *Text generation*, Cambridge : Cambridge Univ. Press.

## SYNTHÈSE DES PHRASES À EXTRACTION

- MEL'ČUK Igor (1964): "Tipy svjazej meždú èlementami teksta i tipologija jazykov" [Types de liens entre les éléments textuels et la typologie linguistique], in L.I. Rojzenzon (éd.), *Materialy konferencii "Aktual'nye voprosy sovremennogo jazykoznanija i lingvističeskoe nasledie E.D. Polivanova"*, Vol. I, Samarkand: SamGU, pp. 57-59.
- MEL'ČUK Igor (1974): *Opyt teorii lingvističeskix modelej «Smysl ⇔ Tekst»* [Esquisse d'une théorie de modèles linguistiques «Sens ⇔ Texte»], Moscou: Nauka.
- MEL'ČUK Igor (1988): *Dependency Syntax: Theory and Practice*, Albany, NY: State Univ. of New York Press.
- MEL'ČUK Igor (1993-1999): *Cours de morphologie générale*, 5 vol., Montréal: Presses de l'Univ. Montréal / Paris: CNRS.
- MEL'ČUK Igor (1997): *Vers une Linguistique Sens-Texte*, Leçon inaugurale au Collège de France, Paris: Collège de France.
- MEL'ČUK Igor (1999): *Communicative Organization in Natural Language (The Semantic-Communicative Structure of Sentences)*, Amsterdam: Benjamins.
- MEL'ČUK Igor & al. (1984, 1988, 1992, 1999): *Dictionnaire explicatif et combinatoire du français contemporain*, vol. 1-4, Montréal: Presses de l'Univ. Montréal.
- MEL'ČUK Igor & PERTSOV Nikolaj (1987): *Surface Syntax of English: A Formal Model within the Meaning-Text Framework*, Amsterdam: Benjamins.
- MEL'ČUK Igor, CLAS André & POLGUÈRE Alain (1995): *Introduction à la lexicologie explicative et combinatoire*, Louvain-la-Neuve: Duculot.
- NOGIER Jean-François (1991): *Génération automatique de langages et graphes conceptuels*, Paris: Hermès.
- POLLARD Carl & SAG Ivan A. (1994): *Head-Driven Phrase Structure Grammar*, CSLI series, Chicago: Univ. Chicago Press.
- POLGUÈRE Alain (1990): *Structuration et mise en jeu procédurale d'un modèle linguistique déclaratif dans un cadre de génération de texte*, Thèse de l'Univ. de Montréal.
- POLGUÈRE Alain (1998): "Pour un modèle stratifié de la lexicalisation en génération de texte", *T.A.L.*, 39:2, pp. 57-76.
- RIZZI Luigi (1978): "Violations of the wh-island constraint and the subjacency condition", *Montreal Working Paper in Linguistics XI* (repris dans Rizzi, *Issues in Italian Syntax*, Dordrecht: Foris, 1982).
- ROSS John (1967): *Constraints on Variables in Syntax*, PhD Thesis, MIT (publié sous le titre *Infinite Syntax!*, Dordrecht: Reidel, 1985).
- SAG Ivan A. & FODOR Janet D. (1996): "Une analyse sans catégories vides des phénomènes d'extraction", *Langages*, 122, Paris: Larousse.
- SOWA John S. (1984): *Conceptual Structures: Information Processing in Mind and Machine*, Reading, MA: Addison-Wesley.
- TESNIÈRE Lucien (1959): *Éléments de syntaxe structurale*, Paris: Klincksieck.
- ŽOLKOVSKIJ Alexandr & MEL'ČUK Igor (1967): "O semantičeskom sinteze [Sur la synthèse sémantique]", *Problemy kibertiki*, 19, pp. 177-238, [Trad. fr.: *T.A. Informations*, 2, pp. 1-85, 1970.]