

Université de Montréal

La place des connaissances lexicales
face aux connaissances du monde
dans le processus d'interprétation des énoncés

par Kornél Robert BANGHA

Thèse présentée à la Faculté des études supérieures
en vue d'obtention du grade de
Philosophiae Doctor (Ph. D.) en linguistique
option Intelligence Artificielle

Août 2003

© Kornél Robert Bangha, 2003

Université de Montréal
Faculté des études supérieures

Cette thèse intitulée :
La place des connaissances lexicales
face aux connaissances du monde
dans le processus d'interprétation des énoncés

présentée par :
Kornél BANGHA

a été évaluée par un jury composé des personnes suivantes :

Richard Kittredge
président-rapporteur

Alain Polguère
directeur de recherche

Marie-Claude L'Homme
membre du jury

Sabine Bergler
examinatrice externe

Philippe Langlais
représentant du doyen de la FES

Résumé

Dans cette thèse, nous poursuivons un double but : un but général et un but spécifique. Le but général de cette thèse est d'étudier la place qu'occupent les connaissances lexicales face aux connaissances du monde dans le processus d'interprétation des énoncés. Plus particulièrement, nous cherchons à déterminer dans quelle mesure les connaissances nécessaires à l'interprétation des énoncés sont encodées dans la langue. Bien entendu, c'est une question fondamentale pour laquelle il n'existe pas de réponse communément admise. Étant donné son caractère général, nous ne pouvons l'étudier de façon détaillée que dans un domaine plus restreint, celui de l'interprétation des expressions définies – c'est notre but spécifique.

Notre but pourrait paraître entièrement théorique. Cependant, il a également un intérêt pratique énorme pour le traitement automatique des langues naturelles. En effet, pour accomplir des tâches intelligentes liées aux langues naturelles – la traduction automatique ou les questions-réponses, pour ne citer que deux exemples – les ordinateurs doivent être dotés de connaissances appropriées. Or, pour pouvoir les doter de telles connaissances, il faut au préalable déterminer quelles sont ces connaissances, quel est leur rôle exact. Toutefois, d'un point de vue pratique, l'important n'est pas de savoir comment l'interprétation des énoncés se fait chez les humains mais de déterminer le rôle qu'un certain type de connaissances, dans notre cas les connaissances lexicales, peut jouer dans ce processus.

Pour atteindre notre but, nous examinons d'abord le rapport entre les connaissances linguistiques et les connaissances du monde et ensuite les projets principaux qui cherchent à les modéliser. Nos recherches sont menées dans le cadre de la Lexicologie Explicative et Combinatoire (LEC) et notre outil principal est le Dictionnaire de Combinatoire. Nous définissons de façon précise ce que nous entendons par l'interprétation des énoncés et nous faisons une étude de cas : nous analysons les expressions définies d'un corpus.

Les résultats de l'analyse montrent notamment que, dans l'interprétation du défini, ce sont les fonctions lexicales (surtout celles qui modélisent les liens actanciels) telles que postulées par la LEC, qui jouent le rôle principal. De façon plus générale, ces analyses suggèrent que les connaissances linguistiques et le contexte linguistique contiennent une très grande partie des informations nécessaires à l'interprétation des énoncés.

Mots clés

Connaissances lexicales, connaissances du monde, interprétation, expressions définies, sémantique, lexicologie, fonction lexicale, actant, linguistique informatique, intelligence artificielle.

Abstract

The aim of this research is, on one hand, general, and on the other hand, specific. The general aim is to study to what extent the process of interpretation of statements is based on lexical knowledge rather than on world knowledge. We are particularly interested in knowing to what extent the knowledge needed for interpretation is part of the language. This question being too general to be properly researched, it was necessary to look at it from a more limited angle: the interpretation of definite expressions – this is our specific aim.

Besides its theoretical interest, this is also an essential issue in Natural Language Processing where computers need a great deal of knowledge. In order to provide the necessary knowledge to computers, we first need to examine what this knowledge consists of, what role it plays. However, from a computational point of view, the most important point is not to know how human interpretation works but rather to determine what piece of knowledge is needed for any linguistic task.

First of all, we make a clear distinction between world knowledge and linguistic knowledge. After that, we present the framework of this study: the Explanatory Combinatorial Lexicology. An entire chapter is concerned about the notion of interpretation. At the end, we conduct a study dedicated to definite expressions and we

analyse a corpus of twelve texts.

The conclusion of our study is that linguistic knowledge (and particularly lexical functions) together with the information contained in the texts we analysed, can grant us most of the knowledge needed for the interpretation of definite expressions.

Key words

Lexical knowledge, world knowledge, interpretation, definite expression, semantics, lexicology, lexical function, actant, computational linguistics, artificial intelligence.

Table des matières

Résumé	III
Abstract	V
Table des matières	VII
Abréviations	XIV
Conventions d'écriture	XIV
1 Introduction	1
1.1 Objectifs de recherche	1
1.2 Méthodologie.....	3
1.3 Originalité et intérêt de notre travail	4
1.4 Remarques sur le cadre théorique	6
2 Connaissances lexicales et connaissances du monde	8
2.1 Peut-on distinguer deux types de connaissances ?	9
2.1.1 Éventail des points de vue	9
2.1.1.1 Haiman : la distinction entre encyclopédies et dictionnaires est théoriquement intenable	10
2.1.1.2 Eco : les sémantiques en forme de dictionnaires sont inconsistantes	12
2.1.1.3 Lakoff : réalisme expérimental et théorie des prototypes	13
2.1.1.4 Lenat & Guha : prédominance des connaissances du monde	14
2.1.1.5 Kayser : pas de véritable frontière entre le linguistique et l'extralinguistique.....	15
2.1.1.6 Fellbaum : relations sémantico-conceptuelles et relations lexicales .	16

2.1.1.7 Murphy : connaissances au sujet des mots.....	17
2.1.1.8 Pustejovsky : le lexique génératif.....	18
2.1.1.9 Katz et Fodor : la frontière entre connaissances linguistiques et connaissances du monde est nécessaire.....	18
2.1.1.10 Quine : vérité analytique et vérité empirique.....	19
2.1.1.11 Wierzbicka : les connaissances linguistiques sont différentes des connaissances encyclopédiques.....	20
2.1.2 Notre position.....	21
2.1.2.1 Remarques sur l'éventail des points de vue.....	21
2.1.2.2 Concepts et lexies.....	23
2.1.2.3 Trois rapports possibles entre les CL et les CM.....	24
2.1.2.4 En résumé.....	26
2.2 Projets de modélisation des connaissances du monde.....	27
2.2.1 Intelligence Artificielle et connaissances du monde.....	27
2.2.2 Cyc.....	30
2.2.2.1 Base de connaissances de Cyc.....	31
2.2.2.2 Cyc et la compréhension des langues naturelles.....	34
2.2.3 Mikrokosmos de Nirenburg.....	35
2.2.3.1 Présentation de l'ontologie de Nirenburg.....	36
2.2.3.2 Remarques sur l'ontologie de Nirenburg.....	40
2.2.4 Projets « unificateurs ».....	41
2.3 Projets de modélisation des connaissances lexicales.....	43
2.3.1 Lexique génératif de Pustejovsky.....	43
2.3.1.1 Présentation.....	44

2.3.1.2 Exemples de générations de sens	48
2.3.1.3 Remarques sur l'approche générative	49
2.3.2 Connaissances lexicales dans les dictionnaires traditionnels	52
2.3.3 WordNet.....	54
2.3.3.1 Présentation de WordNet.....	55
2.3.3.2 Remarques sur les informations lexicales de WordNet.....	56
3 Cadre linguistique de notre recherche : la lexicologie explicative et combinatoire	60
3.1 Présentation sommaire	61
3.1.1 Objet et but de la lexicologie explicative et combinatoire	61
3.1.2 Critères méthodologiques explicites de la LEC	63
3.2 Identification des connaissances encodées dans le lexique.....	72
3.2.1 Introduction à la problématique d'encodage des CM	72
3.2.2 Connaissances lexicalisées	75
3.2.3 Connaissances encodées dans la définition lexicale.....	77
3.2.3.1 Détermination du genre prochain	78
3.2.3.2 Détermination des différences spécifiques.....	79
3.2.3.3 Prise en considération de la pertinence linguistique	82
3.2.4 Connaissances encodées à travers les liens lexicaux.....	83
3.2.4.1 Intérêt des liens lexicaux pour l'encodage des connaissances	84
3.2.4.2 Positions actanciennes d'une lexie et participants d'une situation	85
3.2.4.3 Dérivations sémantiques.....	93
3.2.4.4 Collocations.....	95
3.3 Place du DiCo dans notre recherche.....	98
3.3.1 Pourquoi choisir le DiCo pour notre recherche ?	99

3.3.2	Présentation du DiCo.....	100
3.3.3	Dix champs principaux de description d'une lexie dans le DiCo	102
4	Délimitation de l'interprétation des énoncés.....	105
4.1	Présentation de la notion d'énoncé.....	106
4.1.1	Langue et parole, type et token	106
4.1.2	Définition de l'énoncé.....	107
4.2	Compréhension des énoncés	108
4.2.1	Problématique de la compréhension.....	108
4.2.2	Compréhension : d'un point de vue pratique	110
4.2.3	Compréhension : approches théoriques.....	111
4.2.4	En résumé.....	115
4.3	Interprétation des énoncés	116
4.3.1	Définition de l'interprétation.....	117
4.3.2	Choix de l'interprétation	119
4.4	Contexte des énoncés	122
4.4.1	Définition du contexte.....	123
4.4.2	Illustration du contexte.....	124
4.5	Trois problèmes d'interprétation.....	126
4.5.1	Désambiguïsation	126
4.5.2	Paraphrase	129
4.5.3	Résolution de l'anaphore.....	132
5	Une étude de cas : les expressions définies.....	134
5.1	Trois approches concernant le défini.....	134
5.1.1	Études du défini en philosophie analytique.....	135
5.1.2	Études du défini en grammaire.....	138
5.1.3	Études du défini en linguistique	140
5.2	Caractéristiques du défini.....	146

5.2.1 Trois types de rapport entre le défini et l'indéfini.....	147
5.2.1.1 Opposition effective	147
5.2.1.2 Opposition neutralisée.....	148
5.2.1.3 Absence d'opposition.....	148
5.2.2 Défini et indéfini dans les locutions, les collocations et les expressions génériques.....	150
5.2.2.1 Choix du défini et de l'indéfini dans les locutions et les collocations	150
5.2.2.2 Choix du défini et de l'indéfini dans les expressions génériques....	155
5.2.3 Le défini dans les expressions libres	156
5.2.3.1 Opposition entre défini et indéfini dans les expressions libres	157
5.2.3.2 Descriptions définies	159
5.2.4 Identification contextuelle du référent d'un défini.....	160
5.3 Introduction aux analyses de l'interprétation du défini.....	162
5.3.1 Intérêt des analyses du défini pour notre recherche	162
5.3.2 But et portée des analyses	165
5.3.3 Méthodologie des analyses.....	166
5.3.3.1 Présentation générale de la méthodologie	166
5.3.3.2 Huit types d'identification du défini	170
5.3.4 Connaissances lexicales nécessaires aux analyses	181
5.3.4.1 Connaissances présentes dans le DiCo.....	181
5.3.4.2 Élaboration des CL utilisée dans l'analyse.....	182
5.3.4.3 Vérification d'une partie des connaissances utilisées dans les analyses	186
5.4 Analyse de l'interprétation du défini dans les textes.....	188

5.4.1 Premier lieu : la gare	190
Premier texte.....	190
Deuxième texte.....	196
Troisième texte.....	200
Quatrième texte	207
5.4.2 Deuxième lieu : l'école.....	214
Cinquième texte.....	214
Sixième texte	220
Septième texte	228
5.4.3 Troisième lieu : la mairie.....	234
Huitième texte	234
Neuvième texte.....	241
Dixième texte	247
5.4.4 Quatrième lieu : le jardin public.....	252
Onzième texte.....	252
Douzième texte.....	255
5.5 Évaluation des analyses.....	264
5.5.1 Résultats des analyses	264
5.5.1.1 Statistiques.....	264
5.5.1.2 Commentaires type par type.....	266
5.5.1.3 Évaluation générale des analyses	270
5.5.2 Limites de l'analyse.....	273
6 Conclusion.....	276
6.1 Aperçu du travail accompli	276

6.2 Quatre questions qui restent sans réponse.....	278
Index.....	281
Bibliographie.....	283
Annexe : articles du DiCo	I
Premier article : GARE	I
Deuxième article : ÉCOLE.....	IV
Troisième article : MAIRIE	VIII
Quatrième article : JARDIN PUBLIC.....	X

Abréviations

ASém :	Actant sémantique
ASynS :	Actant syntaxique de surface
CL :	Connaissances lexicales
CM :	Connaissances du monde
DD :	Description définie
DiCo :	Dictionnaire de combinatoire (base de données lexicales)
FL :	Fonction lexicale
FP :	Forme propositionnelle
IA :	Intelligence Artificielle
LEC :	Lexicologie explicative et combinatoire
LG :	Lexique Génératif
NPR :	Le Nouveau Petit Robert
TAL :	Traitement Automatique des Langues
TST :	Théorie Sens-Texte

Conventions d'écriture

Format	Valeur	Exemples
Majuscule	Lexie	GARE, CHEF DE GARE
Souligné	Concept	<u>manger</u> , <u>nourriture</u>
Entre guillemets simples	Sens	‘très’, ‘peur’
Italique	Expression (type)	<i>devant la gare</i>

1 Introduction

1.1 Objectifs de recherche

Le but de notre travail est d'apporter un meilleur éclairage sur la place qu'occupent les connaissances lexicales (CL), face aux connaissances du monde (CM), dans le processus d'interprétation des énoncés. Nous comptons atteindre ce but en trois étapes :

- délimiter ce que sont les CL, notamment par rapport aux CM ;
- définir la notion d'interprétation des énoncés ;
- étudier la place des CL face aux CM dans un domaine à la fois restreint et représentatif : l'interprétation du défini.

Dans ce qui suit, nous présentons brièvement ces trois étapes.

Bien qu'il soit tout à fait trivial de dire que les CL (qui sont un sous-ensemble des connaissances linguistiques) et les connaissances du monde jouent un rôle très important dans toute activité langagière, il n'existe pas du tout de consensus sur la nature exacte de ce rôle. En regardant de près les travaux des linguistes et des chercheurs en sciences cognitives ou en Intelligence Artificielle, on rencontre des posi-

tions diamétralement opposées. Nous examinons d'abord les principales positions concernant ces connaissances et nous nous penchons ensuite plus particulièrement sur les CL, en cherchant à répondre aux questions suivantes :

Que sont les connaissances lexicales ? Notamment : Comment les définir ? Que couvrent-elles ? Quel est leur rapport aux connaissances du monde ?

La notion d'interprétation des énoncés peut paraître intuitive et facile à saisir mais en l'examinant de plus près, on constate qu'il n'existe pas d'unanimité à son sujet. Pour la clarifier, il faut se poser une deuxième série de questions :

Qu'est-ce que l'interprétation des énoncés ? Notamment : Dans quelle mesure est-ce une tâche linguistique ? Dans quelle mesure implique-t-elle une interaction avec nos connaissances du monde ? Peut-on faire une distinction entre interprétation et compréhension ?

Déterminer le rôle exact des connaissances lexicales dans le processus d'interprétation des énoncés en général serait une tâche largement au-dessus des limites d'une thèse. Il nous est toutefois possible de faire des analyses dans un domaine restreint et de postuler que les résultats obtenus sont représentatifs. Ainsi, nous pouvons analyser un corpus afin d'examiner **le rôle que jouent les connaissances lexicales dans l'interprétation des expressions définies.**

1.2 Méthodologie

Les objectifs que nous venons de mentionner déterminent la structure et la méthodologie de notre travail : examiner l'état d'art des recherches concernant les connaissances lexicales et les connaissances du monde, choisir et présenter notre cadre théorique, spécifier la notion d'interprétation et examiner la place des CL dans un domaine restreint.

Le deuxième chapitre présente les principales approches concernant les CL et les CM. La première section de ce chapitre porte sur le rapport entre les deux types de connaissances : est-il nécessaire, voire possible, de les séparer ? Pour quelles raisons faudrait-il les séparer : pratiques ou théoriques ? La deuxième section s'intéresse aux CM : elle examine trois approches d'IA dont le but est de représenter ces connaissances. La troisième section est consacrée aux CL : elle décrit également trois approches dont le but est d'énumérer les CL.

Le troisième chapitre expose le cadre théorique que nous avons adopté pour notre recherche : la Lexicologie explicative et combinatoire (LEC). Bien entendu, la LEC est l'outil principal de ce travail, pas son objet. Après en avoir fait une présentation générale, nous nous intéressons aux CL que la LEC postule : y a-t-il des critères linguistiques pour les définir ? Que contiennent-elles ? Selon quels formalismes les représenter ? La dernière partie du chapitre décrit le Dictionnaire de combinatoire, le DiCo, que nous utilisons dans notre étude de cas pour évaluer la place des CL.

Dans le quatrième chapitre, consacré à l'interprétation, nous justifions d'abord notre intérêt pour l'interprétation, notamment face à la génération. Ensuite, nous proposons une distinction entre l'interprétation et la compréhension : nous considérons la première comme un processus plus restreint, inclus dans la seconde. À la fin du chapitre, nous illustrons les enjeux de l'interprétation dans trois domaines spécifiques : la désambiguïsation, le paraphrasage et la résolution de l'anaphore.

Le cinquième chapitre contient une étude de cas : l'analyse des expressions définies. La première section est consacrée à trois approches concernant le défini : celle des grammairiens, celle des philosophes et celle des linguistes. Nous proposons notre propre analyse du défini dans la deuxième section. La troisième section est une introduction aux analyses présentées dans la quatrième section. Dans la cinquième section, nous faisons une évaluation de ces analyses.

1.3 Originalité et intérêt de notre travail

La question principale que nous nous posons, à savoir la place des CL face aux CM, n'est pas originale en elle-même. L'originalité de notre travail se manifeste, d'une part, dans la méthode utilisée, d'autre part, dans les résultats obtenus.

Dans notre analyse, nous utilisons les CL telles qu'elles sont modélisées par la LEC et énumérées dans un dictionnaire informatisé, le Dictionnaire de Combinatoire (DiCo). La richesse et les critères méthodologiques rigoureux de cette lexicologie nous permet d'inclure dans les CL plus que ce que contiennent la majorité des lexiques des systèmes de traitement de la langue, tout en les distinguant des CM.

Les résultats des analyses nous apportent deux nouveautés. La première consiste à montrer que les CL, telles que décrites par la LEC et le DiCo, jouent (ou peuvent jouer) un rôle considérablement plus important que ce que leur attribuent les autres approches examinées. Deuxièmement, ces résultats semblent nous indiquer que, contrairement à ce que soutiennent notamment de nombreux chercheurs en IA, la quantité des connaissances que couvrent (aussi) les CL lors du processus d'interprétation est bien supérieure à celle qu'incluent seulement des CM. Bien entendu, l'importance des CM nous paraît incontestable et nous ne cherchons point à en démontrer le contraire mais plutôt à montrer que, dans bien des cas, on peut s'en passer si on dispose de suffisamment de CL.

Déterminer la place des CL face aux CM est une question théorique en soi mais elle devient un enjeu central lorsque l'on s'intéresse au traitement automatique des langues (TAL). En effet, dans la mesure où les humains possèdent les connaissances nécessaires à l'accomplissement des tâches linguistiques, il importe peu de savoir la nature exacte de ces connaissances. Par contre, dans le cas des programmes informa-

tiques qui doivent accomplir ces mêmes tâches linguistiques, il est indispensable d'identifier ces connaissances pour pouvoir les fournir à la machine.

1.4 Remarques sur le cadre théorique

Nous nous situons dans un cadre saussurien qui considère la langue comme un système autonome. Par conséquent, les CL sont aussi autonomes par rapport au CM. Cette autonomie a deux aspects. Premièrement, il n'y a pas d'inclusion totale : les CL ne sont pas simplement une sous-classe des CM, un sous-ensemble linguistiquement encodé. Par exemple, connaître le genre grammatical d'un nom est une connaissance lexicale mais, la plupart du temps, n'encode pas une connaissance sur le référent du nom. Le genre féminin attribué à TABLE ou le masculin attribué à BUREAU est complètement indépendant de tout ce que nous savons concernant les tables et les bureaux. Deuxièmement, il n'y a pas d'exclusion totale non plus, les CL et les CM ne sont pas mutuellement exclusives : de nombreuses CM sont encodées dans la langue. C'est notamment le cas de l'hyponymie : nous savons, d'une part, qu'un bureau est une table et qu'une table est un meuble, d'autre part, qu'il existe une lexie BUREAU dont le sens inclut 'table' et qu'il y a aussi une lexie TABLE dont le sens inclut 'meuble'.

Bien que la majorité des recherches scientifiques se situent à l'intérieur d'une discipline précise, il faut parfois se poser des questions qui dépassent nécessairement les

frontières de la discipline dont elles relèvent de façon centrale. Il en est ainsi de la question que nous nous posons concernant la place des CL. En premier lieu, cette question relève de la lexicologie. Cependant, telle qu'elle est formulée (la place des CL face aux CM dans le processus d'interprétation), elle nécessite aussi d'entrer dans d'autres disciplines, notamment dans l'IA.

2 Connaissances lexicales et connaissances du monde

Il est clair que les humains possèdent une quantité immense de connaissances et qu'ils les utilisent pour accomplir des tâches intelligentes variées, y compris des tâches linguistiques, mais la nature et le contenu exacts de ces connaissances restent à déterminer. Sont-elles seulement et tout simplement des connaissances sur le monde qui nous entoure ? Ou bien sont-elles des connaissances encodées dans telle ou telle langue particulière ? Y a-t-il un moyen pour les énumérer et les structurer ? Si oui, selon quels critères ?

La première section de ce chapitre est consacrée à la nature même de ces connaissances. Il faut d'abord se demander s'il est possible de distinguer deux types de connaissances : les CM et les CL. Les avis sur cette question sont très partagés : les uns pensent que oui, les autres pensent que non, certains affirment leur avis de façon très catégorique alors que d'autres le font de façon plus nuancée. Nous présentons au lecteur des positions de chacun de ces camps.

La deuxième et la troisième sections présentent chacune trois projets dont le but est d'énumérer respectivement les CM et les CL. Nous ne nous intéressons pas tant aux travaux qui proposent tout simplement un formalisme pour représenter les connais-

sances mais à des projets qui ont effectivement entrepris l'énumération des connaissances.

2.1 Peut-on distinguer deux types de connaissances ?

Bien qu'il soit fréquemment fait référence à la distinction entre les CL et les CM (pensons aux dictionnaires et aux encyclopédies), elle est loin d'être évidente pour tout le monde. Peut-on vraiment distinguer les CL et les CM ? Nous tenons à illustrer ici la diversité des points de vue concernant cette question, en commençant par ceux qui contestent le plus cette distinction (souvent dans une optique de sciences cognitives) et en allant vers les positions qui distinguent clairement deux types de connaissances. Nous ne souhaitons pas engager une discussion détaillée à leur sujet mais, à la fin de la section, nous présentons notre propre position. Certains écrits de ces auteurs seront repris aux chapitres suivants.

2.1.1 Éventail des points de vue

Notre but est de présenter la diversité (et la divergence) des points de vue concernant la séparation des CL et des CM. Nous ne cherchons pas à résumer les livres et les articles cités mais à présenter quel est l'avis des auteurs sur la question de cette séparation. Si ces présentations sont relativement sommaires, c'est parce que nous tenons à les limiter autant que possible à ce qui a un rapport direct à notre sujet.

Les points de vues présentés sont très diversifiés et forment presque un continuum entre les deux positions extrêmes : d'une part, nier la possibilité d'une distinction théorique entre les CL et les CM, d'autre part, affirmer la nécessité de faire une distinction claire entre elles. Nous avons choisi de présenter ce continuum en allant d'un extrême à l'autre plutôt que d'essayer de regrouper ces positions en catégories plus ou moins arbitraires.

2.1.1.1 Haiman : la distinction entre encyclopédies et dictionnaires est théoriquement intenable

En se penchant sur la distinction entre les dictionnaires et les encyclopédies, Haiman (1980) s'intéresse en premier lieu à la question des définitions que l'on trouve dans des dictionnaires. Il présente la position de la sémantique structurale qui croit qu'il existe un petit nombre de concepts de base à partir desquels tous les autres peuvent être dérivés¹. Si de tels concepts primitifs existent, il est important de les identifier. Haiman voit trois possibilités pour ce faire :

- (a) Les primitifs sont choisis par ceux qui les analysent comme étant les plus simples.
- (b) Les primitifs sont enracinés dans l'expérience du monde.

¹ Cf. Leech (1974 : 5-6 et 207) et Wierzbicka (1972 : 25).

(c) Les primitifs sont innés.

Haiman estime que les deux premières possibilités ne permettent pas de distinguer les dictionnaires des encyclopédies. Il soutient que la simplicité n'est pas une notion objective et différentes personnes font des choix différents, notamment en fonction de leur expérience du monde. La possibilité (b) est basée sur une distinction entre ce que Russell appelle les mots-objets (dont le sens vient directement du monde, par ostension) et les mots de dictionnaire (dont le sens est expliqué à l'aide d'autres mots). Haiman objecte à cette approche qu'il est difficile de distinguer entre les deux types de mots et surtout que, directement ou indirectement, chaque mot dépend entièrement de nos connaissances du monde. La possibilité (c) lui semble inacceptable parce que, soutient-il, sans expérience, il n'y a pas de pensée et certainement pas de langage.

Haiman considère que la dénotation est à la base du sens : les définitions doivent avoir une base empirique, les mots doivent être ancrés dans la réalité. Autrement dit, il soutient que les connaissances sémantiques dérivent des connaissances culturelles : elles ne sont qu'un sous-ensemble de ces dernières avec des frontières floues. Or, la distinction entre dictionnaires et encyclopédies est basée sur la distinction entre sens et dénotation. Une conséquence directe de son point de vue est que la distinction entre dictionnaires et encyclopédies est théoriquement intenable. Il admet cependant que cette distinction a sa raison d'être puisqu'elle fonctionne bien en pratique. Il compare cette situation à la division de la grammaire en phonologie, mor-

phologie, syntaxe et sémantique au début du structuralisme : c'était aussi une question pratique à l'origine mais elle a été, selon Haiman à tort, élevée au statut de principe théorique avec le temps.

2.1.1.2 Eco : les sémantiques en forme de dictionnaires sont inconsistantes

Eco (1988) consacre tout un chapitre à la distinction entre dictionnaires et encyclopédies. Son but est de démontrer qu'un modèle de définition par genre et espèce est intenable. Pour ce faire, il se penche d'abord sur les dictionnaires et les définitions qu'ils contiennent. Une définition est une équation entre un signifié et un définissant et il soutient que ce dernier doit être composé de primitifs sémantiques dont le nombre est limité. Ainsi, les deux problèmes suivants doivent être résolus : comment déterminer les primitifs utilisés et comment limiter leur nombre ?

Eco reprend les trois possibilités mentionnées par Haiman (1980) pour identifier les primitifs et il les considère également intenable. Il trouve cependant intéressant d'envisager une quatrième possibilité où ces primitifs seraient organisés selon un arbre d'hyponymes et d'hyperonymes. Après avoir examiné cette possibilité en détail, il arrive à la conclusion suivante : ou bien les primitifs ne peuvent pas être interprétés et alors on ne peut pas expliquer le sens des termes (on ne définit pas le signifié), ou bien ils peuvent et doivent être interprétés et on ne peut pas limiter leur nombre. Qui plus est, si les primitifs sont interprétés, l'arbre cesse d'être un dictionnaire pour devenir fatalement une encyclopédie. Autrement dit, les sémantiques en

forme de dictionnaires sont inconsistantes, il faut avoir recours aux sémantiques en forme d'encyclopédie.

2.1.1.3 Lakoff : réalisme expérimental et théorie des prototypes

Lakoff (1987) est un plaidoyer de ce que l'auteur appelle le réalisme expérimental : il y prend position contre l'objectivisme. Ce dernier fait une distinction entre les propriétés essentielles d'un objet et ses propriétés accidentelles. Les connaissances définitoires correspondent aux propriétés essentielles des entités et des catégories alors que les connaissances encyclopédiques correspondent à leurs propriétés accidentelles. Il s'agit, bien entendu, de la distinction entre dictionnaires et encyclopédies.

Le réalisme expérimental de Lakoff refuse cette distinction. En effet, il refuse même l'idée selon laquelle les catégories sont basées sur les propriétés essentielles communes de leurs membres. Il préfère une autre théorie de catégorisation, la *théorie des prototypes*, développée par Eleanor Rosch. Cette théorie ne fait pas de distinction entre les propriétés essentielles et les propriétés accidentelles et, du même coup, refuse aussi la distinction entre dictionnaires et encyclopédies.

2.1.1.4 Lenat & Guha : prédominance des connaissances du monde

Nous présenterons dans (2.2.2) le projet Cyc suivant Lenat et Guha (1990). Selon les auteurs, il existe bel et bien une séparation entre les CM et les CL. Toutefois, ils mettent en valeur les CM et considèrent notamment que les connaissances linguistiques ne jouent pas de rôle significatif dans le processus de compréhension. Ils illustrent leur position par les exemples suivants :

(1a) *The ink is in the pen.*

(1b) *The pig is in the pen.*

(2a) *The little girl saw the bicycle in the window. She wanted it.*

(2b) *The little girl saw the bicycle in the window. She pressed her nose up against it.*

Désambiguïser *pen* dans (1a) et (1b), affirment-ils, (instrument d'écriture ou enclos) ou encore trouver l'antécédent de *it* dans (2a) et (2b) demande des CM ; ce ne sont pas nos connaissances de l'anglais qui permettent de le faire. Autrement dit, ils font une distinction entre CM et CL sans toutefois accorder un rôle significatif aux CL.

2.1.1.5 Kayser : pas de véritable frontière entre le linguistique et l'extralinguistique

Kayser (1997) considère qu'il n'existe pas de véritable frontière entre le linguistique et l'extralinguistique, même s'il y a des phénomènes qui se situent clairement d'un côté ou de l'autre. Il admet notamment que les dictionnaires doivent contenir des CL mais il affirme qu'ils présupposent, entre autres, une maîtrise suffisante d'une masse de connaissances présumées connues de tout être humain en âge de les consulter qui ne peuvent pas y figurer. Sans ces connaissances supplémentaires, on ne pourrait pas utiliser les dictionnaires.

Pour illustrer sa position, il donne toute une liste de connaissances concernant l'école qu'il considère comme des CM (voir Kayser (1997 : 31-32)). Il affirme qu'une telle liste de connaissances est hétéroclite, et qu'il n'y a aucun moyen de savoir si elle est raisonnablement complète, ou même seulement représentative de l'ensemble de nos connaissances sur le monde de l'école. Avant de penser à représenter de telles connaissances, il faudrait pouvoir les recenser, ou au moins évaluer leur quantité, et il faut admettre que, pour le moment au moins, nous n'avons guère les moyens de le faire. Kayser ajoute que l'on n'a jamais vu un concept se promener sans ses habits de la langue. Autrement dit, les connaissances du monde sont difficiles à énumérer non seulement parce qu'elles sont potentiellement infinies mais aussi parce que nous n'y avons pas d'accès direct, il faut passer par la langue !

2.1.1.6 Fellbaum : relations sémantico-conceptuelles et relations lexicales

Dans Fellbaum *et al.* (1998), on trouve une présentation de WordNet qui est une base de données lexicales informatisée. Nous en faisons une présentation plus détaillée dans la section (2.3.3). Les auteurs admettent que les dictionnaires sont, en général, des recueils des connaissances des mots et les encyclopédies des recueils des connaissances du monde mais ils considèrent que les frontières entre les deux sont floues. Ils proposent une distinction entre les relations sémantico-conceptuelles et les relations lexicales. Ce sont les ensembles de synonymes – et non pas les mots – qui sont liés par différentes relations sémantico-conceptuelles : synonymie, hyponymie/hyperonymie, **méronymie** (relation entre le tout et ses parties), implication (par exemple, *ronfler* implique *dormir*), etc. Par contre, l'antonymie est une relation lexicale qui concerne les mots plus que les concepts.

Le résultat d'une telle division est que les liens conceptuels sont considérés comme plus forts, plus importants que les liens lexicaux car, dans les cas où les deux types de lien existent, ils sont classés parmi les liens conceptuels. En effet, WordNet ne reconnaît des liens lexicaux que si ces liens n'ont pas d'équivalent sur le plan conceptuel.

2.1.1.7 Murphy : connaissances au sujet des mots

Murphy (2000) s'intéresse à l'organisation du lexique, notamment à savoir quel est le rôle des relations sémantiques dans cette organisation. L'auteur soutient que ces relations font partie non pas du lexique mais de nos connaissances encyclopédiques. Ce qui, pour nous, est particulièrement intéressant dans cet article est la distinction entre trois types de connaissances :

- 1) connaissances lexicales, c'est-à-dire connaissances des mots ;
- 2) connaissances conceptuelles relatives aux mots, c'est-à-dire connaissances au sujet des mots ;
- 3) connaissances conceptuelles relatives aux dénotations des mots, c'est-à-dire connaissances du monde.

Le premier et le troisième type sont assez intuitifs mais le deuxième demande des précisions. Il s'agit très clairement d'un type de CM. Il inclut les connaissances et la réflexion des linguistes ou des locuteurs sur les mots. Il est clair par exemple que savoir que tel ou tel mot est fréquemment utilisé dans les mots croisés n'est pas une connaissance lexicale mais une connaissance conceptuelle relative aux mots.

Curieusement, Murphy considère que les relations lexicales (synonymie, antonymie, etc.) ne devraient pas figurer dans des dictionnaires car elles sont des connaissances conceptuelles au sujet des mots et qu'elles font partie de nos CM.

2.1.1.8 Pustejovsky : le lexique génératif

Le lexique génératif de Pustejovsky est présenté en détail dans la section (2.3.1). Pustejovsky (1995) postule une séparation nette entre les CL et les CM mais l'originalité de son travail réside dans les limites qu'il établit entre ces deux types de connaissances. En effet, il inclut dans le lexique génératif des connaissances que les lexicologues avant lui reléguaient presque toujours aux encyclopédies. (Cf. Peeters (2000) et Taylor (2000)) La structure de qualia notamment contient systématiquement des informations sur, par exemple, comment l'objet dénoté par l'unité lexicale en question a été créé, de quoi il est fait, etc.

2.1.1.9 Katz et Fodor : la frontière entre connaissances linguistiques et connaissances du monde est nécessaire

Katz & Fodor (1963) affirment que la grammaire et le dictionnaire ne sont pas les seules composantes d'une théorie sémantique : il faut aussi ajouter des règles de projection. Les **règles de projection** choisissent, à partir des informations qui se trouvent dans le dictionnaire, le sens approprié pour chaque unité dans l'énoncé. Elles sont indispensables parce qu'un dictionnaire offre plus de sens pour chaque unité lexicale que ce que cette unité pourrait avoir dans un énoncé donné. Ces règles nécessitent que l'on prenne en considération la frontière qui existe entre les connaissances que l'on possède au sujet de la langue et celles au sujet du monde car c'est

cette frontière qui permet de définir l'objet d'une théorie sémantique. Sans cette frontière, la compréhension des énoncés pourrait faire appel à toutes nos connaissances du monde. Les auteurs proposent les énoncés suivants :

(3a) *Our store sells alligator shoes.*

(3b) *Our store sells horse shoes.*

Pour savoir que (3a) est synonyme de *Our store sells shoes made from alligator skins*, il faut savoir que les alligators ne portent pas de chaussures mais qu'il existe des chaussures faites de peau d'alligator. Pour savoir que (3b) est synonyme de *Our store sells shoes for horses*, il faut également savoir que les chevaux ont des sabots. Autrement dit, disent les auteurs, il faut avoir recours à des connaissances du monde. Remarquons cependant que cette analyse est fort contestable puisqu'elle ne tient pas compte de l'existence de l'unité lexicale HORSE SHOE.

2.1.1.10 Quine : vérité analytique et vérité empirique

Quine (1953) établit une distinction claire entre vérité analytique et vérité empirique des énoncés. La première est de nature linguistique, elle est vraie en fonction du sens des unités lexicales qui composent l'énoncé alors que la deuxième relève de nos connaissances du monde, elle est vraie en fonction de ce qu'est le monde. Nous pouvons illustrer cela par des énoncés suivants :

(4a) *9 est supérieur à 7.*

(4b) *Le nombre des planètes du système solaire est égal à 9.*

L'énoncé (4a) décrit une vérité analytique : quiconque connaît la définition de 9, 7 et celle de l'expression *être supérieur à* doit savoir qu'il est vrai. Le sens des unités lexicales qui s'y trouvent garantit sa vérité, nul besoin de savoir comment le monde est. Par contre, l'énoncé (4b) dénote une vérité empirique : pour affirmer qu'il est vrai, il faut savoir comment est le monde, savoir notamment le nombre des planètes dans le système solaire – connaître seulement le sens des unités qui le composent ne permet pas de savoir s'il est vrai. Autrement dit, la vérification de (4a) fait appel aux CL alors que celle de (4b) fait appel aux CM. Bien entendu, tout cela présuppose une séparation claire entre les CL et les CM.

2.1.1.11 Wierzbicka : les connaissances linguistiques sont différentes des connaissances encyclopédiques

En s'intéressant aux primitifs sémantiques, Wierzbicka (1996) examine les définitions de plusieurs dictionnaires et constate qu'elles incluent souvent des informations qui relèvent des encyclopédies et non pas des dictionnaires. Elle admet que la limite entre sens et connaissances encyclopédiques puisse être difficile à établir mais elle soutient qu'en principe cela est faisable.

Wierzbicka affirme qu'en analysant la langue, on peut trouver des preuves qui permettent de faire une distinction claire entre les connaissances linguistiques et les connaissances encyclopédiques. Cela suggère qu'il y a aussi une distinction entre un « dictionnaire mental » et une « encyclopédie mentale » dans le cerveau.

2.1.2 Notre position

Nous résumons ici notre position concernant la distinction entre CM et CL – une position qui sera notre postulat de travail. En premier lieu, nous revenons brièvement sur les points de vue mentionnés plus haut, ensuite, nous nous penchons sur la distinction entre concepts et lexies.

2.1.2.1 Remarques sur l'éventail des points de vue

Tout en admettant que la frontière entre le linguistique et l'extralinguistique puisse ne pas être facile à établir dans certains cas, nous sommes d'accord avec Wierzbicka pour dire qu'il existe des méthodes qui permettent de distinguer avec précision ce qui est une connaissance linguistique et ce qui n'en est pas une.

Contrairement à Haiman, Eco ou Lakoff, nous avons la conviction qu'il existe des CL, clairement distinctes des CM, sans toutefois être mutuellement exclusives. Notre position est basée sur l'observation du caractère fondamentalement arbitraire de la langue. Les locutions et les collocations en sont des exemples par excellence.

Bien entendu, tout cela ne signifie pas qu'il n'existe pas de parallèle entre les CL et les CM, au contraire. Notre position consiste à dire que les CM seules ne permettent pas de prévoir ce qui est encodé dans la langue et surtout comment se fait l'encodage.

Contrairement à Kayser, ou aux auteurs de Cyc et de WordNet, nous croyons qu'une bonne partie des connaissances du monde a son équivalent lexical et doit se trouver dans un bon dictionnaire. Nous considérons que la position de ces auteurs est erronée et qu'elle sous-estime l'importance des CL. Nous verrons plus tard que les connaissances concernant l'école que Kayser énumère sont, en grande partie, encodées dans la langue aussi. On peut par exemple penser à la lexie CLASSE qui dénote, par définition, un regroupement d'élèves (*Il a 35 élèves dans sa classe.*) ou aux lexies BULLETIN (DE NOTE) et EXAMEN qui couvrent également une partie des connaissances que Kayser considère comme non linguistiques. Quant à l'approche de Cyc, elle est contestable pour la simple raison qu'elle n'est basée sur aucune théorie linguistique qui lui permettrait de démontrer les véritables frontières des CL.

Comme Pustejovsky, Katz et Fodor, Quine ou Wierzbicka, nous croyons qu'il est incontournable de modéliser les CL et de les utiliser dans le traitement du langage naturel. Dans le troisième chapitre, nous présentons le cadre théorique que nous utilisons pour le faire. Nous tenons simplement à souligner dans ce qui suit un dernier aspect de la distinction entre CL et CM : la distinction entre concepts et lexies.

2.1.2.2 Concepts et lexies

La distinction entre les CM et les CL peut être liée à la distinction entre les concepts et les lexies. En effet, les CM s'articulent autour des concepts comme les CL s'articulent autour des lexies. Or, le flou qui existe à propos des CM existe également à propos des concepts.

Selon *Le Nouveau Petit Robert*, un concept est une représentation mentale et générale d'un objet. L'étude des concepts devrait aller bien plus loin que cette définition : elle devrait notamment énumérer les concepts et définir chacun d'entre eux.

L'énumération des concepts est déjà problématique. Il en existe combien ? Personne ne le sait. Existe-t-il un concept manger ? Beaucoup diraient, intuitivement, que oui. Pourtant, on pourrait aussi argumenter, en s'inspirant de Onyshkevych et Nirenburg (1995) que manger n'est pas un concept mais une composition de deux éléments conceptuels : celui d'ingérer et celui de nourriture. Cette analyse ne fait que poser d'autres questions : ingérer et nourriture sont-ils des concepts eux-mêmes ou la combinaison d'autres concepts ?

Même lorsque l'on a isolé un concept, il est très difficile, pour ne pas dire impossible, d'établir ce qu'il contient. Selon quels critères peut-on définir un concept ? Bien entendu, on peut toujours donner une réponse ad hoc, plus ou moins intuitive, mais nous ne connaissons aucune réponse précise à cette question. Lenat et Guha (1990)

relèvent ce problème en parlant de ce qu'ils appellent un **objet riche** : c'est un objet dont on pourrait dire une infinité de choses. Ils ajoutent que, dans le monde, il existe très peu de choses dignes d'intérêt qui ne soient pas des objets riches. Les concepts sont un peu comme ces objets riches : on pourrait en dire une infinité de choses.

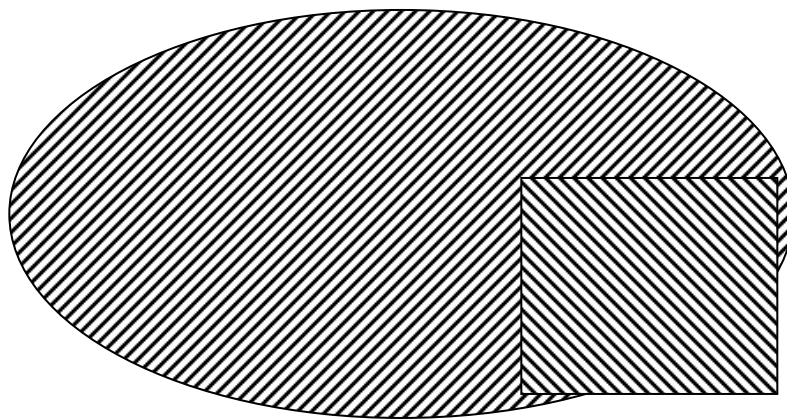
Contrairement aux concepts, les unités lexicales sont plutôt tangibles. Nous avons à notre disposition une quantité innombrable de phrases écrites et orales qui les contiennent. Nous pouvons les analyser en détail, y compter l'occurrence de chaque unité, etc. Bien entendu, en donner une description complète demande un travail énorme, mais cependant réalisable. La division des vocables en lexies² peut se faire selon des critères rigoureux et il existe également des critères rigoureux pour définir chaque lexie – c'est ce que nous allons voir dans le troisième chapitre.

2.1.2.3 Trois rapports possibles entre les CL et les CM

Notre position concernant le rapport entre les CM et les CL peut être illustrée par le schéma 1 (l'ellipse représente les CM et le rectangle représente les CL) :

² La définition exacte de *vocable* et *lexie* est donnée au début du troisième chapitre.

Schéma 1 : CM et CL.



Ce schéma divise nos connaissances en trois parties :

1) La grande partie de nos CM n'est pas encodée dans la langue. Nous savons par exemple qu'une maison coûte – en général – bien plus cher qu'un ordinateur de bureau, nous savons que Napoléon a perdu la bataille de Waterloo, etc. Ces connaissances ne sont pas encodées dans la langue, n'ont pas leurs équivalents linguistiques. Bien entendu, nous pouvons les exprimer à l'aide de la langue : en formulant des énoncés ou en écrivant des textes.

2) Certains CM sont encodées dans la langue, ont leur équivalent dans la langue. C'est notamment le cas de la relation de l'hyponymie : d'une part, nous savons par

exemple que le chien, le chat, le cheval etc. sont des animaux, d'autre part, la lexie ANIMAL est un hyperonyme des lexies CHIEN, CHAT, CHEVAL, etc.

3) Il existe des CL qui ne peuvent pas être considérées comme l'encodage linguistique des CM. C'est notamment le cas du genre grammatical pour la majorité des noms ou, de façon encore plus générale, de ce que la TST appelle la combinatoire lexicale. Par exemple, le genre grammatical des lexies TABLE ou BUREAU, respectivement féminin et masculin, n'encode pas une CM sur les objets que sont les tables et les bureaux.

2.1.2.4 En résumé

Nous croyons que la distinction entre CM et CL est possible et même nécessaire aussi bien pour la modélisation du processus de compréhension (et celui de la génération) que pour le traitement du langage naturel par ordinateur. Sans exclure les CM, les CL doivent avoir une place prioritaire dans l'étude du processus d'interprétation car elles sont plus concrètes et plus directement accessibles.

Dire que la distinction est possible entre ces deux types de connaissances ne signifie pas que ces connaissances soient mutuellement exclusives mais qu'il est possible de décider si une connaissance fait partie de nos CL, de nos CM ou des deux à la fois – même si cette décision n'est pas toujours facile. Dans le troisième chapitre, nous présentons la Lexicologie Explicative et Combinatoire, qui propose de nombreux

critères et tests permettant de rédiger des articles de dictionnaire en distinguant ce qui est une CL de ce qui n'en est pas une.

2.2 Projets de modélisation des connaissances du monde

Cette section est consacrée aux approches qui s'intéressent à la modélisation des CM et qui, par conséquent, pourraient apporter une contribution significative au problème de l'interprétation des énoncés. Nous examinons d'abord quel est le rôle que les chercheurs en IA attribuent aux connaissances du monde. Nous présentons deux projets, à savoir Cyc et Mikrokosmos, dont le but est de créer effectivement des bases de connaissances générales. Pour terminer, nous nous penchons aussi sur des projets de base de connaissances plus spécifiques qui pourraient avoir un véritable intérêt s'il était possible de les réunir pour former une base de connaissances générales.

2.2.1 Intelligence Artificielle et connaissances du monde

Le but de l'IA est de créer des ordinateurs capables d'accomplir des tâches relativement complexes qui, de la part des humains, nécessiteraient une forme d'intelligence. Pour y arriver, les ordinateurs doivent, d'une part, posséder des connaissances, d'autre part, avoir la capacité de les gérer. On peut distinguer deux types de méthodes : les méthodes fortes et les méthodes faibles. Les méthodes fortes

cherchent à accomplir les tâches qui leur sont demandées avant tout à l'aide de nombreuses connaissances, souvent limitées à un domaine précis. Les méthodes faibles, quant à elles, mettent l'accent sur la gestion des connaissances, les raisonnements, etc. sans chercher à accumuler une grande quantité de connaissances.

Comme nos recherches visent à déterminer la nature des connaissances utilisées pour la compréhension, nous nous penchons sur les méthodes fortes. Luger (2002) considère que la nécessité d'utiliser des bases de connaissances pour résoudre des problèmes complexes se justifie par une simple observation : la cause de la bonne performance des experts humains est qu'ils possèdent une grande quantité de connaissances dans leur domaine. Toutefois, si l'on veut procéder ainsi pour la compréhension des énoncés en général (c'est-à-dire des énoncés non spécifiques à un domaine précis), on se heurte à un énorme problème : il faut recueillir et représenter toutes les connaissances susceptibles d'être utiles. Étant donné qu'il faut déjà une grande quantité de connaissances pour un seul domaine spécifique, on peut facilement imaginer qu'il faudra des efforts gigantesques pour créer une base de données qui ne soit pas limitée à un domaine précis.

On peut concevoir la création d'une telle base de connaissances de deux façons différentes : soit on la crée directement, soit on crée d'abord des bases de connaissances spécifiques et on cherche à les unir ensuite. Quelle que soit l'option choisie, la réalisation de ces projets pourrait être divisée en deux phases : il y a d'abord une phase préliminaire qui consiste à élaborer une méthodologie, une manière de repré-

sentation ; ensuite, on a une deuxième phase, qui est la construction même des bases de connaissances.

On trouve de nombreux travaux consacrés à la phase préliminaire. En général, ils proposent de représenter les connaissances de façon explicite, à l'aide de modèles formels divers : réseaux sémantiques (Collins & Quillian (1969), Brachman (1979)), dépendances conceptuelles (Schank & Rieger (1974)), calcul des prédicats, scénarios (Bartlett (1932), Schank & Abelson (1977)), cadres (Minsky (1975)), graphes conceptuels (Sowa (1984)), etc. Il existe aussi des alternatives à une représentation explicite (Brooks (1991), Mitchell (1993), Hofstadter (1995)). Toutefois, comme le remarque Onyshkevich & Nirenburg (1995), il ne suffit pas d'utiliser des formalismes pour construire un modèle de représentation des connaissances : il faut aussi construire effectivement des bases de connaissances.

Or, la construction effective de ces bases reste le plus gros problème. Certes, il existe des projets dans des domaines spécifiques – et par conséquent réduits – qui ont complété la phase deux. Malheureusement, aucun d'entre eux n'a créé une base générale complète jusqu'à maintenant. Une telle base de connaissances devrait contenir toutes les connaissances pouvant être nécessaires à la compréhension des énoncés interprétables par un locuteur moyen. Même le projet Cyc, qui semble être le plus avancé, n'y est pas arrivé.

Nous avons donc vu que, de façon générale, l'IA met l'accent sur les connaissances du monde pour accomplir des tâches intelligentes. Quelle est sa position exacte concernant le traitement des langues naturelles ? Les chercheurs en IA affirment qu'une grande quantité de nos connaissances du monde dans lequel nous vivons est nécessaire pour le traitement automatique des langues naturelles. Cette affirmation mérite d'être prise au sérieux : nous considérons qu'il existe des cas où elle est la plus adéquate. Toutefois, son principal défaut est, à nos yeux, qu'elle sous-estime l'ampleur des connaissances linguistiques et notamment celle des CL que les humains utilisent. Il est important de souligner que l'approche IA fait une affirmation bien plus forte que de dire tout simplement que les CM peuvent être utiles pour la compréhension. Leur position pourrait plutôt se résumer ainsi :

Un ordinateur doté de CL (ou, plus généralement, de connaissances linguistiques) mais dépourvu de CM ne pourrait pas accomplir de manière fiable et adéquate des tâches relatives au traitement des langues naturelles.

2.2.2 Cyc

Cette sous-section est consacrée au projet Cyc : nous faisons d'abord une présentation générale de sa vision et de sa façon de procéder, ensuite nous examinons plus particulièrement son positionnement face à la compréhension des énoncés.

2.2.2.1 Base de connaissances de Cyc

Lenat et Guha (1990) constatent que plusieurs programmes dont le but était d'exécuter des tâches « intelligentes » ont échoué là où les humaines réussissent sans difficulté. Mentionnons l'exemple du système expert bancaire qui ignorait que quelqu'un de 18 ans ne peut pas travailler dans le même endroit depuis 20 ans³. Contrairement à la machine, un humain aurait tout de suite vu que cela est impossible. Qu'est-ce qui explique cette différence ? Les humains ont à leur disposition un ensemble de connaissances (générales et spécifiques) à consulter – ce qui n'est pas le cas des ordinateurs. Ces connaissances contiennent non seulement des faits statiques mais aussi des méthodes pour résoudre des problèmes. Pour construire des ordinateurs intelligents, il faut donc les doter de ces connaissances.

On a surtout l'habitude de construire des systèmes experts qui sont limités à un domaine bien précis et qui contiennent des connaissances sur ce domaine. Or, avoir plusieurs petites bases de connaissances, chacune décrivant un domaine précis, n'équivaut pas à en avoir une seule grande qui couvrirait plusieurs domaines. En effet, en cherchant à unir deux (ou plusieurs) systèmes experts, on pourrait rencontrer deux sources d'incompatibilités. La première est que chaque règle a en elle les hypothèses et les simplifications de son propre système expert et la deuxième est que la terminologie et le découpage du monde peuvent varier d'un système à l'autre.

³ Pour d'autres exemples, voir Lenat & Guha (1990 : 1-3).

Par conséquent, il vaudrait mieux construire une grande base de connaissances dès le départ.

La base de connaissances de Cyc contient les connaissances qui sont nécessaires à la compréhension d'une encyclopédie ou d'un journal par exemple. Il ne s'agit pas avant tout du contenu de ces écrits, mais des connaissances que les auteurs supposent que leurs lecteurs ont déjà. Elle contient aussi des connaissances qui sont nécessaires pour répondre à des questions auxquelles n'importe qui pourrait répondre après avoir lu un texte. On y trouve également des connaissances qui expliquent pourquoi certaines histoires peuvent nous paraître incroyables. Ces connaissances, comme dans la plupart des programmes d'IA, peuvent se diviser en deux classes : les faits et les règles d'inférence. En effet, Cyc contient aussi de nombreux mécanismes d'inférence.

Comment Cyc procède-t-il ? Une des hypothèses de base en IA est que l'on peut simuler l'intelligence dans un « micro-monde » particulier en manipulant un ensemble de symboles qui représentent ce micro-monde. On peut s'y prendre de deux façons différentes. Une approche possible est de choisir un domaine particulier et y résoudre des problèmes précis. L'approche opposée serait de chercher un mécanisme de raisonnement, construit indépendamment de tout domaine et de toute connaissance particulière. La logique formelle préconise la deuxième approche alors que les systèmes experts favorisent la première. Cyc choisit une solution intermé-

diaire et son monde est l'ensemble des domaines de toutes les tâches particulières, jusqu'à un niveau pré-expert.

Cyc voit ce monde comme étant composé d'un certain nombre de choses reliées entre elles. Ces choses sont regroupées dans différentes collections. À part la collection universelle, Chose, chaque collection est un sous-ensemble propre d'une autre collection. Chaque unité doit être le membre d'au moins une collection nommée, la collection la plus générale étant Chose. La catégorie Chose peut être divisée d'une part en Collection et en Objet Individuel, d'autre part en Objet Tangible, Objet Intangible et Objet Complexe Tangible & Intangible.

Le monde est représenté dans Cyc à l'aide d'un langage de cadre (*frame-based language*) : toutes les assertions au sujet d'un objet (par exemple, Texas) sont rassemblées dans une structure de données que l'on appelle cadre. Ces unités ont des emplacements (*slot*) et chaque emplacement a une valeur qui est un ensemble.

Par exemple :

```
Texas
  capital :          ( Austin )
  résidents :       ( Doug Guha Mary )
  état de :         ( États Unis d'Amérique )
```

L'unité Texas a ici trois emplacements, chacun correspond à un prédicat.

2.2.2.2 Cyc et la compréhension des langues naturelles

Lenat et Guha soutiennent que l'on ne peut pas comprendre les langues naturelles tant que l'on ne connaît pas la grande partie de ce qu'ils appellent la « réalité consensuelle », c'est-à-dire un ensemble de connaissances que tout le monde est supposé posséder. (Bien entendu, le projet Cyc ne s'intéresse pas seulement aux connaissances partagées par tous mais aussi à celles que seulement tel ou tel groupe d'humains possèdent.) Reprenons les exemples donnés dans (2.1.1.4) :

(5a) *The ink is in the pen.*

(5b) *The pig is in the pen.*

(6a) *The little girl saw the bicycle in the window. She wanted it.*

(6b) *The little girl saw the bicycle in the window. She pressed her nose up against it.*

L'occurrence de *pen* dans (5a) dénote un stylo et celle dans (5b) dénote un enclos. Comment le sait-on ? Comment sait-on que la fillette voulait la bicyclette dans (6a) et non pas la vitrine ? Ce ne sont pas nos connaissances de la langue qui nous le disent – selon Lenat et Guha – mais nos connaissances des choses du monde (qu'est-ce qui contient de l'encre, où sont les cochons, quels sont les objets qui attirent les fillettes, et ainsi de suite) : la désambiguïsation de l'occurrence de *pen* dans (5a) et

dans (5b) ainsi que l'identification de l'antécédent du pronom *it* dans (6a) ou dans (6b) se fait en s'appuyant sur nos connaissances du monde.

Il est important de remarquer que les auteurs du projet Cyc font une double affirmation ici : d'une part, concernant les connaissances du monde, d'autre part concernant les connaissances linguistiques. La première partie de cette affirmation est assez évidente : les connaissances nécessaires aux tâches mentionnées dans le paragraphe précédent font (aussi) partie de nos CM. Par contre, la deuxième partie est problématique : à savoir que nos connaissances linguistiques ne nous aident pas à désambiguïser ces énoncés. Nous soutenons que la désambiguïstation est possible, au moins pour (5a) et (5b), à l'aide des CL.

Comme le projet Cyc n'est pas encore terminé, on ne peut pas vraiment évaluer dans quelle mesure il couvre les connaissances nécessaires à l'interprétation des énoncés. Qui plus est, Cyc s'intéresse aux CM sans s'intéresser à des CL. Par conséquent, il ne pourrait pas nous aider pour discerner quelle est la part respective de ces deux types de connaissances.

2.2.3 Mikrokosmos de Nirenburg

D'abord nous présentons le projet Mikrokosmos de Nirenburg, ensuite nous faisons des remarques concernant la sémantique qui y est utilisée.

2.2.3.1 Présentation de l'ontologie de Nirenburg

Le projet Mikrokosmos de Nirenburg *et al.* (Onyshkevych et Nirenburg (1995), voir aussi Carlson et Nirenburg (1992) et Nirenburg *et al.* (1992)) est avant tout un projet de traduction automatique à l'aide d'une base de connaissances (*Knowledge-Based Machine Translation* ou KBMT). Ce projet, comme d'autres projets KBMT, utilise une base de connaissances pour permettre aux ordinateurs de faire de la traduction – il est ainsi différent des projets de Traduction Automatique qui fonctionnent sur des statistiques. Il vise une compréhension fonctionnelle des textes en vue d'une traduction efficace. Par compréhension fonctionnelle, il faut entendre une représentation du sens qui permet la traduction vers plusieurs langues. Il ne s'agit pas d'une compréhension totale qui inclurait toute information implicite et explicite.

De façon générale, Mikrokosmos ressemble à d'autres projets de KBMT : étant donné un texte en langue source, après l'avoir analysé par plusieurs programmes en utilisant les connaissances pragmatiques, lexicales et grammaticales de la langue en question, on en extrait son sens et on le représente dans un langage neutre, l'interlingua. La différence entre ces projets réside dans la précision avec laquelle l'interlingua représente le texte source. L'interlingua de Mikrokosmos est un langage de cadre (*frame-based language*) : les cadres dénotent les instances des objets ontologiques et les emplacements (*frame slots*) dénotent les propriétés des concepts. Ces emplacements peuvent avoir quatre types de valeurs : *value*, *default*, *sem* et *saliance*. Le premier type dénote la valeur actuelle de la propriété alors que le second

dénote la valeur par défaut. Ces deux types de valeur sont en distribution complémentaire : la valeur par défaut est mentionnée si et seulement si la valeur actuelle n'est pas connue. Le troisième type de valeur apporte des contraintes sémantiques sur les deux premiers. Le quatrième type de valeur indique l'importance de cette propriété. Nous pouvons illustrer tout cela par l'exemple (7) :

(7)

```
(TABLE
...
      (MADE-OF (default *wood)
                (sem (*solid)
                    (salience 0.7)))
...
)
```

Cet exemple est un extrait de la représentation du concept de la table ; il décrit la propriété « fait de quoi ». Ce formalisme doit être compris de la façon suivante : par défaut, une table est faite de bois. Elle peut être faite d'autres matériaux aussi mais ces matériaux ont la contrainte sémantique⁴ d'être solides. La valeur d'importance de cette propriété est élevée (0.7), ce qui signifie qu'il s'agit d'une propriété importante du concept. Par contre, la propriété qu'une table a, entre autre, en général qua-

⁴ Pour nous, cette contrainte semble être plutôt conceptuelle que sémantique. Pour Nirenburg, les deux sont inséparables.

tre pieds (qui est également une propriété par défaut) n'aurait qu'une importance réduite.

Un énoncé comme (8) pourrait être représenté par (9) – les entités et les événements sont numérotés pour identification :

(8) *Fido mange une saucisse.*

(9)

```
(DOG323
  (INSTANCE-OF *DOG)
  (NAME « FIDO »))
(SAUSAGE23
  (INSTANCE-OF *SAUSAGE))
(INGEST17
  (INSTANCE-OF *INGEST)
  (AGENT      (VALUE DOG323)
              (SEM *ANIMAL)))
  (THEME      (VALUE SAUSAGE23)
              (SEM *INGESTIBLE)))
```

Ce formalisme dit que :

- il y a un chien (numéroté 323) qui s'appelle *Fido*,
- il y a une saucisse (numéroté 23),
- il y a un événement d'ingestion (numéroté 17) dont l'agent est le chien numéro 323 et cet agent doit être un animal. En outre, le thème de cet événement est la saucisse numéro 23 qui doit être digestible.

Nirenburg soutient que l'interlingua doit être interprétée en terme d'un modèle de monde indépendant. Ce modèle est nécessaire aussi bien pour des raisons théoriques que pour des raisons pratiques. En sémantique formelle, une des approches les plus répandues est celle où les énoncés sont interprétés en fonction d'un modèle du monde. Dans les applications informatiques, l'interprétation des unités textuelles n'est faisable qu'en présence d'un modèle du monde détaillé dont les éléments sont liés, directement ou indirectement, à différentes parties du texte par un lien « être une signification pour ».

Les connaissances dans le modèle du monde sont séparées en deux bases de données reliées entre elles. La première est la véritable base ontologique qui contient des connaissances concernant les concepts. Elle est organisée de façon taxonomique pour des raisons de stockage et d'efficacité. La deuxième est une base onomastique : elle est la collection d'instances spécifiques des concepts ontologiques. Par exemple, les concepts président américain ou marque de voiture sont dans l'ontologie alors que les connaissances au sujet du président américain Harry Truman ou de la marque de voiture Saab sont dans la base onomastique.

Les éléments, les entités de ce modèle du monde, sont eux-mêmes connectés les uns aux autres à travers un ensemble de liens ontologiques bien définis. Ces liens sont des propriétés des entités et ils font partie du modèle. En effet, au plus haut niveau de la classification ontologique, les concepts sont classés en deux catégories : entités

et propriétés. On pourrait cependant considérer que même les propriétés sont des entités puisqu'il faut décrire leur nature ontologique aussi bien que celle des entités.

Comme la sémantique doit être basée sur une ontologie (cf. la *Discourse Representation Theory* (DRT) de Kamp (1981)), le sens lexical doit référer à un concept ontologique. Plus précisément, le sens lexical doit référer à un concept ontologique dans des cas simples. Dans des cas plus complexes, le sens est une combinaison de concepts ontologiques avec des contraintes. Par exemple, le sens de la lexie MANGER serait le concept ingérer avec deux contraintes : le sens de l'agent doit faire référence au concept animal et le sens de l'objet doit faire référence au concept objet physique.

2.2.3.2 Remarques sur l'ontologie de Nirenburg

Le projet de Nirenburg repose sur plusieurs idées intéressantes, notamment la séparation des connaissances en une base ontologique et une base onomastique. Toutefois, on n'y trouve pas de séparation entre les CM et les CL – ce qui serait le plus important de notre point de vue. Certes, il est question de sémantique dans ce projet mais cette sémantique ne fait qu'une avec l'ontologie. En effet, le seul contenu (et le seul but) des connaissances lexicales serait de lier des lexies à des concepts (ou combinaisons de concepts) dans l'ontologie. Il ne pourrait y avoir aucun lien direct entre deux lexies : il ne pourrait y avoir que des liens indirects, à travers les concepts

qu'elles dénotent (qui peuvent effectivement être liés les uns aux autres). Qui plus est, le projet semble être arrêté sans être terminé et utilisable.

2.2.4 Projets « unificateurs »

Nous venons d'examiner deux projets, Cyc et Mikrokosmos, dont le but est de recueillir et représenter des connaissances générales. Dans le cas de la majorité des applications informatiques, qui ont des objectifs pratiques, tel n'est pas le cas : les systèmes experts, par exemple, s'intéressent plutôt à des connaissances liées à un domaine précis selon des critères spécifiques. On peut se demander s'il serait possible d'unir un certain nombre de bases de connaissances spécifiques afin d'en créer une grande base générale. Lenat (voir plus haut) a déjà avancé des arguments contre cette idée.

Schank et Abelson (1977) sont également opposés à l'unification de plusieurs bases de connaissances. Ils considèrent que lors d'un échange linguistique, les interlocuteurs se servent des connaissances du monde de façon implicite. Pour pouvoir les transmettre à l'ordinateur, ils proposent de les formaliser à l'aide de scénarios (*script*). Ils contestent l'idée de pouvoir organiser les connaissances de façon hiérarchique, avec des ensembles et des membres. Dans une telle organisation, le principal lien entre les éléments serait l'appartenance : par exemple, les canaris font partie des oiseaux, les oiseaux font partie des animaux, etc. Toutefois, ils considèrent que même s'il est assez aisé de catégoriser, de hiérarchiser les noms qui dénotent des

entités concrètes, il serait très difficile d'en faire autant avec les noms dénotant les entités abstraites ou avec les verbes. De façon générale, ils restent sceptiques face à un système universel des connaissances. Ils ne voient pas pourquoi les connaissances sur les sujets aussi différents que les équations mathématiques, la contrebande de stupéfiants, l'écriture des poèmes ou la préparation de repas devraient être représentées selon le même schéma. Les critères de généralité et d'élégance ne sont pas suffisants.

Toutefois, ce scepticisme face à la possibilité d'unification de plusieurs bases de connaissances n'est pas partagé par tous les chercheurs. Il existe en effet de nombreuses tentatives de définir des standards et de créer des outils qui permettraient d'unir plusieurs bases de connaissances et ontologies créées indépendamment les unes des autres – cf. Farquhar *et al.* (1996) pour Ontolingua ou Swartout *et al.* (1997) pour Ontosaurus.

Les objectifs du projet *Semantic Web*, développé par le *Word Wide Web Consortium* (<http://www.w3.org/2001/sw/>) sont semblables. Les auteurs du projet constatent que le Web actuel est entièrement destiné à des lecteurs humains et que les ordinateurs sont limités à transmettre et à présenter les informations sur le Web, sans avoir la possibilité de nous aider à les traiter. Leur but est de créer un Web où les informations peuvent être « comprises » aussi bien par la machine que par les humains. La meilleure façon de le faire leur semble être d'encoder les informations sur le Web dans un formalisme que la machine pourra traiter. Ils proposent d'utiliser le format

RDF (*Resource Description Framework*), qui est une forme de XML. Ainsi, le Web deviendrait une énorme base de connaissances que l'on pourrait interroger de façon précise. Remarquons cependant que, malgré son nom, *Semantic Web* s'intéresse peu à la sémantique et cherche plutôt à élaborer un formalisme de représentation.

Une fois de plus, comme ces projets ne sont pas encore achevés, il nous est impossible d'évaluer dans quelle mesure ils pourraient être utiles à l'interprétation des énoncés.

2.3 Projets de modélisation des connaissances lexicales

Dans la section précédente, nous avons vu quelques-uns des projets les plus représentatifs qui s'intéressent aux connaissances du monde. Dans cette section, nous examinons trois approches qui cherchent à modéliser les connaissances linguistiques, plus particulièrement les connaissances lexicales : le lexique génératif de Pustejovsky, les dictionnaires traditionnels et WordNet.

2.3.1 Lexique génératif de Pustejovsky

Nous faisons d'abord une présentation de la structure lexicale que postule le lexique génératif de Pustejovsky (1995). Ensuite nous montrons des mécanismes qui permettent de générer de nouveaux sens à partir de cette structure. Finalement, nous

avançons des considérations concernant cette approche qui expliquent pourquoi nous ne l'avons pas utilisée dans cette recherche pour évaluer la place des CL dans le processus d'interprétation.

2.3.1.1 Présentation

Un **lexique génératif** (LG) est un lexique où les unités lexicales ont un sens de base avec une structure interne relativement complexe. Cette structure permet de générer, grâce à des mécanismes bien précis, un ensemble plus large de sens lorsque l'unité lexicale est combinée avec d'autres unités dans le syntagme ou dans la phrase. Selon Pustejovsky, la structure interne des unités lexicales peut être représentée à l'aide de plusieurs sous-structures dont les trois plus importantes sont la structure argumentale, la structure événementielle et la structure relationnelle.

La **structure argumentale** d'une unité lexicale est la spécification minimale de sa sémantique lexicale. Elle est constituée de quatre types d'arguments. Les véritables arguments doivent être exprimés syntaxiquement (*Jean est arrivé en retard*). Les arguments par défaut font partie de la structure logique mais ne sont pas nécessairement exprimés en syntaxe (*Jean lit un livre*). Les arguments intégrés (*shadow arguments*) sont incorporés dans l'unité lexicale et ne peuvent être exprimés que par spécification (*Jean pleure de l'œil gauche* est possible mais pas **Jean pleure de l'oeil*). Les véritables adjoints modifient l'expression logique, mais ne font pas partie de l'interprétation situationnelle et ne sont pas liés à la représentation d'aucune

unité lexicale (*Marie est allée à New York **mardi***). Les arguments d'une unité lexicale sont représentés dans une structure où le type d'argument est directement encodé.

Nous pouvons illustrer la structure argumentale des deux unités lexicales CONSTRUIRE et BEURRER par les schémas suivants⁵ :

```
(S1)
CONSTRUIRE
  ARGSTR :
    ARG1 = être_animé
    ARG2 = artefact
    ARG_DEFAULT = matériel
```

```
(S2)
BEURRER
  ARGSTR :
    ARG1 = humain
    ARG2 = objet_physique
    ARG_INTEGRE : beurre
```

Le schéma (S1) encode l'information qu'il y a trois arguments dans la structure argumentale du verbe CONSTRUIRE : le premier argument est un être animé, le deuxième un artefact alors que le troisième est un matériel. Les deux premiers sont

⁵ Nous avons légèrement simplifié ces schémas afin d'en faciliter la lecture.

de véritables arguments alors que le troisième est un argument par défaut. Selon (S2), le verbe BEURRER a également trois arguments ; les deux premiers sont de véritables arguments alors que le troisième est un argument intégré.

La **structure événementielle** spécifie les verbes (et tout ce qui peut être un prédicat) qui peuvent être classés notamment en processus (par exemple, COURIR, DORMIR), état (CROYANCE, GRAND) ou les deux à la fois (CONSTRUIRE, CASSER). La causation directe relève également de ce dernier cas car elle contient deux événements : un processus où le premier argument agit sur le deuxième (*Pierre a tué Marie*) et un état dans lequel le deuxième argument se trouve (*Marie est morte*).

(S3)

COURIR

EVENTSTR :

E1 = processus

(S4)

CONSTRUIRE

EVENTSTR :

E1 = processus

E2 = état

RESTR = parties ordonnées exhaustives

TETE = E1

La structure événementielle du verbe COURIR dit simplement qu'il s'agit d'un processus. Celle du verbe CONSTRUIRE est bien plus complexe : elle est composée de deux sous-événements dont le premier (E1) est un processus et le deuxième (E2) un état. Ces deux sous-événements sont ordonnés dans le temps (le premier précède le second) et il n'y a pas d'autre sous-événement. Le marqueur TETE exprime le sous-événement le plus important.

La **structure relationnelle** (*qualia*) spécifie quatre aspects essentiels du sens d'une unité lexicale. L'aspect **constitutionnel** spécifie la relation entre un objet et ses parties (un roman contient une narration, une bibliothèque contient des livres). Le **formel** montre ce qui distingue une unité lexicale dans un domaine plus large (la femme est un être humain de sexe féminin, un roman est un livre). Le **telique** exprime le but (la nourriture doit être mangée, le roman doit être lu, le dictionnaire doit être consulté) et la fonction (le couteau sert à couper quelque chose). L'**agentif** contient les facteurs impliquant les origines de l'objet dénoté (un roman est écrit alors qu'un dictionnaire est compilé).

(S5)

ROMAN

QUALIA

CONST = narration

FORMAL = livre

TELIC = lecture

AGENT = écriture

```

(S6)
CONSTRUIRE
  QUALIA
    FORMAL = existe(ARG2)
    AGENTIF = action_de_construire(E1, ARG1,
ARG2)

```

Dans (S6), le formel indique que le deuxième argument de construire (l'artefact) existe alors que l'agentif montre que l'action de construire est un processus (E1) où le premier argument agit sur le deuxième.

2.3.1.2 Exemples de générations de sens

Pustejovsky considère que les structures que nous venons de présenter sont le point de départ pour de nombreuses opérations sémantiques (coercition de type, co-composition, liage sélectif, sélection sémantique). La structure relationnelle suggère notamment une interprétation contextuelle. Comment ces opérations fonctionnent-elles ?

Nous nous contentons ici d'illustrer le mécanisme de coercition de type. Les verbes comme VOULOIR, COMMENCER ou AIMER sont sémantiquement légers, sous-spécifiés. Lorsque leur complément direct est un nom, le sens du verbe est spécifié par le rôle telique de son argument. Prenons comme exemple les énoncés suivants :

(10) *Bernard a commencé le roman.*

(11) *Agnès aime la bière.*

Le telique de ROMAN est d'être lu, celui de BIÈRE est d'être bue. Par conséquent, ces énoncés seront compris comme ceci :

(12) *Bernard a commencé à lire le roman.*

(13) *Agnès aime boire de la bière.*

Il existe toutefois des cas spécifiques :

(14) *Les vendeurs de livre préfèrent les livres de cuisine aux romans avant Noël.*

Dans cet énoncé, il s'agit de préférer vendre les livres plutôt que de les lire, le telique de l'agent (*vendeur de livres*) prime celui de l'objet (*livre*).

2.3.1.3 Remarques sur l'approche générative

Notre remarque principale concernant le LG porte sur l'absence de dictionnaire type génératif. Il faut constater qu'il existe déjà de nombreux dictionnaires (la sous-section suivante leur est consacrée) qui, contrairement à l'approche générative,

étaient compilés suivant une approche « d'énumération de sens »⁶. Par contre, il n'existe pas de dictionnaire basé sur l'approche LG, c'est-à-dire qui contiendrait un seul sens par vocable et qui permettrait de générer d'autres sens en fonction du contexte.

D'un point de vue pratique, cela pose un problème majeur pour nos analyses du défini dans le chapitre 5 car ces analyses nécessitent une ressource lexicale, ne serait-ce qu'incomplète. Cette absence pose également problème au niveau théorique car seul un tel dictionnaire pourrait donner une réponse convaincante à certaines questions concernant le LG.

La première de ces questions concerne l'absence de CL suffisamment spécifiques dans un LG. Le but d'un lexique génératif est d'utiliser le plus souvent possible des mécanismes de génération afin d'alléger le plus possible le dictionnaire : ne pas inclure dans toute une série d'articles de dictionnaire les informations qui pourraient être obtenues à l'aide des règles générales. Par exemple, plutôt que de contenir des informations spécifiques à chaque lexie (mentionner l'expression *commencer un livre* dans l'article LIVRE ou *prendre le train* dans l'article TRAIN), le LG contiendrait des règles générales : commencer un artefact signifie, par défaut, commencer à

⁶ Il faut entendre par *énumération de sens* ce que l'on désigne en langage courant en disant qu'un mot peut avoir plusieurs sens. Dans le cadre de la LEC, on dirait qu'un vocable contient plusieurs lexies et chaque lexie a un sens différent.

utiliser cet artefact en fonction du telique de l'unité lexicale correspondante. Toutefois, de telles règles, en elles-mêmes, sont beaucoup trop générales car elles ne rendent pas compte de la différence d'interprétabilité qui existe entre expressions similaires. Par exemple, *Jean a commencé un livre* est facilement interprétable (comme *Jean a commencé à lire un livre*) mais *Jean a commencé un journal, une voiture ou un film* ne le sont pas. Bien entendu, ces irrégularités n'impliquent pas que de telles règles ne puissent pas exister. Verspoor (1997) se penche sur ce genre d'exemples et examine notamment la possibilité de concilier les règles génératives et de telles exceptions par des contraintes, des restrictions sélectionnelles. Elle estime que de telles contraintes peuvent rendre compte d'une partie des exceptions seulement. Sa conclusion est que même si la compréhension de la langue peut se faire en général selon certaines règles, ce sont des particularités lexicales qui gouvernent. Nous pouvons en conclure qu'il est de la plus haute importance d'avoir un dictionnaire qui contienne ces particularités.

Un dictionnaire génératif pourrait également apporter un meilleur éclairage sur la distinction entre ce qui est considéré linguistique et ce qui est considéré extralinguistique chez Pustejovsky. Certes, une matrice de Pustejovsky pourrait être vue comme une bonne approximation du sens de l'unité lexicale qu'elle décrit. Toutefois, elle contient aussi de nombreuses informations qui ne relèvent pas toujours des CL mais que l'on devrait plutôt considérer comme des simples CM. On peut penser notamment à la structure relationnelle (*qualia*) qui contiendrait les informations suivant-

tes⁷ : l'aspect constitutionnel inclut le matériel et le poids, le formel inclut la forme et la couleur, l'agentif inclut le créateur, etc. Dans certaines circonstances, il peut être utile de connaître le matériel utilisé pour la construction d'une maison (bois, pierres, briques et autres), reconnaître sa forme ou savoir qui était le constructeur. Il est clair cependant que ces informations relèvent dans de nombreux cas seulement des CM et non pas des CL (Cf. Peeters (2000) et Taylor (2000)).

On constate également que les valeurs que prennent les matrices ne dépendent d'aucune langue spécifique. Il suffirait de définir la lexie française GARE et cette même définition s'appliquerait à la lexie anglaise RAILWAY STATION, à la lexie allemande BAHNHOF, à la lexie italienne STAZIONE FERROVIALE, etc. Même si cela pouvait être une bonne approximation et même si les dictionnaires standard ne donnent pas une définition plus précise, cela n'est pas suffisant lorsque l'on cherche une réponse à notre question principale.

2.3.2 Connaissances lexicales dans les dictionnaires traditionnels

Comme nous l'avons remarqué plus haut, aucun projet de type lexique génératif n'a abouti à la création d'un dictionnaire avec une large couverture. Tous les dictionnaires traditionnels traitent la polysémie de la même façon : pour chaque vocable, ils énumèrent des sens distincts sans mécanisme génératif. Étant donné qu'il existe des

⁷ Pustejovsky(1995 : 85-86).

dictionnaires avec une couverture très large (plusieurs dizaines de milliers de vocables), on peut se demander dans quelle mesure ils pourraient nous éclairer sur l'importance des CL dans le processus d'interprétation.

Il existe d'assez bons dictionnaires actuellement, tels *Le Robert*, *Le Larousse* ou *Le Trésor de la Langue Française* en français ou bien le *Collins Cobuild English Dictionary*, *Oxford English Dictionary*, *Longman Dictionary*, *Merriam-Webster's Unabridged Dictionary*, etc. en anglais. Nous considérons que ces dictionnaires contiennent une partie des connaissances en question. De nombreuses relations lexicales sont incluses déjà dans la définition de la lexie vedette. Par exemple, un des sens de MAIRIE – selon *Le Nouveau Petit Robert* – est 'administration municipale' : cela indique clairement qu'il y a un lien lexical entre MAIRIE et MUNICIPAL. En outre de la définition, la plupart des dictionnaires contiennent aussi des exemples où l'on peut trouver des lexies, des locutions ou des collocations qui ont un lien lexical avec la lexie vedette. Pour GARE par exemple, *Le Nouveau Petit Robert* mentionne effectivement CONSIGNE, HALL, BUFFET, BUVETTE, QUAI – pour ne citer que cinq lexies. Assez souvent, les dictionnaires contiennent aussi des synonymes, des antonymes, des hyperonymes, etc.

Malheureusement, ces dictionnaires ont aussi de sérieux inconvénients. Comme le remarque Le Pesant (1996), ce sont des dictionnaires non explicites. Ils ont été conçus pour être exploités par des êtres humains possédant la langue dans laquelle ils ont été rédigés. Le cerveau humain a des aptitudes telles que les informations

n'ont pas besoin d'être fournies sous une forme entièrement explicite. Qui plus est, une trop grande explicitation des informations rendrait ces ouvrages difficile à consulter par l'humain.

Tout en reconnaissant que ces dictionnaires puissent avoir une certaine utilité par rapport à notre question, nous croyons qu'ils ne peuvent pas nous donner une réponse précise – et cela pour trois raisons :

- un article de dictionnaire « classique » ne contient pas forcément toutes les informations lexicales relatives à la vedette ;
- l'article peut aussi contenir des informations qui n'ont pas de lien lexical avec la vedette (cf. Wierzbicka (1996 : 262ssq)) ;
- même si de nombreuses informations sont présentes dans un article, elles ne sont pas encodées de façon explicite.

En effet, les auteurs de dictionnaires ne se sont pas donné comme but d'énumérer explicitement dans un article toutes les lexies (et seulement celles-là) qui ont un lien sémantique avec la lexie vedette et d'encoder ces liens.

2.3.3 WordNet

De nos jours, WordNet est devenu la ressource lexicale informatisée la plus utilisée – et probablement même la plus efficace. Nous le présenterons à travers le point de

vue de Fellbaum *et al.* (1998). Ensuite, nous ferons des remarques concernant les liens qu'il couvre et la séparation de différents types de connaissances.

2.3.3.1 Présentation de WordNet

Qu'est-ce WordNet au juste ? Un dictionnaire ? Un thésaurus ? Les dictionnaires contiennent en général des connaissances sur des lexies alors que les encyclopédies contiennent des connaissances sur le monde. Les limites entre les deux ne sont pas toujours faciles à établir. Quant aux thésaurus, ils sont construits autour des concepts et sont destinés à aider l'utilisateur à trouver l'unité lexicale la plus appropriée lorsqu'il a un concept en tête. WordNet n'est ni un dictionnaire classique ni un thésaurus : il est une combinaison des traits de chacune de ces deux ressources lexicales.

D'un certain point de vue, WordNet ressemble à un thésaurus car il est construit autour des ensembles de synonymes (*synset*). Un ensemble de synonymes contient toutes les lexies qui sont des quasi-synonymes ou, au moins, possèdent une intersection de sens importante. Ainsi, l'utilisateur qui a un concept en tête peut trouver, à partir d'une des lexies qui l'expriment, d'autres unités lexicales qui expriment le même concept. D'un autre point de vue, WordNet ressemble à un dictionnaire traditionnel. Par exemple, on y trouve des définitions et des exemples pour presque chaque ensemble de synonymes et, dans certains cas, même des informations morphologiques.

WordNet fait une distinction entre les relations sémantico-conceptuelles et les relations lexicales. Ce sont les ensembles de synonymes – et non pas les lexies – qui sont liés par différentes relations sémantico-conceptuelles : synonymie, hyponymie/hyperonymie, méronymie (relation entre le tout et ses parties), implication (par exemple, ronfler implique dormir), etc. Par contre, l'antonymie est une relation lexicale qui concerne les lexies plus que les concepts.

Les quatre principales catégories syntaxiques (nom, verbe, adjectif, adverbe) sont représentées séparément dans WordNet. La raison en est que les relations entre les lexies et les concepts varient dépendamment des catégories syntaxiques. Par exemple, bien que les noms et les verbes soient organisés hiérarchiquement, les noms sont liés par l'hyponymie/hyperonymie alors que les verbes le sont par la troponymie. Cette dernière se définit de la manière suivante : V1 est un **troponyme** de V2 si et seulement si V2-er signifie V1-er d'une certaine façon. Par exemple, boiter, c'est marcher d'une certaine manière, déguster, c'est manger d'une certaine manière.

2.3.3.2 Remarques sur les informations lexicales de WordNet

L'organisation hiérarchique et l'héritage dans WordNet permettent d'établir certains liens entre les lexies. Bien que ces liens puissent nous être utiles pour l'interprétation du défini, ils risquent d'être insuffisants. En effet, un des problèmes de WordNet est ce que Fellbaum *et al.* (1998) appellent « le problème de tennis » : comment lier entre eux les lexies du même champ sémantique ? Il est clair qu'il

existe des liens par exemple entre TENNIS, RAQUETTE et BALLEs, ou encore entre HÔPITAL et MÉDECIN. Malheureusement, WordNet n'arrive pas à établir ces liens de façon claire et explicite. Un problème similaire pourrait venir du fait que dans WordNet les verbes, les adjectifs, les adverbes et les noms sont traités séparément : ainsi, il n'y aurait pas de relation explicite entre REPAS et MANGER...

Bien entendu, on ne peut pas nier que ces liens sont présents dans WordNet, comme ils sont présents dans de nombreux dictionnaires. En effet, les définitions de WordNet contiennent ces liens, tantôt directement (BALL et RACQUET sont dans la définition de TENNIS), tantôt indirectement (HOSPITAL et DOCTOR pourraient être liées par le fait que tous les deux contiennent MEDICAL dans leur définition) :

RACQUET: a sports implement (usually consisting of a handle and an oval frame with a tightly interlaced network of strings) used to strike a ball (or shuttlecock) in various games

TENNIS: a game played with rackets by two or four players who hit a ball back and forth over a net that divides the court

BALL: round object that is hit or thrown or kicked in games

HOSPITAL: a medical institution where sick or injured people are given medical or surgical care

MEDICAL: relating to the study or practice of medicine

DOCTOR: a licensed medical practitioner

DINNER: the main meal of the day served in the evening or at midday

MEAL: the food served and eaten at one time

EAT1: take in solid food

EAT2: eat a meal; take a meal

Ces définitions indiquent un rapport entre HOSPITAL et DOCTOR, mais ce rapport est loin d'être aussi formel (et calculable) qu'une relation de synonymie, antonymie ou hyperonymie. Les relations qui sont dans WordNet de façon formelle et calculable, telles la synonymie, l'hyperonymie/l'hyponymie ou la méronymie, ne permettent pas de démonter le lien étroit qui existe entre ces lexies.

Notre deuxième série de remarques porte sur une des questions les plus fondamentales pour tous projets semblables : quelles informations inclure dans le lexique ? La plupart des liens que l'on trouve dans WordNet sont considérés par ses auteurs comme des liens conceptuels et non pas comme des liens lexicaux. Il nous semblerait plus exact de dire que ces liens sont – très souvent – à la fois conceptuels et lexicaux. Prenons comme exemple HOUSE et ses hyperonymes et hyponymes. Il est clair qu'il existe *aussi* un lien lexical entre HOUSE et CONSTRUCTION ou encore entre HOUSE et COTTAGE – et non seulement entre les concepts que les occurrences de ces lexies pourraient dénoter. Le problème est que pour WordNet, ces liens sont uniquement conceptuels. En effet, pour WordNet, les liens conceptuels et lexi-

caux sont mutuellement exclusifs : tous les liens qui peuvent être conceptuels ne sont pas considérés comme lexicaux.

3 Cadre linguistique de notre recherche :

la lexicologie explicative et combinatoire

Nous proposons dans ce chapitre une présentation de la Lexicologie explicative et combinatoire (LEC) qui fournit le cadre principal de nos analyses. Notre but n'est pas tant de la décrire de façon exhaustive que d'en présenter les principales caractéristiques qui la rendent particulièrement utile pour notre recherche. Après avoir fait une présentation générale de la LEC, nous nous intéressons à la façon dont elle traite les connaissances lexicales. À la fin du chapitre, nous décrivons le DiCo, la base de connaissances lexicales, qui est l'outil de nos analyses.

Cette présentation ne peut donner qu'un aperçu général. Pour en savoir davantage, on pourra consulter la littérature sur la TST, la LEC et le DiCo donnée dans la bibliographie : Mel'čuk (1993), Mel'čuk (1995), Mel'čuk *et al.* (1984, 1988, 1992, 1999), Mel'čuk, Clas et Polguère (1995), Polguère (1998), Polguère (2000) et Polguère (2003).

3.1 Présentation sommaire

Nous présentons d'abord l'objet d'étude et le but de la lexicologie, tels qu'ils sont définis par la LEC. Ensuite, nous illustrons le caractère rigoureux de la LEC qui en fait le cadre théorique par excellence de nos recherches. Bien entendu, cette présentation reste très sommaire : certains points qui y sont mentionnés seront explicités dans la section (3.2).

3.1.1 Objet et but de la lexicologie explicative et combinatoire

Pour pouvoir décrire le rôle de la lexicologie dans la théorie linguistique que nous utilisons, il faut nous mettre d'accord sur les définitions suivantes :

Un **lexème** est un mot pris dans une acception bien déterminée et munie de tous les renseignements qui spécifient totalement son comportement dans un texte. Exemples : GARE, TRAIN, COURIR, PRÉPARER.

Une **locution** est une expression entièrement idiomatique, c'est-à-dire une expression dont le sens ne peut pas être décrit en fonction du sens d'aucun de ses constituants. Par exemple, le sens de la locution JETER L'ÉPONGE ne contient ni le sens de la lexie JETER ni celui de la lexie ÉPONGE.

Une **lexie** est soit un lexème soit une locution. Elle est une unité trilatérale, elle a un sens, une forme phonique/graphique et un ensemble de propriétés de combinatoires (propriétés qui permettent de combiner cette lexie avec d'autres lexies dans un texte). Exemples : MANGER, GARE MARITIME

Un **vocable** est un ensemble de lexies qui sont associées aux mêmes signifiants et possèdent une intersection de sens significative. Par exemple, le vocable BUREAU contient cinq lexies BUREAU – cf. Mel'čuk et *al.* (1995 : 158).

Le **lexique** d'une langue L est l'ensemble de toutes les lexies de L.

La **lexicologie** est l'étude du lexique. Les lexies y doivent être considérées simultanément sous leur facette sémantique, syntaxique et lexico-combinatoire.

Ces définitions sont à la base non seulement de la LEC mais aussi de la Théorie Sens-Texte (TST). La TST est une théorie linguistique qui, comme toute autre théorie linguistique, étudie et cherche à décrire les langues naturelles. La LEC en est la composante lexicale. Elle peut être considérée comme la composante centrale parce que, suivant la TST, le lexique d'une langue constitue la partie primordiale de cette langue.

Comme l'objet de la LEC est l'étude des lexies des langues naturelles, elle doit aboutir à l'élaboration de dictionnaires. En effet, la lexicologie théorique détermine le contenu et la forme des informations à assigner aux lexies. Une lexie munie de toutes ces données, organisée de façon systématique et cohérente, constitue un article de dictionnaire. Le regroupement des lexies ainsi décrites forme donc un dictionnaire. Le dictionnaire est à la fois l'objet d'étude et le produit final de la lexicologie.

La LEC est explicative parce que, pour chaque lexie, elle permet d'élaborer une explication sémantique formelle. Cette explication inclut non seulement la définition de la lexie en question mais aussi les liens précis qu'elle entretient avec d'autres lexies de la langue. La LEC est combinatoire parce qu'elle accorde une valeur particulière à la présentation rigoureuse et exhaustive de la combinatoire lexicale.

Le produit final dont nous venons de parler est le *Dictionnaire Explicatif et Combinatoire* (DEC), dont quatre volumes sont déjà parus. Il existe aussi deux autres types de dictionnaires en construction selon les principes de la LEC : le *Dictionnaire de Combinatoire* (DiCo) et le *Lexique Actif du Français* (LAF).

3.1.2 Critères méthodologiques explicites de la LEC

L'approche de la LEC nous paraît particulièrement intéressante car elle a des critères méthodologiques explicites et rigoureux qui sont nécessaires pour veiller à ce que le

dictionnaire ne contienne que ce qui est encodé dans la langue et, indirectement, pour distinguer ainsi les CL et les CM. En effet, elle propose des méthodes de description formelle, cohérente, uniforme et exhaustive du lexique, cf. Mel'čuk et *al.* (1995 : 34-43). Il convient d'expliquer brièvement ici ce qu'il faut entendre par ces quatre qualificatifs.

- 1) La description du lexique que fait la LEC est **formelle** car elle est écrite dans un métalangage préétabli et explicite.
- 2) Les articles de dictionnaire élaborés par la LEC sont **cohérents** car le sémantisme, la syntaxe et la cooccurrence lexicale des lexies sont bien mis en correspondance.
- 3) Le traitement des lexies est **uniforme** dans la LEC car la description est similaire pour toutes les unités d'un même champ sémantique.
- 4) La description de la LEC est **exhaustive** car elle présente la totalité des informations lexicographiques pour chaque lexie, y compris les références à toutes les lexies apparentées.

La première tâche de cette description est de délimiter les lexies à l'intérieur d'un vocable polysémique – rappelons que la majorité des vocables sont polysémiques. Malheureusement, on ne peut pas dire que cette tâche soit accomplie de façon rigoureuse dans tous les dictionnaires. Pour y arriver, la LEC propose les critères suivants : le critère d'interprétation multiple, le critère de différence sémantique lo-

cale/globale, le critère de cooccurrence compatible, le critère de cooccurrence différentielle et le critère de dérivation différentielle, cf. Mel'čuk et *al.* (1995 : 61-69).

Le **critère d'interprétation multiple** dit que si pour une lexie potentielle L ayant comme sens '...s1 ou s2...', on peut construire une phrase qui a deux interprétations uniquement à cause de L telles qu'une de ces interprétation fait intervenir le sens 's1' et l'autre fait intervenir 's2', alors L doit être scindée de sorte qu'au lieu d'une lexie L, on a deux lexies L1 ayant comme sens '...s1...' et L2 ayant comme sens '...s2...' – sauf indication contraire du critère de cooccurrence compatible. Par exemple, *Jean leur a proposé un voyage* peut être interprété soit comme *Jean leur a offert un voyage* soit comme *Jean leur a suggéré de faire un voyage* : il faut donc distinguer (au moins) deux lexies PROPOSER.

Le **critère de cooccurrence compatible** dit que si, pour la lexie potentielle L ayant comme sens '...s1 ou s2...', on peut construire une phrase normale à cooccurrence compatible, alors L ne doit pas être scindée en deux mais garder la disjonction dans sa définition. Par exemple, étant donné la phrase *Des avions et des navires bombardent l'usine*, la lexie BOMBARDER doit contenir une disjonction : 'lancer des bombes ou des obus'.

Le **critère de cooccurrence différentielle** dit que si, pour la lexie potentielle L ayant comme sens '... s1 ou s2...', on peut dégager deux ensembles disjoints de cooccurrents morphologiques, syntaxiques ou lexicaux tels que l'un correspond à

‘...s1...’ et l’autre à ‘...s2...’, alors L doit être scindée de sorte qu’au lieu d’avoir L, on a deux lexies L1 et L2. Par exemple, *Il s’occupe de thé* signifie ‘Il travaille dans l’industrie et/ou le commerce de thé’ alors que la phrase *Il s’occupe du thé* signifie ‘Il prend soin du thé’. Autrement dit, il faut distinguer (au moins) deux lexies S’OCCUPER.

Le **critère de dérivation différentielle** dit que si, pour la lexie potentielle L ayant comme sens ‘... s1 ou s2...’, on peut dégager deux ensembles disjoints de dérivés morphologiques ou lexicaux tels que l’un correspond à ‘...s1...’ et l’autre à ‘...s2...’, alors L doit être scindée de sorte qu’au lieu d’avoir L, on a deux lexies L1 et L2. Par exemple, bien que l’on puisse dire *Julie élève des enfants* et *Julie élève des cochons*, dire que *Julie est éleveuse de cochons* est possible alors que *Julie est éleveuse d’enfants* ne l’est pas.

Une fois la lexie délimitée, il faut écrire l’article. Un article de dictionnaire du DEC se compose de trois zones majeures : zone phonologique, zone sémantique et zone de combinatoire. À ces trois zones majeures s’ajoutent deux zones auxiliaires : la zone d’illustration et la zone phraséologique.

L’élément le plus important d’un article de dictionnaire se trouve sans nul doute dans la zone sémantique : c’est la définition lexicographique. La LEC propose cinq règles qui déterminent le métalangage de la définition lexicographique, cf. Mel’čuk et al. (1995 : 78-95).

- 1) La **règle de la forme propositionnelle** dit que pour une lexie prédicative L, le défini⁸ doit être une forme propositionnelle où les variables représentent les actants sémantiques de L : *X reproche Y à Z* pour REPROCHER, *X est le père de Y* pour PÈRE, etc.
- 2) La **règle de décomposition** dit que la définition d'une lexie L doit être faite par des lexies qui sont chacune sémantiquement plus simple que L. Cette règle vise à éviter la circularité, les cercles vicieux dans les définitions. Par exemple, si la définition de CUISINIER inclut 'cuisiner' alors la définition de CUISINER ne peut pas inclure 'cuisinier'.
- 3) La **règle du bloc maximal** dit que la définition d'une lexie L doit être faite au moyen de lexies telles qu'aucune configuration de ces lexies ne peut être remplacée par une seule lexie sémantiquement équivalente. En d'autres termes, les définitions doivent être les moins décomposées possibles.
- 4) La **règle de standardisation** dit que la définition d'une lexie doit être faite de façon à éviter dans le définissant les termes ambigus et les termes synonymes. Pour ce faire, le définissant est décrit dans un métalangage qui, au besoin, peut s'écarter de la langue standard. Par exemple, le DEC utilise les exposants numériques pour distinguer les lexies homonymes : VOLER¹ et VOLER².
- 5) La **règle de substituabilité** dit qu'une lexie et sa définition doivent être réciproquement substituables dans tous les énoncés, sans modification du sens exprimé.

⁸ Une définition est composée de deux parties : le défini et le définissant.

Le but de ces règles est de s'assurer que la définition de la lexie L est formellement correcte. Bien entendu, ceci est une condition nécessaire pour arriver à une définition précise mais elle n'est pas suffisante : il faut aussi que la définition de L soit conforme aux faits de la langue dont L fait partie. Pour cette raison, la LEC propose quatre critères pour la formulation de la définition lexicographique, cf. Mel'čuk et *al.* (1995 : 95-102).

1) Le **critère de pertinence linguistique** dit que la définition de la lexie L1 doit inclure la composante 'c' si et seulement si on peut trouver dans la langue étudiée une autre lexie L2 qui est formellement liée à L1 et dont la définition comprend 'c' de façon indiscutable. Ce critère sera présenté plus en détail dans (3.2.3.3).

2) Le **critère de cooccurrence avec modificateurs qualificatifs** dit que la définition de L doit refléter explicitement la cooccurrence de L avec les modificateurs qualitatifs. Autrement dit, la définition doit comporter une composante sémantique qui est en mesure d'accepter le sens des modificateurs en question. Par exemple, étant donné que l'on peut dire *applaudissements nourris*, la définition de la lexie APPLAUDISSEMENT doit comporter une composante que l'adjectif NOURRI peut intensifier – cette composante sera '...la force ou la fréquence...'

3) Le **critère de cooccurrence avec quantificateurs** dit que la définition de L doit refléter explicitement la cooccurrence de L avec les quantificateurs. Il s'agit surtout des numéraux si L est un nom et de l'expression *n fois* si L est un verbe. Par exem-

ple, la définition de AIL et OIGNON doit rendre compte du fait que l'on peut dire *Apporte-moi un oignon* mais pas **Apporte-moi un ail*.

4) Le **critère de cooccurrence avec négation** dit que la définition de L doit refléter explicitement la façon dont L se combine avec la négation. Par exemple, en disant *Isabelle n'est pas veuve*, on ne nie pas qu'Isabelle est une femme en âge d'être mariée – par conséquent, la composante 'femme en âge d'être mariée' ne fait pas partie du corps de la définition de VEUVE mais elle est une contrainte sur une variable dans la définition.

Une **définition lexicographique** est une paraphrase de la lexie et elle a la même valeur syntaxique. Elle se présente comme une expression homogène mais elle comprend des composantes aux propriétés et comportements fort différents. Pour rendre compte de ces différences, la LEC distingue six types de composantes sémantiques dans la définition d'une lexie L : composante ordinaire, composante générique (=le genre prochain), composante faible (=qui peut s'estomper dans des conditions précises), composante optionnelle (=qui peut être neutralisée par un contexte contradictoire explicite), contraintes sur les variables (=qui caractérisent les actants sémantiques) et partie présuppositionnelle (=qui n'est pas affectée par la négation).

La définition n'est cependant pas le seul élément de la zone sémantique : on y trouve également la connotation lexicographique. La LEC définit la **connotation lexicographique** comme une caractéristique sémantique que la lexie L attribue à son référent et qui n'appartient pas à la définition lexicographique de L. Il est évident que la

connotation et la définition d'une lexie sont distinctes et pourtant proches l'une de l'autre – d'où l'importance de les distinguer. La LEC propose deux critères pour décider si un élément sémantique est une composante de la définition de la lexie L ou s'il est une des connotations de L : critère de sens antonymique et critère de « mauvais état ».

Nous pouvons illustrer ces deux critères en nous inspirant des exemples proposés par Mel'čuk et *al.* (1995 : 111-117). Imaginons qu'un lexicographe veuille élaborer le sens de la lexie L = VENT (*le vent souffle*) et que son intuition lui dit qu'il existe un lien entre cette lexie et le sens S = 'changeant (de direction)'. En effet, ce sens S est inclus dans les locutions LE VENT TOURNE, ALLER OU LE VENT SOUFFLE, TOURNER À TOUT VENT, etc. La question qu'il se pose donc est de savoir s'il faut inclure ce S dans la définition de VENT ou si c'est plutôt une connotation de VENT. Le **critère de sens antonymique** dit que si l'adjonction d'un modificateur à L qui contient la négation de S crée une contradiction logique, alors S est une composante de la définition de L, sinon, S est une connotation de L. Dans notre exemple, l'expression *un vent constant* ne crée pas de contradiction : cela signifie que 'changeant (de direction)' est une connotation de VENT, ne doit pas être inclus dans la définition de VENT.

Le critère de « mauvais état » peut être illustré par les lexies TÊTEI.1a ('partie supérieure du corps d'une personne...') et CŒURI.1a ('organe principal de la circulation sanguine d'une personne...'). En français, on associe à la première lexie le sens

‘organe de la raison’ et à la deuxième le sens ‘organe des sentiments’. La question que le lexicographe doit se poser est la suivante : ces sens appartiennent-ils à la définition des lexies en question ou à leur connotation ? Le critère de « mauvais état » permet de décider. Considérons en effet les deux exemples suivants :

(1a) *J’ai une douleur à la tête, je n’arrive plus à réfléchir.*

(1b) **J’ai une douleur au cœur, je ne peux plus aimer.*

La première phrase est toute naturelle : des maux de tête peuvent empêcher la réflexion. La deuxième ne l’est pas : le mauvais état du cœur n’empêche personne d’aimer. Nous pouvons donc conclure, grâce à ce critère, que la définition de TÊTE doit inclure le sens ‘organe de la raison’, cela fait partie de sa définition. Par contre, la définition de CŒUR n’inclut pas le sens ‘organe de l’amour’, cela fait partie plutôt de sa connotation.

La zone de combinatoire est également élaborée de façon rigoureuse. Elle a deux parties principales, consacrées respectivement à la combinatoire syntaxique et à la combinatoire lexicale. La combinatoire lexicale de la lexie inclut les **fonctions lexicales** (FL) qui en décrivent les actants, les dérivations sémantiques et les collocations. On peut distinguer 56 fonctions lexicales standard simples : elles couvrent, dans toutes les langues, les relations paradigmatiques (synonymie, antonymie, générique...) entre les lexies ainsi que leurs relations syntagmatiques (intensificateurs, verbes phasiques...). Le lecteur peut trouver une présentation détaillée des FL dans

Mel'čuk *et al.* (1995 : 125-152). Les FL nous seront très utiles notamment pour établir des liens lexicaux lors des analyses des expressions définies.

3.2 Identification des connaissances encodées dans le

lexique

Après une introduction sommaire à la problématique d'encodage des CM, nous présentons dans cette section une méthodologie qui permet d'établir les CM encodées dans la langue.

3.2.1 Introduction à la problématique d'encodage des CM

Comment la rigueur de la LEC, ses principes descriptifs et sa méthodologie que nous venons de présenter nous aident-ils à discerner les connaissances lexicales utilisées dans le processus d'interprétation des énoncés ? Comment décide-t-elle si une connaissance donnée est linguistique – et plus particulièrement lexicale – ou non ? Quelles sont les procédures à appliquer, les tests à faire ? Comment ces choix sont-ils motivés ? Pour donner des exemples concrets, nous savons que, typiquement, la neige est blanche, que l'on trouve des tables dans un restaurant ou encore que les tables ont souvent quatre pieds. Comment la LEC décide-t-elle si ces connaissances sont encodées dans la langue ou pas ? Dire qu'**une connaissance X est encodée**

dans une langue Y signifie pour nous qu'il existe un parallèle entre X et une connaissance (règle) qui fait partie de Y, en tant que système. Par exemple, c'est une CM que de savoir qu'une fourchette est un ustensile. Par contre, savoir que la définition de FOURCHETTE inclut 'ustensile' (ou encore, savoir que USTENSILE est un hyperonyme de FOURCHETTE) fait partie de nos connaissances du français – c'est une CL. Il est clair que ces deux connaissances, bien qu'elles soient de types différents, peuvent être mises en parallèle : cette CM est encodée en français.

Le but de cette section est de montrer qu'une quantité immense de nos connaissances du monde est encodée lexicalement. Pour ce faire, nous proposons une grille d'analyse en trois étapes qui permet de justifier l'inclusion de ces connaissances dans le dictionnaire.

- 1) La première question à poser est celle de la lexicalisation : quelles sont les lexies qui peuvent dénoter nos connaissances dans un certain domaine ? Comment les délimiter ?⁹
- 2) Ensuite, il faut examiner la définition des lexies, élaborée en fonction du genre prochain, des différences spécifiques et de la pertinence linguistique. Cette définition encode une bonne partie de nos connaissances sur ce que la lexie dénote.

⁹ Les critères de la LEC pour identifier et délimiter les lexies sont avant tout sémantiques. Pour une approche morphologique, voir Fradin (1996).

3) La dernière partie de cette analyse concerne les liens lexicaux qu'entretiennent les lexies les unes avec les autres : ces liens encodent également une partie considérable de nos connaissances du monde.

Il est important de souligner que les connaissances lexicales dont nous parlons ici sont, la plupart du temps (mais pas toujours), des connaissances du monde encodées dans le lexique. L'encodage linguistique de certaines connaissances est conceptuellement prévisible : par exemple, quiconque sait ce que c'est qu'un restaurant s'attendra, à juste titre, à ce que la définition de la lexie RESTAURANT contienne le sens 'manger'. Bien entendu, cela ne signifie pas que l'on y trouvera forcément la lexie MANGER : cette définition peut aussi contenir une autre lexie (par exemple REPAS) qui inclut ce sens. Il existe aussi des connaissances dont l'encodage linguistique est conceptuellement imprévisible : par exemple, pour parler d'un remerciement dont le degré est élevé, on peut dire des phrases comme *Je vous remercie infiniment* ou *Je l'ai remercié chaleureusement* mais les phrases **Je l'ai remercié infiniment* ou **Je vous remercie chaleureusement* seraient bizarres en français. Finalement, il arrive aussi que certaines connaissances linguistiques n'aient pas leur équivalent dans nos connaissances du monde : on peut penser par exemple au genre grammatical des noms français qui ne dénotent pas des entités sexuées (féminin pour TABLE, masculin pour BUREAU).

Pourquoi telle ou telle partie de nos connaissances sont encodées dans la langue et pas d'autres ? Dans quelle mesure ce qu'est le monde actuellement (ou ce qu'il a été

auparavant) justifie les choix d'encodage et dans quelle mesure c'est arbitraire ? Faute de place, nous ne pouvons pas examiner ces questions dans ce travail. Ce qui est important pour nous, c'est la rétroaction de la langue sur nos connaissances : parce que telle ou telle connaissance est encodée dans la langue, elle peut devenir plus importante, plus présente dans notre esprit. Par exemple, nous savons que, typiquement, dans un restaurant, il y a des employés qui font le service. Cette connaissance est également encodée dans la langue à travers les lexies RESTAURANT et SERVEUR et les liens lexicaux qui les unissent : notamment, RESTAURANT est un actant¹⁰ de SERVEUR – et SERVEUR est aussi un actant de RESTAURANT. D'un point de vue conceptuel, la situation est semblable dans le cas d'un guichet d'informations/renseignements : typiquement, on y trouve aussi un employé dont le travail consiste à donner des renseignements mais il n'existe pas de lexie en français pour désigner cet employé. L'encodage lexical d'une connaissance peut fortifier les connaissances du monde au sujet de leur référent.

3.2.2 Connaissances lexicalisées

La première étape de notre grille d'analyse est de déterminer tous les concepts, toutes les entités qui sont **lexicalisés** dans un certain domaine : il nous faut identifier toutes les lexies qui dénotent les éléments de ce domaine.

¹⁰ La notion d'actant sera présentée en détails dans (3.2.4.2).

Regardons de près un exemple concret, celui de l'école. Lorsque l'on parle de l'école, on peut tout de suite penser à des lexies comme BULLETIN [de note], CANTINE, CLASSE, COLLE, COURS, DIPLÔME, DIPLÔMÉ, DIRECTEUR, ÉCOLIER, ÉLÈVE, ENSEIGNER, ENSEIGNANT, FINISSANT, INSTITUTEUR, MAÎTRE, NOTE, PRINCIPAL, PROVISEUR, RENTRÉE, SCOLAIRE, SÉCHER [l'école], RETENUE, etc.

Quelle est l'importance de toutes ces lexies énumérées plus haut par rapport à notre sujet ? Nous savons par exemple qu'il y a des personnes qui étudient dans des écoles et ces personnes ont un nom : ÉCOLIER ou ÉLÈVE. L'existence de ces lexies montre clairement que cette connaissance est encodée en français.

Aux yeux de certains lecteurs, il pourrait paraître trivial qu'il y ait un nom spécifique pour désigner les personnes qui étudient dans des écoles. Or, dans une optique saussurienne, cela n'a rien d'évident : dans cette optique, la langue est considérée comme un système relativement autonome et arbitraire – et non pas comme un simple reflet, un simple encodage du monde ou de nos CM. Bien entendu, nous défendons la position que la langue encode une grande partie de nos connaissances du monde mais elle le fait de façon non prévisible à partir de nos CM.

Nous pouvons illustrer cet encodage non prévisible en comparant l'école et le parking. Nous savons que dans une école, il y a des élèves et nous savons aussi que dans un parking, il y a des voitures. Ces deux connaissances sont conceptuellement

comparables mais elles ne sont pas encodées de la même façon en français. La première est lexicalisée (c'est-à-dire, il existe un nom typique pour dénoter spécifiquement les personnes qui étudient à l'école) alors que la seconde ne l'est pas. Autrement dit, le fait qu'il n'existe pas de nom spécifique pour les voitures qui sont stationnées dans un parking montre clairement que l'existence des lexies ÉCOLIER et ÉLÈVE n'est pas une nécessité absolue non plus.

Bien entendu, le fait qu'il n'y ait pas un nom spécifique pour les voitures qui sont stationnées dans un parking ne signifie pas qu'aucune information à ce sujet n'est encodée dans la langue. Il est clair que la définition de la lexie PARKING doit contenir des éléments relatifs au stationnement des voitures. Nous arrivons ainsi à la deuxième étape de notre grille d'analyse qui consiste à examiner les définitions des lexies concernées.

3.2.3 Connaissances encodées dans la définition lexicale

Nous avons mentionné plus haut combien il est difficile – voire impossible – de délimiter le contenu des concepts. Qu'en est-il des lexies ? Comment distinguer entre, d'une part, ce que nous savons sur l'entité ou le fait dénoté par une lexie et, d'autre part, ce que la définition de cette lexie doit inclure ? La LEC a-t-elle des critères suffisamment rigoureux qui nous permettent de dire que tel ou tel élément fait partie de la définition de la lexie et n'est pas seulement un élément de nos connaissances du monde ? Ou bien, doit-on se limiter à une définition plus ou moins intuitive ?

L'élaboration des définitions est une tâche ardue. En effet, quiconque compare deux dictionnaires peut constater que sous une même entrée, le nombre d'acceptations peut varier d'un dictionnaire à l'autre : par exemple, ÉCOLE a 5 acceptations selon *Le Nouveau Petit Robert* alors qu'elle en a 8 selon *Le Petit Larousse*. Qui plus est, dans les cas où l'on trouve le même nombre d'acceptations, les définitions ne sont pas identiques – même si elles sont souvent semblables.

Malgré toutes les difficultés de l'élaboration des définitions des lexies, nous croyons que la LEC peut accomplir cette tâche de façon rigoureuse, suivant des critères précis. Nous en examinons ici les trois étapes qui ont le plus de rapport avec l'encodage des CM : déterminer le genre prochain, déterminer les différences spécifiques et prendre en considération la pertinence linguistique. La définition hypéronymiques (genre prochain et différence spécifique) est souvent privilégiée en lexicographie (cf. Henry (1996 : 24)) mais la pertinence linguistique n'a pas reçu la place qu'elle mériterait.

3.2.3.1 Détermination du genre prochain

Lorsque l'on cherche à représenter les connaissances du monde selon une approche IA, une des relations les plus fondamentales entre les concepts, les objets de ce monde est la relation appelée IS_A qui indique l'appartenance d'un élément à un ensemble. Elle indique par exemple qu'un restaurant est un établissement, un magazine est une publication, etc. Or, quiconque pense aux définitions lexicographiques,

se rend immédiatement compte que cette relation a son équivalent linguistique qui encode des connaissances sinon identiques au moins en étroite relation : le genre prochain.

Le genre prochain est la composante centrale de la définition de la lexie : il en est la paraphrase minimale, l'hyperonyme le plus proche. Par exemple, *Le Nouveau Petit Robert* définit RESTAURANT comme 'établissement où l'on sert des repas moyennant paiement' et MAGAZINE comme 'publication périodique, généralement illustrée'. Suivant ces définitions, 'établissement' est le genre prochain de RESTAURANT et 'publication' est celui de MAGAZINE. Cela signifie que nos connaissances du monde qui indiquent qu'un magazine est une publication ou un restaurant est un établissement sont encodées dans la langue à travers le genre prochain de MAGAZINE et RESTAURANT.

3.2.3.2 Détermination des différences spécifiques

Il est clair que nos connaissances du monde ne se limitent pas à savoir qu'un restaurant est un établissement ou qu'un magazine est une publication. Parmi tout ce que nous savons à leur sujet, que faut-il encore inclure dans leur définition ? Malgré toute l'importance du genre prochain, il ne peut pas suffire à lui seul pour définir une lexie. Dans le cas de deux lexies qui ne sont pas des synonymes absolus (phénomène extrêmement rare dans des langues naturelles, rappelons-le !), la définition de ces deux lexies doit refléter cette différence. C'est pour cette raison que l'on doit

inclure également dans la définition un ensemble de composantes que l'on appelle les différences spécifiques : elles permettent, d'une part, de faire une distinction sémantique entre les lexies qui ont le même genre prochain, d'autre part, de caractériser la combinatoire sémantique d'une lexie.

Revenons à notre exemple concernant les magazines. Nous possédons de nombreuses connaissances à leur sujet mais que faut-il inclure dans la définition de la lexie MAGAZINE ? Nous avons déjà identifié son genre prochain : 'publication'. Un livre, par exemple, est également une publication : pour distinguer entre MAGAZINE et LIVRE, la définition de ce premier doit inclure 'périodique'. Toutefois, définir MAGAZINE comme 'publication périodique' ne la distinguerait pas de BULLETIN qui peut également avoir un sens qui inclut 'publication périodique'. La différence entre MAGAZINE et BULLETIN est qu'un magazine est généralement illustré alors que la présence d'illustrations n'est pas caractéristique du bulletin. Nous sommes ainsi arrivés à la définition 'publication périodique, généralement illustrée' que propose *Le Nouveau Petit Robert* pour MAGAZINE.

Cette définition n'est cependant pas encore satisfaisante puisque nous savons aussi qu'il faut distinguer entre magazines, d'une part, et revues et journaux illustrés d'autre part. La différence entre magazines et journaux – mêmes illustrés – réside dans le fait que les pages d'un magazine sont attachées les unes aux autres. Qu'en est-il de la différence entre MAGAZINE et REVUE ? Elle semble résider dans le fait que le sens de MAGAZINE inclut aussi 'divertir'. En effet, il est naturel de dire

revue mathématique mais il serait bizarre d'utiliser l'expression ?*magazine mathématique*.

Récapitulons l'exemple du magazine : parmi toutes les connaissances que nous en possédons, quelles sont celles qui sont linguistiques ? Plus précisément, lesquelles doivent être encodées dans sa définition, comment définir la lexie MAGAZINE ? Afin de la différencier d'autres lexies comme LIVRE, BULLETIN, JOURNAL et REVUE, nous sommes amené à la définir comme 'publication périodique, sous forme d'un cahier, généralement illustrée, destinée à divertir'. Pour illustrer l'intérêt qu'une telle définition peut avoir pour nous, nous pouvons tout simplement penser à des énoncés comme (2) :

(2) *Samedi après-midi, j'ai passé une heure entière à lire des magazines politiques. Ce petit divertissement m'a fait beaucoup de bien.*

Les humains ont à leur disposition à la fois des CM et des CL pour savoir que la lecture des magazines peut être un divertissement. Toutefois, ce qu'il faut remarquer ici est que des machines qui seraient dotées des CL mais pas des CM, pourraient également le savoir, notamment grâce à la définition de MAGAZINE – étant donné que les définitions font partie de nos CL.

3.2.3.3 Prise en considération de la pertinence linguistique

Nous venons de voir qu'une partie de nos connaissances du monde doit également être encodée dans la définition d'une lexie afin de pouvoir la distinguer d'autres lexies sémantiquement proches. Cependant, la définition d'une lexie ne doit pas se limiter à des composantes qui permettent cette distinction : elle doit aussi inclure tout ce qui est pertinent d'un point de vue linguistique – comme le stipule le critère de la pertinence linguistique mentionné dans (3.1.2). Or, ces connaissances linguistiquement pertinentes sont les correspondantes d'une partie de nos CM – comme le sont le genre prochain et les différences spécifiques. Nous pouvons illustrer cela à travers deux exemples : celui de NUAGE et celui de NEIGE, tirés de Mel'čuk et *al.* (1995).

Nous savons par exemple, que les nuages peuvent cacher le ciel. Est-ce que cette connaissance a quelque chose à voir avec nos connaissances du français ? Plus particulièrement la définition de la lexie NUAGE dans le sens de 'amas de vapeur...' doit-elle inclure la composante '...qui cache (partiellement) le ciel...' ? Nous pensons que oui parce qu'il existe une autre lexie NUAGE dont le sens est 'sentiment ou événement déplaisant... qui obscurcit le fait X [comme un nuage cache (partiellement) le ciel...]'. C'est en effet cette deuxième lexie NUAGE que l'on trouve dans des énoncés comme (3) :

(3) *Au début, leur bonheur était sans nuage.*

L'inclusion de la composante '...qui cache (partiellement) le ciel...' permet d'expliciter le lien métaphorique qui existe entre les deux lexies NUAGE.

Nous savons également que la neige est blanche. S'agit-il d'une connaissance linguistiquement motivée ou d'une simple connaissance du monde ? Il est vrai qu'il n'est pas nécessaire d'inclure la couleur blanche dans la définition de NEIGE (dans le sens de 'eau congelée...') pour la distinguer d'autres lexies dénotant des substances du même type. Toutefois, si l'on veut rendre compte de l'existence de l'expression semi-figée *blanc comme neige* ou de celle d'une autre lexie NEIGE dans le sens 'cocaïne' (qui est en lien métaphorique avec la première lexie NEIGE), on doit inclure la couleur blanche dans la définition. Il est important de remarquer que, contrairement à NEIGE, la définition des lexies SUCRE et RIZ ne mentionne pas la blancheur, même si les substances qu'elles dénotent sont, typiquement, blanches.

3.2.4 Connaissances encodées à travers les liens lexicaux

Nous avons vu qu'un grand nombre de connaissances du monde a son équivalent encodé linguistiquement à l'intérieur de la définition des lexies. Dans ce qui suit, nous montrons qu'il y a aussi des connaissances du monde qui sont encodées dans la langue à travers les liens que les lexies ont les unes avec les autres.

3.2.4.1 Intérêt des liens lexicaux pour l'encodage des connaissances

Nous avons parlé plus haut de la relation conceptuelle IS_A et de son équivalent linguistique qui est l'hyponymie. Nos connaissances du monde nous indiquent par exemple qu'un chalet est une sorte de maison et nos connaissances de la langue indiquent que la lexie MAISON est l'hyperonyme de la lexie CHALET. Bien entendu, à chaque relation hyperonymique, on peut associer une relation hyponymique : en occurrence, CHALET est hyponyme de MAISON.

Les exemples de liens lexicaux que nous venons de donner pourraient paraître conceptuellement prévisibles à certains lecteurs. Il est important de souligner que nos connaissances du monde ne nous permettent pas de prévoir ce qui sera encodé dans la langue et surtout comment cela sera encodé. Nous savons par exemple qu'il y a des personnes qui vont dans des restaurants pour manger, pour prendre des repas. D'un point de vue conceptuel, il n'y aurait pas de raison de ne pas pouvoir les appeler des *mangeurs* ou des *preneurs de repas*. Toutefois, de telles appellations seraient insolites en français même si on pourrait les comprendre dans certains contextes : on s'attend à parler des *clients* plutôt (bien que cela ne désigne pas spécifiquement les clients d'un restaurant). La situation est semblable dans le cas de ceux qui empruntent des livres d'une bibliothèque : on les appelle des *usagers* (ou les *lecteurs*) de la bibliothèque et non pas les *emprunteurs*. Pourtant, on appelle bien des *conducteurs* ceux qui conduisent une voiture !

Nous pouvons nous attendre à ce que certains types de connaissances du monde soient encodés dans une langue mais il nous est impossible de prévoir de quelle manière cela se fera sans connaître la langue en question. Nous savons par exemple que dans un bateau, dans un royaume, dans une école, dans une faculté, dans une gare, dans une bibliothèque, etc. il y a une personne qui en est responsable mais seulement nos connaissances de la langue nous permettent de savoir que cette personne s'appelle respectivement CAPITAINE, ROI, DIRECTEUR/PROVISEUR, DOYEN, CHEF DE GARE, etc. en français. Autrement dit, la connaissance qu'il existe un responsable pour chacun des lieux cités est encodée à travers les liens lexicaux qui existent entre BATEAU et CAPITAINE, ROYAUME et ROI, FACULTÉ et DOYEN, etc.

Il est important de souligner le caractère linguistique de ces liens : comme nous l'avons dit, ils ne sont pas toujours prévisibles d'un point de vue purement conceptuel. Nous avons beau avoir des connaissances sur ces lieux et leur responsable, lorsque nous apprenons une langue nouvelle, nous devons aussi apprendre les lexies correspondantes si celles-ci existent.

3.2.4.2 Positions actancielles d'une lexie et participants d'une situation

Nous proposons ici une présentation succincte de ce que la TST appelle les actants et les positions actancielles. Bien qu'il n'y ait pas de correspondance exacte entre les positions actancielles d'une lexie et les participants de la situation que cette lexie

peut dénoter, de nombreuses connaissances du monde concernant ces participants peuvent être encodées dans la langue à travers les positions actancielles.

Le lecteur trouvera une description détaillée de la notion d'actant dans Mel'čuk (2003) et Mel'čuk (à paraître). Dans ce qui suit, nous en présenterons les grandes lignes. Pour faire cette présentation, il est important de préciser d'abord ce qu'il faut entendre par situation linguistique, d'une part, et par les participants de cette situation linguistique, d'autre part.

Une **situation linguistique** est une situation décrite par la langue (souvent par une unité lexicale qui a un sens prédicatif). Nous entendons *situation décrite* au sens large : à proprement parler, une lexie comme PROFESSEUR peut servir à dénoter un individu plutôt qu'une situation mais, d'une manière ou d'une autre, cet individu doit être lié à une situation et cette situation a des participants (quelqu'un enseigne quelque chose à quelqu'un). Une situation linguistique est distincte d'une situation du monde réelle. Nous pouvons illustrer cette notion à travers l'exemple de la situation linguistique que dénote le verbe VENDRE. Cette situation linguistique met en jeu une personne (le vendeur = l'ancien propriétaire) qui transmet quelque chose (une marchandise) à une autre personne (l'acheteur = le nouveau propriétaire) contre une somme d'argent (le montant de la transaction). Les éléments vendeur, acheteur, marchandise et montant sont nécessaires à cette situation : ils sont des participants obligatoires de cette situation. Bien entendu, un participant obligatoire n'est pas forcément une personne ou un objet : il peut également être un processus, une action,

etc. De façon plus générale, nous pouvons dire qu'un élément E de la situation S dénotée par la lexie L est un **participant obligatoire** si et seulement si sans E la lexie L ne pourrait plus dénoter S. Par exemple, en enlevant la somme d'argent de cette situation, on obtiendrait une autre situation qui serait décrite non pas par VENDRE mais par DONNER. La lexie LIRE dénote une situation avec deux participants obligatoires : une personne X lit un texte Y. On pourrait facilement inventer une situation où il y aurait un troisième participant : *Chaque après-midi, Luc lit des journaux à un monsieur aveugle*. Il est clair que le troisième participant (*un monsieur aveugle*) n'est pas obligatoire : en l'enlevant, on aurait toujours une situation (*Chaque après-midi, Luc lit des journaux.*) que LIRE pourrait dénoter. Par contre, si l'on enlevait la personne X (qui lit), il ne resterait plus de situation, il resterait seulement un texte Y.

Dans certaines situations, nous pouvons rencontrer des participants qui ne sont pas obligatoires. Par exemple, dans la situation prototypique dénotée par le verbe BOIRE, il est également question d'un contenant : on boit dans un verre, dans une tasse, à la bouteille, etc. Il est clair cependant que le contenant n'est pas un participant obligatoire : on peut également boire directement d'une fontaine, d'une rivière, etc. Les participants typiques mais non obligatoires d'une situation (comme le contenant dans la situation dénotée par BOIRE) sont appelés des **participants optionnels**.

Les participants de la situation linguistique dénotée par la lexie L sont, la plupart du temps, encodés dans la définition même de la lexie L à travers les positions actanciennes. Les **positions actanciennes** de L sont des places vides dans la définition de L et qui, dans une phrase, sont occupées par des actants. Elles sont identifiées par des variables (X, Y, Z, U, V et W), comme l'illustre la définition suivante :

(4) *X vend Y à Z pour W* = X transmet à Z les droits de possession complète et permanente de Y en échange d'une somme d'argent W que X demande à Z pour Y.

Nous tenons à mentionner ici que les actants sémantiques d'une lexie ne sont pas ses seuls actants : elle a également des actants syntaxiques profonds et des actants syntaxiques de surface. En effet, dans la description d'une langue, les linguistes distinguent souvent plusieurs niveaux : sémantique, syntaxique, morphologique et phonétique/phonologique. La TST en distingue sept : sémantique, syntaxique profond, syntaxique de surface, morphologique profond, morphologique de surface, phonologique profond et phonologique de surface. Bien entendu, les actants sémantiques font partie du niveau sémantique, les actants syntaxiques profonds font partie du niveau syntaxique profond et les actants syntaxiques de surface font partie du niveau syntaxique de surface. Dans les cas non marqués, il existe une correspondance entre ces trois types d'actants. Toutefois, on trouve aussi des cas spécifiques où il y a un décalage entre ces trois niveaux d'actants – cf. Mel'čuk (2003 : 11-13).

Les actants sémantiques d'une lexie sont spécifiés dans la définition lexicographique de la lexie, ses actants syntaxiques profonds et ses actants syntaxiques de surface sont spécifiés par son tableau de régime. Le **tableau de régime** d'une lexie décrit de façon formelle la correspondance entre les actants sémantiques, les actants syntaxiques profonds et les actants syntaxiques de surface. Les circonstanciels de temps, de lieu, etc. sont des entités non spécifiées, ni par la définition, ni par le tableau de régime. En annexe, nous présentons quatre articles du DiCo où le lecteur peut trouver des exemples de tableaux de régime. Nous en citons un dans (5), celui de GARE, précédé par la forme propositionnelle` :

(5)

f_p :

~ DU lieu Y [POUR LE train X]

tr :

X = I = ---

Y = II = de N, A-poss, Adj | Adj = A0(N)

La forme propositionnelle dit que GARE a deux actants, le premier (X) étant TRAIN, le second (Y) étant LIEU. Le tableau de régime ne contient pas d'information particulière sur X mais dit que l'actant Y peut être réalisé de trois façons au niveau syntaxique de surface : soit comme un nom introduit par la préposition DE (*la gare de Dijon*), soit comme un adjectif possessif (*sa gare*), soit comme un adjectif qui est un dérivé nominal (*la gare dijonnaise*).

Nous avons mentionné que, la plupart du temps, il y a une correspondance entre les actants d'une lexie et les participants de la situation dénotée par cette lexie. Un actant sémantique est nécessairement un participant de la situation linguistique mais un participant n'est pas forcément un actant. En effet, il arrive parfois qu'un participant de la situation dénotée par une lexie L ne soit pas **exprimable auprès de L**. La lexie COÛTER en est une bonne illustration. Il est clair que la situation linguistique dénotée par cette lexie est la même que celle dénotée par PAYER et elle implique quatre participants : l'acheteur, le vendeur, la marchandise et le prix. On peut les exprimer tous les quatre avec PAYER (*Julie a payé quarante dollars à Marie pour ce manteau*) mais on ne peut en exprimer que trois avec COÛTER (*Ce manteau a coûté quarante dollars à Julie*). Ainsi, bien que la situation dénotée par COÛTER ait quatre participants, la lexie COÛTER n'aura que trois positions actanciennes. Bien entendu, dans la définition de COÛTER, on trouvera aussi le vendeur : il ne sera pas un actant (une variable actancielle) mais une variable sémantique. Le lecteur peut d'ailleurs remarquer un certain parallélisme entre les participants non exprimables d'une lexie L suivant la LEC et les arguments intégrés d'une lexie L suivant le Lexique Génératif (voir 2.3.1) : ils ne sont pas, en général, exprimables auprès de la lexie L.

Il faut préciser que la contrainte d'être exprimable auprès de L signifie bien plus que la simple présence dans la même phrase que L. Cette contrainte inclut soit une relation syntaxique forte (notamment une relation de sujet, COD ou COI), soit une rela-

tion paradigmatique ou la lexie exprimable auprès de L est un dérivé nominal de L, cf. Mel'čuk (2003). Considérons les exemples suivants :

(6) *Cet ordinateur coûte très cher, il faut être un bon vendeur pour le vendre.*

(7) *Ce DVD coûte 19\$ chez Renaud-Bray.*

Dans le premier énoncé, *vendeur* est syntaxiquement indépendant de *coûter*. Dans le deuxième, même si *Renaud-Bray* exprime le vendeur, il est – d'un point de vue syntaxique – un circonstanciel de lieu.

Maintenant, nous pouvons donner une définition plus formelle de ce qu'il faut entendre par actant sémantique. X est un **actant sémantique** de la lexie L si et seulement si les deux conditions suivantes sont satisfaites : X correspond à un participant de la situation dénotée par L et X est exprimable dans le texte auprès de L. X sera un actant obligatoire s'il correspond à un participant obligatoire et il sera un actant optionnel s'il correspond à un participant optionnel.

Deux remarques s'imposent concernant cette définition :

a) Les variables qui encodent les positions actancielles sémantiques peuvent être typées, c'est à dire sémantiquement contraintes. Dans la définition de VENDRE, donnée plus haut dans (3), seule la variable W était typée (somme d'argent W). Il

aurait été possible, mais pas nécessaire, de préciser aussi le type des autres actants : X et Z sont des personnes, Y est une marchandise.

b) Il arrive que deux (trois) participants d'une situation correspondent à un seul actant. Dans ce cas, la variable dénotant cet actant est scindée en deux (trois). Par exemple, la situation dénotée par FUMER a trois actants: celui qui fume, une substance qui brûle et un artefact dans lequel la substance brûle mais il n'a que deux actants dont un scindé qui inclut à la fois la substance et l'artefact. Ainsi, la définition de FUMER contiendra les éléments suivants :

(8) *X fume Y* = X aspire la fumée du tabac Y1 brûlé dans l'artefact Y2...

Typiquement, les deux participants correspondant à une variable scindée ont un lien de métonymie (dans notre exemple, le tabac est dans l'artefact) et ils peuvent être exprimés chacun à leur tour mais pas simultanément :

(8a) *Benoît fume du tabac.*

(8b) *...Benoît fume une pipe.*

(8c) **Benoît fume du tabac dans une pipe¹¹.*

¹¹ Signalons toutefois que certains locuteurs peuvent juger (8c) acceptables.

3.2.4.3 Dérivations sémantiques

Comme les liens actanciels, les liens de dérivations sémantiques sont aussi intéressants pour nous parce qu'ils encodent l'équivalent de toute une série de connaissances, souvent l'équivalent des CM.

Lorsque l'on parle de dérivation en linguistique, on a tendance à penser avant tout à la dérivation morphologique (cf. Henry (1996 : 19)). On voit tout de suite que les noms ÉDUCATION ou ÉDUCATEUR dérivent morphologiquement du verbe ÉDQUER ou que BLANCHIR et BLANCHEUR dérivent de BLANC. Il est clair cependant que le lien entre ces lexies n'est pas seulement morphologique mais aussi sémantique : on peut établir une relation de sens entre ces lexies, trouver des composantes sémantiques communes.

Contrairement aux exemples que nous venons de donner, il existe de nombreux cas où l'on peut constater une dérivation sémantique qui n'est pas accompagnée de dérivation morphologique. Par exemple, il existe un lien sémantique fort entre MANGER et NOURRITURE ou entre CONCOURS et CANDIDAT bien que ces lexies n'aient pas de lien morphologique.

Ces exemples montrent que la dérivation sémantique peut être marquée morphologiquement comme elle peut ne pas l'être – cela n'a pas d'importance ici. Ce qui

compte pour nous, c'est le caractère sémantique de la dérivation. Dans ce qui suit, nous proposons de la définir suivant Mel'čuk et Polguère (à paraître).

Une **dérivation sémantique** est une relation lexicale entre deux lexies basée sur une parenté de sens, selon les trois cas suivants :

1. Les deux lexies ont (approximativement) le même sens. Elles peuvent être des synonymes exacts (VÉLO et BICYCLETTE), des synonymes approximatifs (OBTENIR et ACQUÉRIR), des conversifs (DONNER et RECEVOIR), des hyponymes/hyperonymes (MAISON et CHALET) lorsqu'elles appartiennent à la même catégorie syntaxique. Une de ces lexies peut également être un dérivé nominal, verbal, adjectival ou adverbial de l'autre (AGRESSER / AGRESSION / AGRESSIF / AGRESSIVEMENT)
2. Les deux lexies sont des antonymes exacts ou approximatifs, c'est-à-dire qu'elles ont des sens opposés (CHAUD et FROID, ARRIVER et PARTIR).
3. Une des lexies désigne un élément de la situation désignée par l'autre lexie. Il peut s'agir d'un participant (CANDIDAT pour CONCOURS, LECTEUR pour LIRE), d'un circonstant (PISCINE pour NAGER, ÉTABLE pour TRAIRE) ou même d'une caractéristique d'un participant ou d'un circonstant (CUIT ou CRU pour MANGER, VAINQUEUR et VAINCU pour GUERRE).

Certaines dérivations sémantiques sont propres à une langue donnée (Cf. Polguère (à paraître)) mais il en existe une cinquantaine que l'on retrouve à travers toutes les langues. La Théorie Sens-Texte les modélise à l'aide des Fonctions Lexicales, mentionnées plus haut.

Pour conclure, nous proposons d'illustrer l'intérêt des dérivations sémantiques par l'exemple suivant :

(9) *Jean a voulu nager ce matin mais la piscine était fermée.*

De quelle piscine s'agit-il ? La dérivation sémantique qui existe entre NAGER et PISCINE peut nous aider à dire qu'il s'agit de la piscine où Jean voulait nager. Bien entendu, il est clair que cette connaissance lexicale correspond aussi à une connaissance du monde, à savoir qu'il y a une relation entre la piscine et l'action de nager. Ce qui est important pour nous ici, c'est que les CL (en occurrence une dérivation sémantique) pourraient être suffisantes même sans les CM pour établir ce lien.

3.2.4.4 Collocations

Les collocations ont un double intérêt pour notre recherche. Premièrement, elles aussi, elles encodent de nombreuses connaissances, y compris l'équivalent des connaissances du monde. Deuxièmement, elles ont tendance à contrôler le choix de

l'article défini/indéfini – ce qui devient très important pour nous lors de l'étude des expressions définies.

Nous proposons d'introduire la notion de collocation à travers les exemples suivants :

(10a) *Isabelle mange la pomme.*

(10b) *Julie regarde la télé.*

Bien que la structure syntaxique de ces deux phrases soit identique, la nature de l'expression verbale qui s'y trouve est très différente. Dans (10a), on a une expression libre : *manger la pomme*. Les lexie MANGER et POMME y ont chacune leur sens plein, indépendant l'un de l'autre. En effet, remplacer POMME par POIRE ou par VIANDE dans cette phrase n'affecterait aucunement le sens de MANGER. De la même façon, remplacer MANGER par COUPER ou ACHETER ne changerait point le sens de POMME. Dans (10b), nous avons une expression semi-figée¹² : *regarder la télé*. TÉLÉ y a son sens plein mais le sens de REGARDER dépend de celui de TÉLÉ. Si dans (10b), on remplaçait TÉLÉ par TABLE ou par CUISINIÈRE,

¹² On pourrait nous objecter que (10b) est ambiguë : elle peut contenir soit une expression semi-figée soit une expression libre, comme dans *regarder une table*. Il faut souligner que la première interprétation est bien plus naturelle en l'absence d'autres informations. Nos remarques concernent seulement cette première interprétation.

le sens de REGARDER ne pourrait plus rester le même : on obtiendrait une expression libre.

On retrouve le même type d'expression semi-figée dans *café noir*, *peur bleue*, *sommeil léger*, etc. La TST appelle de telles expressions des collocations. Nous nous limitons ici à en donner la définition (pour plus de détails, Mel'čuk et Polguère (à paraître) ou dans Polguère (2000)) :

Une **collocation** est une expression semi-figée qui est constituée d'une **base** que le locuteur choisit librement en fonction de son sens (TÉLÉ dans notre exemple) et d'un **collocatif** (REGARDER) choisi pour exprimer un sens donné en fonction de la base, c'est-à-dire de façon contrainte.

Il n'est peut-être pas inutile de souligner encore une fois le caractère linguistique, et conceptuellement imprévisible, des collocations – et la nécessité de les mettre dans un dictionnaire. Nous nous contentons de mentionner trois raisons de cette nécessité :

a) Contrairement à ce qui est le cas pour les expressions libres, il n'est pas toujours possible de calculer le sens d'une collocation à partir du sens des éléments qui la constituent.

b) Lorsque l'on apprend une langue étrangère, on ne peut pas faire une traduction automatique, mot à mot, d'une collocation. Par exemple, l'équivalent anglais de *regarder la télé* serait *watch tv* et non pas **look at (the) TV* – on remarque notamment la disparition de l'article.

c) On ne peut pas postuler l'existence d'un mécanisme génératif (comme la métonymie) qui générerait une sorte de sens figuré ou contextuel. En effet, si un tel mécanisme existait, on voit mal pourquoi des expressions comme *regarder la télé* ou *écouter la radio* peuvent être interprétées dans un sens non littéral alors que tel n'est pas le cas avec **regarder l'écran* ou **écouter les haut-parleurs*.

3.3 Place du DiCo dans notre recherche

Nous avons vu dans la section (3.2) qu'il existe des critères précis qui nous permettent d'identifier ce que sont les CL, sans y inclure des CM non encodées dans la langue. Or, ces CL doivent être inventoriées dans un dictionnaire. Cette section est consacrée au DiCo, le dictionnaire que nous avons choisi pour notre analyse.

En premier lieu, nous examinons les motivations qui nous ont fait préférer le DiCo à d'autres dictionnaires pour étudier la place des CL dans le processus d'interprétation (3.3.1). Ensuite, nous faisons une présentation rapide de ce dictionnaire (3.3.2). À la

fin de la section, nous présenterons les dix champs lexicaux qui constituent les articles de DiCo (3.3.3).

3.3.1 Pourquoi choisir le DiCo pour notre recherche ?

Pourquoi nous intéresser au DiCo plutôt qu'à un autre dictionnaire pour étudier la place des CL dans la compréhension ? Nous avons mentionné que les dictionnaires existants, tels *Le Robert* pour le français ou bien le *Collins Cobuild English Dictionary* pour l'anglais, contiennent de nombreuses CL. Toutefois, l'utilisation de ces dictionnaires entraînerait des inconvénients majeurs : d'une part, malgré tous les efforts de leurs auteurs, ils n'ont pas toujours réussi à en exclure les CM qui ne font pas partie de la langue, d'autre part, les CL qui s'y trouvent ne sont pas toujours marquées de façon explicite. En effet, il arrive souvent que des informations pertinentes soient données à travers les exemples seulement.

Les critères d'élaboration du *Dictionnaire Explicatif et Combinatoire*, le DEC (Mel'čuk *et al.* (1984, 1988, 1992, 1999)), pourraient en faire un dictionnaire idéal pour notre but. Malheureusement, le DEC présente également quelques inconvénients majeurs – au moins pour l'instant. Le premier inconvénient est que les quatre volumes déjà publiés ne contiennent pas assez de vocables pour permettre un traitement suffisant de la langue. Bien entendu, d'autres volumes sortiront et couvriront d'autres vocables mais, étant donné le caractère détaillé des informations que le DEC contient, tout cela prend un temps considérable. Un autre inconvénient du

DEC – de notre point de vue – est qu’il est destiné à des humains. Or, nous ne cachons pas que la place des CL dans le processus d’interprétation nous intéresse entre autre dans le but de rendre les machines capables d’interpréter, de comprendre les énoncés à moyen ou à long terme. Pour le réaliser, il faut les doter des connaissances nécessaires à ce processus – ce qui suppose que l’on soit capable d’énumérer ces connaissances et de les transmettre aux machines dans un format qui leur est accessible.

L’utilisation du DiCo nous permet d’éviter à la fois les inconvénients des dictionnaires traditionnels et ceux du DEC : on y trouve des CL de façon explicite, des CM non encodées dans la langue en sont exclus, il couvre un lexique plus large que le DEC, sa nomenclature croît plus rapidement et il peut être utilisé par les ordinateurs.

3.3.2 Présentation du DiCo

Le DiCo est une base de données lexicales du français, développée à l’Observatoire de linguistique Sens-Texte (OLST) par Igor Mel’čuk et Alain Polguère. Les membres de l’OLST travaillent dans le cadre de la Théorie Sens-Texte (cf. Mel’čuk 1997). Rappelons que la composante lexicale de cette théorie linguistique est la lexicologie explicative et combinatoire qui propose les principes d’élaboration du DEC dont quatre volumes sont déjà publiés (Mel’čuk 1984, 1988, 1992 et 1999). Le DiCo est une version simplifiée et plus formalisée du DEC.

Le but du DiCo est double. En premier lieu, il doit permettre la production automatique des lexiques pour des systèmes de Traitement Automatique de la Langue (TAL), d'où l'importance de sa formalisation. Ensuite, le DiCo doit également contenir toute l'information nécessaire pour en dériver des versions « vulgarisées » à l'usage du grand public, tel le *Lexique Actif du Français* (LAF).

Pour atteindre ce double but, le DiCo décrit chaque lexie selon deux axes : les *dérivations sémantiques* qu'elle entretient avec d'autres lexies de la langue et les *collocations* (expressions semi-idiomatiques) qu'elle contrôle. Bien que le DiCo ne contienne pas la définition lexicographique des lexies (comme le fait le DEC) pour le moment, on y trouve les deux éléments les plus importants de cette définition : les étiquettes sémantiques (correspondant au genre prochain) et la structure actancielle.

Un article de dictionnaire entier est consacré à chaque lexie, aussi bien dans le DiCo que dans le DEC ; on appelle cette lexie le **mot-clé** (de l'article). La description d'un vocable est la somme des articles des lexies qu'il contient.

Un article du DiCo contient toute une série d'informations lexicales dont plusieurs nous sont très utiles, notamment le genre prochain, la structure actancielle (= la forme propositionnelle), les dérivations sémantiques et les collocations du mot-clé.

3.3.3 Dix champs principaux de description d'une lexie dans le DiCo

La description d'une lexie dans le DiCo contient dix champs principaux¹³. Nous en proposons ici une présentation très sommaire.

1) *Nom de vocable : Nom*

Ce champ contient le nom du vocable auquel appartient la lexie.

2) *Numéro de lexie*

Ce champ contient le numéro distinctif des lexies du même vocable. Il est vide dans le cas d'un vocable monosémique.

3) *Caractérisation grammaticale : cg*

La caractérisation grammaticale d'une lexie inclut ses propriétés de combinatoire grammaticale, notamment la partie du discours, le genre pour les noms, et éventuellement d'autres informations, telles l'absence du pluriel ou de la forme passive, etc.

¹³ La version de DiCo que nous utilisons contient dix champs mais dans une version future, le champ de caractérisation sémantique sera scindé en deux : il y aura un champ pour l'étiquette sémantique et un autre pour la structure actancielle.

4) *Nota Bene : nb*

Ce champ s'adresse aux lecteurs humains ; il contient de l'information utile mais non nécessairement formalisée.

5) *Caractérisation sémantique : cs*

Bien que le DiCo ne contienne pas de définition complète des lexies, la caractérisation sémantique inclut la partie combinatoire de la définition : l'étiquette sémantique et la forme propositionnelle. L'**étiquette sémantique** d'une lexie est une formule encodant la paraphrase minimale de cette lexie, c'est-à-dire le genre prochain de sa définition. La **structure actancielle** (ou forme propositionnelle) d'une lexie est la formalisation d'une expression ayant la même valeur syntaxique que la lexie décrite : comme son nom l'indique, elle décrit la relation entre les actants de la lexie.

6) *Tableau de régime : tr*

Le tableau de régime établit une correspondance entre les actants sémantiques, les actants syntaxiques profonds et les actants syntaxiques de surface d'une lexie.

7) *Dérivés sémantiques synonymes : syn*

Ce champ contient les synonymes {Syn}, les quasi-synonymes {QSyn} et les hyperonymes {Gener} de la lexie.

8) Liens de fonctions lexicales non synonymes : fl

C'est ce champ qui contient les dérivations sémantiques et les collocations contrôlées par le mot-clé.

9) Exemples : ex

Ce champ contient des exemples dont la fonction est purement illustrative : toute l'information sur le mot-clé est explicitement donnée et encodée dans les autres champs.

10) Pointeurs vers les locutions (phrasèmes) : ph

On énumère dans ce champ les locutions dont le mot-clé est un des constituants lexicaux.

Dans l'annexe, nous présentons quatre articles du DiCo qui dénotent les quatre lieux publics que nous étudions.

La présentation de notre cadre théorique et l'outil principal de nos analyses se termine ici (nous les utiliserons lors des analyses du défini) et nous pouvons examiner maintenant le problème de l'interprétation des énoncés.

4 Délimitation de l'interprétation des énoncés

Ce chapitre est consacré tout particulièrement au problème de l'interprétation des énoncés. En premier lieu, nous présentons la notion d'énoncé. Ensuite, nous nous intéressons respectivement à la compréhension et à l'interprétation : ce sont pour nous deux processus à la fois liés et distincts. Nous devons également nous pencher sur la notion de contexte car l'interprétation des énoncés ne peut pas se faire hors contexte. À la fin de ce chapitre, nous présentons brièvement trois problèmes qui relèvent de l'interprétation : la désambiguïsation, le paraphrasage et la résolution de l'anaphore. (Dans le chapitre suivant, nous examinons en détail un autre cas particulier d'interprétation, à savoir l'interprétation du défini.)

Bien entendu, toutes ces notions ont déjà été étudiées maintes fois. Notre but n'est pas tant de proposer de nouvelles définitions que de clarifier celles qui existent – ceci est une nécessité absolue pour bien effectuer notre travail. Nous nous efforçons d'utiliser une terminologie aussi explicite, logique et cohérente avec d'autres écrits que possible.

4.1 Présentation de la notion d'énoncé

Notre définition de l'énoncé est basée sur l'opposition saussurienne entre langue et parole – une opposition qui peut être mise en correspondance avec celle entre type et token, avancée par Peirce.

4.1.1 Langue et parole, type et token

Dans le *Cours de linguistique générale*, Saussure fait une distinction entre langue et parole. La **langue** se définit comme un code, une mise en correspondance de « concepts » et « d'images acoustiques ». La **parole** est l'utilisation, la mise en œuvre de ce code par les sujets parlants. La langue n'est pas une activité mais une collection de lois, de règles. Toute activité liée à la langue appartient, à proprement parler, à la parole.

Cela implique une distinction claire entre les éléments de la langue et ceux de la parole. Les éléments de la langue sont des entités abstraites alors que des éléments de la parole sont leurs réalisations respectives, produites par des locuteurs à des moments et dans des lieux déterminés.

Peirce¹⁴ établit une distinction semblable entre type et token. Pour lui, le **type** est le signe comme valeur abstraite – par exemple, l’unité lexicale figurant au dictionnaire ou dans la langue. Le **token** est la réalisation individuelle de ce signe – par exemple, une occurrence spécifique d’une unité lexicale dans un acte de parole. On pourrait ainsi établir un parallélisme, d’une part, entre le type chez Peirce et la langue chez Saussure, d’autre part, entre le token chez Peirce et la parole chez Saussure.

4.1.2 Définition de l’énoncé

L’**énoncé** est l’élément de base de la parole : il est grammaticalement autonome. D’un point de vue syntaxique, l’énoncé prototypique correspond à une phrase mais un énoncé peut aussi correspondre à une unité syntaxique plus petite que la phrase (un syntagme, voire une seule lexie). Pour illustrer cela, considérons le dialogue (1) :

(1) *Où vas-tu ? - À la maison.*

Dans ce dialogue, nous trouvons deux énoncés : la question *Où vas-tu ?* est la réalisation d’une phrase complète alors que la réponse *À la maison* est la réalisation d’un syntagme.

¹⁴ Cf Peirce (1978) et Fisette (1990).

On voit que l'énoncé est proche du token de Pierce. La différence principale entre les deux réside dans le fait que Pierce associe le token à un seul signe alors que l'énoncé est le plus souvent la combinaison de plusieurs signes, plusieurs unités lexicales.

4.2 Compréhension des énoncés

Les termes *interprétation* et *compréhension* sont parfois utilisés comme synonymes. On les rencontre souvent (surtout le deuxième) chez les linguistes mais il est extrêmement difficile d'en trouver une définition précise : la plupart du temps, leur définition est laissée à l'intuition des lecteurs. C'est d'ailleurs ce que nous avons fait jusqu'ici. Il est temps maintenant de préciser ce que nous entendons par ces termes.

Nous proposons d'établir une distinction entre la notion de compréhension et d'interprétation : la première, conformément à une terminologie et un usage répandu, désigne un processus cognitif très complexe alors que la seconde désigne un processus plus restreint et avant tout (mais pas exclusivement) linguistique.

4.2.1 Problématique de la compréhension

Comme nous venons de l'affirmer, la **compréhension** des énoncés est un processus cognitif très complexe. Il est évident que la compréhension d'un énoncé inclut

l'identification du sens de la phrase ou du syntagme dont cet énoncé est la réalisation particulière. Par exemple, des personnes différentes, dans des situations différentes, peuvent dire *J'ai faim*. Tous ces énoncés ont en commun le sens de la phrase *J'ai faim*. Toutefois, contrairement à cette phrase qui est un *type*, qui fait partie de la langue, les énoncés sont des *token*, ils font partie de la parole et ils ont aussi un contenu spécifique qu'ils ne partagent pas avec les autres énoncés – c'est notamment le cas de leur référent.

On a également tendance à inclure dans la compréhension l'identification du référent. Imaginons par exemple un dialogue entre deux locuteurs, A et B. A fait une affirmation au sujet de quelqu'un en utilisant un groupe nominal : *le plus grand mathématicien du monde, l'ami de Pierre, l'étudiante de Philippe*, etc. B peut l'interrompre en lui disant qu'il n'a pas compris de qui il est question. Clairement, dans une telle situation, cela signifie une incapacité d'identifier le référent de l'expression.

Cependant, on a l'habitude de ne pas limiter la compréhension d'un énoncé à identifier son sens et son référent. Dans ce qui suit, nous illustrons cette notion de compréhension au sens plus large d'abord à travers les approches pratiques ensuite à travers les approches théoriques.

4.2.2 Compréhension : d'un point de vue pratique

Sanford & Garrod (1981 : 4) affirment qu'un programme peut être considéré comme capable de comprendre s'il peut paraphraser ou résumer un discours, le traduire dans une autre langue et donner des réponses intelligentes à des questions concernant ce discours. Un de leurs exemples est le suivant :

(2) *Jean est tombé malade. Marie a appelé un médecin.*

Le lien de causalité entre ces deux énoncés n'est pas marqué de façon explicite. Sanford et Garrod considèrent cependant que pouvoir établir ce lien fait clairement partie de la compréhension. Autrement dit, la compréhension inclut la capacité de donner à la question (3), portant sur (2), une réponse comme (4) :

(3) *Pourquoi Marie a-t-elle appelé le médecin ?*

(4) *Parce que Jean est tombé malade.*

Reiter (1990 : 257) affirme que lorsqu'une personne entend un énoncé, elle utilise ses connaissances du monde pour faire des inférences et élargir le contenu sémantique de cet énoncé. Il donne l'exemple suivant :

(5) *Il y a un requin dans l'eau.*

Celui qui entend (5) ne s'arrêtera pas à son sens mais, en faisant appel à ses connaissances concernant les requins, fera aussi des inférences, notamment qu'il n'est pas conseillé de nager dans cette eau. Il est clair aussi que le locuteur attend de son auditeur qu'il fasse des inférences. On peut ainsi se demander s'il faut considérer que ces inférences obtenues à l'aide de nos CM font partie de la compréhension.

C'est notamment le cas de Lenat & Guha (1990 : 2). Les auteurs donnent l'exemple d'un système expert accordant un prêt à quelqu'un qui déclare avoir le même emploi depuis 20 ans bien qu'il ait également déclaré qu'il n'avait que 18 ans. Ils considèrent que la machine n'a pas compris ce que c'est que d'avoir 18 ans et d'avoir travaillé depuis 20 ans. Or, pour être exact, ce que la machine n'a pas fait c'est de vérifier la compatibilité de ces deux affirmations : elle n'a pas fait des inférences à partir de ces affirmations en utilisant des CM.

4.2.3 Compréhension : approches théoriques

D'un point de vue théorique, il est clair que la compréhension des énoncés inclut des éléments de natures différentes. Austin (1970) considère qu'en produisant un énoncé, on accomplit trois actes : locutoire, illocutoire et perlocutoire. **L'acte locutoire** est le fait de parler, de produire un énoncé. **L'acte illocutoire** est l'ensemble des actes (demander, promettre, informer...) accomplis directement par l'utilisation de

la parole. **L'acte perlocutoire** est un acte accompli indirectement : l'auditeur peut ne pas le comprendre et l'énonciateur peut nier l'avoir accompli intentionnellement.

Ducrot (1991) affirme également qu'un énoncé contient des éléments de différentes natures. Il distingue entre le composant linguistique et le composant rhétorique d'un énoncé. Le composant linguistique contient le posé et le présupposé. Le **posé** est l'objet d'un acte d'affirmation particulier alors que le **présupposé** est présenté comme une donnée à partir de laquelle on parle, mais qui n'est pas directement en jeu dans la parole. Le présupposé est l'élément sémantique commun à un énoncé *A*, à sa transformation interrogative *Est-ce que A ?* et à sa négation *Il est faux que A*. Les **sous-entendus** font partie du composant rhétorique et sont dépendants du contexte. Ils sont découverts par une sorte de raisonnement qui s'appuie sur l'événement qui constitue l'énonciation. Le sous-entendu a la particularité de pouvoir toujours être nié : le locuteur peut ne pas le prendre en charge.

On pourrait ainsi soutenir qu'identifier les actes locutoires, illocutoires et perlocutoires, et aussi le posé, le présupposé et les sous-entendus d'un énoncé fait partie de sa compréhension. Toutefois, ce qui nous intéresse le plus ici est de savoir s'il existe un fondement théorique pour inclure dans la compréhension d'un énoncé (toutes) les inférences que l'on peut faire à partir de cet énoncé et des connaissances du monde qui pourraient y être jointes – comme le suggèrent certaines approches pratiques, notamment le projet Cyc, mentionné dans (2.2.2).

Inclure les inférences dans la compréhension peut paraître étrange dans un cadre saussurien mais il est tout naturel, pour ne pas dire la seule façon de faire possible, dans un cadre peircéen. En effet, Saussure s'intéresse avant tout à l'étude interne de la langue : le signe dont il est question dans le *Cours de Linguistique Générale* est, par conséquent, le signe en tant qu'élément de la langue. C'est une unité codée, figée, définie par son appartenance à un système, à une langue. Cela correspond à la vision classique sur le signe : un signe est quelque chose qui vaut pour quelque chose d'autre.

Peirce, par contre, ne se limite pas à la langue mais il s'intéresse à tous les signes : aux signes linguistiques aussi bien qu'aux signes non linguistiques. Pour lui, le signe est quelque chose dont la connaissance nous permet de connaître quelque chose de plus. Le signe est ainsi un état temporel, un moment dans un processus sans fin d'acquisition de savoir. Il appelle ce processus théoriquement illimité la *sémiosis ad infinitum*.

Si le signe saussurien est statique alors que le signe peircéen est dynamique, c'est parce que Saussure se limite à la langue alors que Peirce inclut dans ses recherches aussi la parole et bien plus que la parole (même s'il n'utilise pas ce terme). À proprement parler, il n'y a pas d'opposition entre la conception de Saussure et celle de Peirce car ils n'ont pas le même objet d'étude : l'objet d'étude de Saussure est inclus dans celui de Peirce. Le contenu d'une lexie, d'une phrase ou d'un texte est figé mais le contenu d'une occurrence d'une lexie, d'un énoncé ou d'un discours ne l'est pas.

Soulignons cependant que même s'il est possible de faire une infinité d'inférences à partir d'un énoncé, toutes ces inférences ne sont pas pertinentes. La question de la pertinence était au cœur des travaux de Sperber et Wilson (1986) qui distinguent deux modèles de communication – en se basant entre autre sur les recherches de Peirce. Le premier, utilisé par Aristote aussi bien que par les sémioticiens modernes, est le **modèle du code** : selon ce modèle, communiquer, c'est coder et décoder des messages. Le deuxième, développé par Paul Grice et d'autres philosophes modernes, est le **modèle inférentiel** : selon ce modèle, communiquer, c'est produire et interpréter des indices¹⁵. Les auteurs cherchent à incorporer ces deux modèles dans leur théorie.

Ils appellent **contexte** l'ensemble des prémisses utilisées pour la compréhension d'un énoncé. C'est une construction psychologique, un sous-ensemble des hypothèses de l'auditeur sur le monde. Un contexte n'est pas ce que l'on appelle le **savoir partagé**, il ne contient pas seulement de l'information sur l'environnement physique immédiat ou sur les énoncés précédents : des prévisions, des hypothèses scientifiques, des croyances religieuses, des souvenirs, des préjugés culturels sont susceptibles de jouer un rôle dans l'interprétation.

¹⁵ Par *indice*, il faut entendre un signe qui indique quelque chose avec probabilité dans une situation donnée.

La **pertinence** d'un énoncé dépend de son coût et de son effet cognitif. Le **coût** est l'effort nécessaire pour faire des inférences et l'**effet cognitif** est l'ensemble des propositions que l'on peut inférer de l'énoncé joint à un contexte et que l'on n'inférerait pas du seul contexte. La pertinence d'un énoncé est proportionnelle à son effet cognitif et inversement proportionnel au coût.

Pour Sperber et Wilson, la compréhension d'un énoncé dans un contexte donné est l'ensemble des propositions inférables de l'énoncé et qui le rendent le plus pertinent possible.

4.2.4 En résumé

Nous venons de voir qu'il y a des arguments aussi bien pratiques que théoriques pour inclure dans la compréhension des énoncés les inférences que l'on pourrait faire à partir de ces énoncés. Toutefois, on pourrait nous faire deux objections contre cette inclusion : la première serait de dire que ce n'est plus de la linguistique, la deuxième consisterait à rejeter une telle définition de la compréhension sous prétexte qu'elle est infinie et ne pourrait jamais être atteinte, modélisée.

La première objection est facile à réfuter : même si la compréhension ainsi définie ne peut plus être traitée par l'étude interne de la langue, il est possible de faire l'étude externe de la parole – une étude qui peut inclure toutes les disciplines qui, de près ou de loin, ont un rapport avec la parole. Cela serait tout à fait naturel lorsque

l'objectif n'est pas de faire une recherche théorique à l'intérieur d'une discipline donnée mais de résoudre un problème concret : modéliser ce que font les humains et le reproduire par la machine.

La deuxième objection semble être plus difficile à réfuter. Il est vrai que l'interprétation ainsi définie est potentiellement infinie et risque d'être difficilement décrite. Il est également clair que lorsque quelqu'un comprend un énoncé, il n'accomplit pas un processus infini. On peut cependant répondre à cette objection que l'infini est du domaine de la théorie et le fini est du domaine de la pratique. Autrement dit, même si la compréhension d'un énoncé pouvait être théoriquement infinie, la compréhension accomplie par l'auditeur reste un processus fini. Remarquons que la situation est semblable concernant les phrases d'une langue : bien que leur nombre soit théoriquement infini, il n'a jamais été (et ne sera jamais) produit qu'un nombre fini de phrases de cette langue. La raison de cette compréhension finie est pratique : seulement une partie des inférences possibles est pertinente. Autrement dit, les difficultés posées par la compréhension (potentiellement) infinie disparaissent si on arrive à la rendre finie par une théorie de la pertinence.

4.3 Interprétation des énoncés

Dans la section précédente, nous avons vu combien la notion de compréhension des énoncés est large, pouvant dénoter un processus infini. Il nous paraît important de la

distinguer ici d'une autre notion, plus restreinte ; l'interprétation des énoncés. L'interprétation et la compréhension des énoncés ont en commun le fait qu'elles vont toutes les deux du signifiant vers le signifié – contrairement à la génération qui va du signifié vers le signifiant.

4.3.1 Définition de l'interprétation

Nous appelons **interprétation d'un énoncé** un processus cognitif qui consiste à identifier le sens et le référent de cet énoncé.

Le sens, rappelons-le, est le signifié saussurien. Il est le contenu d'un élément de la langue (lexie, syntagme, phrase...) et il s'oppose à la signification qui est le contenu d'un élément de la parole¹⁶. Lorsque nous entendons un énoncé, nous n'avons un accès direct qu'à son signifiant – c'est à travers le processus d'interprétation que nous pouvons accéder à son sens. Prenons un exemple concret :

(6) *Pierre cherche la cuisinière.*

¹⁶ Notre terminologie concernant « sens » et « signification » nous paraît conforme à la terminologie de nombreux auteurs : Fisette (1990 : 6), Sabah (1991 : 259), Fuchs (1982 : 75), etc. D'autres les utilisent dans le sens inverse : Ducrot (1991 : 111).

L'énoncé (6) pourrait avoir de nombreux sens, étant donné que chacun des signifiants qui s'y trouvent pourrait être le signifiant de plusieurs lexies. La structure syntaxique de la phrase élimine certaines de ces possibilités : *la* ne peut être ni le signifiant de la lexie dont le sens est 'note musicale...' ni celui du pronom, etc. Toutefois, les contraintes syntaxiques ne permettent pas une désambiguïsation totale de (6). Par exemple, quel est le sens de l'occurrence de CUISINIÈRE ? Est-ce 'personne qui...' ? Ou bien, est-ce plutôt 'appareil destiné à...' ? L'énoncé seul ne nous permet pas de trancher, il faudrait examiner son contexte d'utilisation et peut-être même utiliser des CM.

Une fois ce sens établi, l'auditeur peut s'attaquer au problème du référent. Dans notre exemple, il s'agit de répondre à la question suivante : de quelle cuisinière parle-t-on ? Une tâche ambitieuse serait de chercher à lier ce référent à un élément d'une base de connaissances. Une tâche un peu plus modeste pourrait être d'établir des liens entre les référents de différentes occurrences des lexies ou des expressions. C'est plutôt cette deuxième tâche que nous visons dans ce travail. Nous pouvons illustrer cela par l'énoncé suivant :

(7) *Je suis allé dans un restaurant indien hier et j'ai beaucoup apprécié le repas.*

Bien entendu, on peut facilement trouver le lien lexical qui existe entre les lexies RESTAURANT et REPAS. Toutefois, au niveau des référents, une autre question se pose : quel lien peut-on trouver entre le repas et le restaurant en question ? Dans

notre cas, la réponse est simple : il s'agit du repas qui a été servi dans le restaurant indien mentionné. Bien entendu, il existe de nombreux cas où la réponse est nettement plus difficile à trouver.

4.3.2 Choix de l'interprétation

Le lecteur pourrait se poser trois questions concernant le choix de l'interprétation :

- (Q1) Pourquoi s'intéresser à l'interprétation des énoncés ?
- (Q2) Pourquoi s'intéresser à l'interprétation des énoncés plutôt qu'à la génération ?
- (Q3) Pourquoi s'intéresser à l'interprétation des énoncés plutôt qu'à la compréhension ?

Dans cette sous-section, nous proposons une réponse à chacune de ces questions.

Pourquoi s'intéresser à l'interprétation ? Au-delà de son importance théorique dans l'étude de la langue, l'interprétation a aussi une importance certaine qui est d'ordre pratique : un nombre grandissant d'applications informatiques en a besoin. Bien entendu, ce besoin ne signifie pas que ces applications soient inconcevables sans une capacité d'interprétation : les moteurs de recherche peuvent utiliser tout simplement des requêtes par mots-clés, la traduction assistée par ordinateur peut être basée sur des méthodes statistiques, etc. Il est clair cependant qu'une capacité d'interprétation

(comparable à celle des humains) pourrait substantiellement améliorer ces applications.

L'interprétation joue un rôle essentiel dans la recherche d'information – ce qui est probablement l'application la plus en demande actuellement : pensons à des moteurs de recherches comme Google (google.com), AltaVista (altavista.com) ou Ask Jeeves (ask.com). Dans la traduction assistée par ordinateur (Reverso – softissimo.com), elle est aussi importante que la génération. Elle peut également jouer un rôle dans la reconnaissance de la parole (ViaVoice – ibm.com ; Dragon Naturally Speaking – lhsl.com), la correction grammaticale (Antidote - druide.com ; Correcteur – documens.com), etc. Or, nous croyons que la réussite de toutes ces applications dépend avant tout non pas d'un travail d'ingénierie astucieux mais d'une bonne modélisation de la langue en général et de l'interprétation en particulier.

Pourquoi s'intéresser à l'interprétation plutôt qu'à la génération ? Nous n'avons qu'une réponse subjective à cette question : notre préférence relève avant tout de l'intérêt personnel. Nous considérons que l'interprétation et la génération sont d'importance sinon égales au moins comparables et ce aussi bien sur le plan théorique que sur le plan pratique (voir Dale *et al.* (1990) pour la génération, Sanford et Garrod (1981) pour l'interprétation/compréhension et Sabah (1988) pour les deux).

D'un point de vue théorique, il est clair que la description exhaustive de la langue doit inclure aussi bien l'étude de la génération que celle de l'interprétation. Par ail-

leurs, aucune de ces deux études ne présuppose l'étude préalable de l'autre. On pourrait nous objecter que, dans les faits, la génération précède l'interprétation. Malgré cette différence chronologique (un énoncé est d'abord généré puis interprété), les deux processus nous semblent inséparables. Lors de la génération d'un énoncé, l'énonciateur doit prendre en considération son interprétation par le destinataire et, de la même façon, lors de son interprétation, le destinataire doit prendre en considération sa génération par l'énonciateur.

D'un point de vue pratique, notamment en linguistique informatique, la situation est semblable : les tâches que nous venons de mentionner, notamment les questions-réponses, la traduction automatique, etc. à des degrés divers, nécessitent la génération et/ou l'interprétation.

À la troisième question, concernant le choix de l'interprétation plutôt que celui de la compréhension, nous avons une réponse de l'ordre logique : l'interprétation peut être vue comme la première étape de la compréhension. En effet, la compréhension telle que nous l'avons présentée dans la section précédente nécessite des inférences à partir des énoncés et des CM. Pour ce faire, il faut au préalable établir le sens et le référent de ces énoncés – autrement dit, il faut les interpréter.

Notre choix de l'interprétation s'explique aussi par des considérations pratiques. Il est clair qu'en incluant dans la compréhension des énoncés les inférences faites à partir de ces énoncés et des CM qui y sont jointes, nous pourrions nous retrouver

face à un processus infini. Nous croyons qu'une théorie de la pertinence permettant de réduire considérablement cet infini est possible. Toutefois, étudier la compréhension (ou même la partie pertinente de la compréhension) serait un travail immense, dépassant les cadres d'une thèse – au moins pour deux raisons. Premièrement, la théorie de la pertinence n'est encore qu'en état embryonnaire : la développer pourrait largement remplir une thèse. Ensuite, étudier la compréhension nécessiterait la prise en considération d'une immense quantité de CM – or, nous n'avons pas de base de données adéquate pour le faire.

4.4 Contexte des énoncés

La présentation de l'interprétation des énoncés ne saurait être complète sans examiner le contexte des énoncés. Il est évident qu'il n'y a pas d'énoncé hors contexte. Il est important de ne pas perdre cela de vue sinon on risque de commettre des erreurs méthodologiques : par exemple, attribuer à la sémantique les informations qui sont dans le contexte ou l'inverse. Selon Ducrot (1991 : 107), ce que l'on appelle une occurrence hors contexte n'est qu'une occurrence dans un contexte artificiellement simplifié et il n'est nullement nécessaire que la signification constatée dans ces conditions permette de comprendre celles qui sont enregistrées dans des contextes naturels.

4.4.1 Définition du contexte

Nous considérons que le **contexte d'un énoncé** est constitué de quatre parties : le cotexte, les connaissances linguistiques activées par l'énoncé et le cotexte, la situation et les connaissances du monde activées par la situation.

La première partie du contexte d'un énoncé est le **cotexte**, c'est-à-dire l'entourage linguistique de l'énoncé. Nous pouvons le définir de la façon suivante :

Étant donné un texte T et une partie P de ce texte, le cotexte de P est le reste de T.

Autrement dit : Cotexte de P = T – P.

Les connaissances linguistiques activées par l'énoncé et le cotexte sont la deuxième partie du contexte d'un énoncé. Les interlocuteurs possèdent une quantité immense de connaissances linguistiques mais ils n'en utilisent qu'une infime partie pour comprendre un énoncé. Or, cette partie des connaissances dépend justement de l'énoncé (et notamment des lexies qu'il contient) et du cotexte.

La troisième partie du contexte d'un énoncé est la **situation**, c'est-à-dire l'entourage extralinguistique de l'énoncé. Traditionnellement, on considère les connaissances sur le locuteur (MOI), le destinataire (TOI), le moment (MAINTENANT) et le lieu (ICI) de la parole comme les **paramètres principaux de l'énonciation**. Ce sont ces quatre paramètres qui déterminent la situation.

Les connaissances du monde activées par l'énoncé et la situation forment la quatrième partie du contexte d'un énoncé. Les interlocuteurs possèdent non seulement des connaissances linguistiques mais aussi une quantité immense de connaissances du monde. Bien évidemment, ils n'en utilisent qu'une partie extrêmement petite pour comprendre un énoncé : les connaissances qui sont liées à la situation et au référent des énoncés.

S'il peut paraître naturel d'inclure dans le contexte la situation et le cotexte, il est peut-être un peu surprenant d'en faire autant avec les connaissances linguistiques et les connaissances du monde. On pourrait nous objecter que même si l'ensemble de ces connaissances peut varier d'une personne à l'autre, elles ne sont pas dépendantes d'un énoncé particulier, comme c'est le cas de la situation et du cotexte. Nous pouvons répondre à une telle objection en soulignant que nous n'avons pas inclus dans le contexte toutes ces connaissances, seulement une partie qui est activée par les énoncés, qui est dépendante des énoncés en question.

4.4.2 Illustration du contexte

Il existe un contexte par défaut, i.e. un contexte neutre, non marqué : il correspond à la situation de l'énoncé. Sauf indication contraire, on comprendra (8) comme (9) :

(8) *Il pleut.*

(9) *Il pleut ici et maintenant.*

Par contre, si (8) est une réponse à la question (10), ce qui crée un contexte différent d'un contexte par défaut, alors on le comprendra comme (11) :

(10) *Quel temps fait-il à Paris ?*

(11) *Il pleut à Paris.*

Il arrive qu'une expression définie soit précédée par un groupe nominal qui a le même référent. Dans un tel cas, le défini est anaphorique, il doit être interprété par rapport à son antécédent :

(12) *Pierre a rencontré un journaliste et un étudiant. L'étudiant lui a posé des questions difficiles.*

Dans bien des cas cependant, le cotexte tout seul est insuffisant pour créer un nouveau contexte : il doit être jumelé avec certaines connaissances liées aux énoncés à interpréter :

(13) *Nous sommes allés au théâtre jeudi dernier. Les acteurs étaient bons.*

Nous soutenons qu'il existe un lien lexical entre la lexie THÉÂTRE et la lexie ACTEUR. C'est une connaissance lexicale qui, jointe à la première partie de (13),

crée un contexte avec un théâtre, un spectacle, des acteurs, etc. D'autres diraient que ce sont avant tout nos connaissances du monde concernant le théâtre qui créent ce contexte. Ce qui est certain, c'est qu'après avoir parlé de théâtre, le locuteur peut parler des acteurs comme s'ils étaient déjà connus, identifiés par son auditeur. Quelles connaissances, quels types de connaissances sont en jeu dans des cas pareils ? Cette question est au cœur de nos recherches et nous y revenons dans (5.3), en cherchant à déterminer le rôle exact que jouent les CL dans la construction du contexte nécessaire à l'interprétation du défini.

4.5 Trois problèmes d'interprétation

Dans le chapitre suivant, nous examinons la place des CL face aux CM dans le processus d'interprétation des expressions définies et cet examen sera appuyé par l'analyse des expressions définies d'un corpus. Cependant, nous tenons à souligner que des questions semblables concernant la place des CL face aux CM se posent dans d'autres domaines aussi et nous en mentionnons trois ici : celui de la désambiguïsation, celui du paraphrasage et celui de la résolution de l'anaphore.

4.5.1 Désambiguïsation

Mel'čuk *et al.* (1995 : 60) considèrent qu'il y a ambiguïté lexicale dans un énoncé lorsqu'il existe plusieurs lexies dont le signifiant est identique mais qui ont des si-

gnifiés et des syntactiques différents et l'énoncé contient un tel signifiant. Reprenons les exemples donnés en (2.1.1), tirés de Lenat & Guha (1990) :

(14a) *The ink is in the pen.*

(14b) *The pig is in the pen.*

Ces énoncés sont potentiellement ambigus, étant donnée qu'il existe au moins deux lexies anglaises PEN¹⁷ définis respectivement par le dictionnaire *Merriam-Webster* comme 'an implement for writing or drawing with ink or a similar fluid' et 'a small enclosure for animals'. Toutefois, l'auditeur peut en faire une interprétation sans même être conscient de l'ambiguïté. En effet, il aura tendance à considérer qu'il s'agit d'un stylo dans (14a) et d'un enclos dans (14b) et pas l'inverse, même si cela n'a pas été précisé auparavant. Comment cela se fait-il ? La réponse la plus directe est que nous possédons des connaissances appropriées, indépendamment de ces énoncés. Pour les ordinateurs qui ne possèdent pas ces connaissances, par exemple pour les programmes de traduction actuellement disponibles, un tel choix n'a rien d'évident.¹⁸

¹⁷ Il en existe plus que deux mais pour notre propos ces deux lexies sont suffisantes.

¹⁸ Nous avons essayé plusieurs sites qui proposent des traductions sur Internet : voir <http://www.freetranslation.com>, <http://www.systransoft.com>, ou encore <http://www.reverso.net>. La traduction de (14a) et celle de (14b) contenaient *stylo* et non pas *enclos*.

Ces connaissances sont-elles des CL ou des CM ? Lenat et Guha soutiennent que ces connaissances ne sont pas linguistiques : pour eux, il s'agit des CM. Pourtant, la première définition de *Merriam-Webster* nous permet d'établir un lien lexical direct entre INK et PEN, et la deuxième un lien lexical indirect entre PIG et PEN (via ANIMAL). Autrement dit, les définitions même de PEN suggèrent un encodage linguistique de ces connaissances. Certes, on pourrait nous objecter qu'il y a des cas où l'ambiguïté lexicale est levée sans l'aide de CL, comme dans (15) :

(15) *The car is in the pen.*

Dans (15), on ne peut pas établir de lien lexical entre CAR et PEN, ce sont plutôt nos CM, notamment nos connaissances sur la taille habituelle des choses, qui nous permettent de lever (ou même éviter) l'ambiguïté.

Il arrive, quoique rarement, que nos connaissances suggèrent deux interprétations :

(16) *Le matelot a mangé un sous-marin.*

D'une part, on pourrait établir un lien entre MATELOT et SOUS-MARIN (dans le sens de 'engin de navigation conçu et équipé pour naviguer sous l'eau'), d'autre part, entre MANGER et SOUS-MARIN (dans le sens de 'sandwich dont la forme peut faire penser à celle d'un sous-marin', utilisé en français québécois). Même dans de tels cas, on a tendance à préférer une des interprétations. On pourrait se deman-

der, une fois de plus, si cette préférence est due aux CL ou aux CM. Rien ne nous empêche de postuler l'utilisation simultanée de ces deux types de connaissances.

4.5.2 Paraphrase

Fuchs (1982) établit une distinction entre le phénomène de paraphrase et la synonymie. Elle affirme qu'une phrase ou un texte Y constitue une **paraphrase** d'une autre phrase ou d'un autre texte X lorsque l'on considère que Y reformule le contenu de X ; autrement dit, lorsque Y et X peuvent être tenus pour deux formulations différentes d'un contenu identique, pour deux manières différentes de dire la même chose. Elle rejette l'idée d'enfermer la paraphrase à l'intérieur du système de la langue : pour elle, c'est une notion plus large que la synonymie, qui concerne les unités linguistiques ayant le même sens.

Elle distingue trois types de paraphrase : linguistique, référentielle et pragmatique. Ce qui nous intéresse ici est la paraphrase référentielle : comme son nom indique, elle concerne les unités linguistiques qui ont le même référent. Elle fait partie de la parole et non pas de la langue. Fuchs soutient que l'établissement d'une paraphrase référentielle se fonde sur la connaissance de la réalité extralinguistique. Considérons les exemples suivants :

(17a) *Jacques Chirac est un habile politicien.*

(17b) *Le président de la République est un habile politicien.*

Dans un univers de discours où Jacques Chirac est le président de la République, (17a) et (17b) sont des paraphrases référentielles, bien que leur sens soient différents.

On peut parler de paraphrase référentielle non seulement dans le cas des individus mais aussi dans le cas des événements. En s'inspirant de Pustejovsky (1995), on peut soutenir par exemple que (18a) peut être interprété comme équivalent à (18b) ou à (18c) sans que cela soit dû à une ambiguïté lexicale :

(18a) *Pierre a commencé un nouveau livre.*

(18b) *Pierre a commencé à lire un nouveau livre.*

(18c) *Pierre a commencé à écrire un nouveau livre.*

Peeters (2002) mentionne un problème semblable :

(19) *Pierre est dans l'annuaire.*

À proprement parler, ce n'est pas Pierre qui est dans l'annuaire mais son nom et son numéro de téléphone¹⁹. Autrement dit, l'énoncé (19) peut être paraphrasé, re-interprété par (20) :

(20) *Le nom et le numéro de Pierre sont dans l'annuaire.*

Une telle paraphrase ne serait plus possible dans (21) qui contient *l'enveloppe* plutôt que *Pierre* :

(21) *L'enveloppe est dans l'annuaire.*

Nous pouvons ajouter un dernier exemple à cette série – l'énoncé (22a) pourrait être équivalent tantôt à (22b) tantôt à (22c) :

(22a) *L'appartement de Philippe est très beau.*

(22b) *L'appartement où habite Philippe est très beau.*

(22c) *L'appartement que possède Philippe est très beau.*

¹⁹ Dans le cadre de la LEC, nous dirions que l'expression *être dans l'annuaire* est une collocation de ANNUAIRE. Ainsi, il s'agirait d'un problème de désambiguïsation : comment distinguer entre une expression libre et une expression semi idiomatique.

Tout comme Fuchs, Peeters et Pustejovsky s'interrogent sur la place des CL et des CM dans de tels processus et, tout en admettant que les CM puissent avoir de l'influence dans certains cas, considèrent que le rôle principal revient aux CL.

4.5.3 Résolution de l'anaphore

Nous avons parlé plus haut de la paraphrase référentielle. La résolution de l'anaphore est un autre problème concernant la co-référence : l'occurrence d'une expression anaphorique et celle de son antécédent ont, par définition, le même référent, mais comment trouver cet antécédent ? Cela est assez simple dans le cas où l'expression anaphorique est une répétition exacte de son antécédent :

(23) *Marie a rencontré un journaliste et un avocat hier. Le journaliste lui a posé toute une série de questions.*

Toutefois, cette tâche d'interprétation peut être plus difficile lorsque le défini et son antécédent contient le même nominal mais ont des modificateurs différents (*le premier ministre du Canada / le premier ministre canadien / le premier ministre Chrétien*) ou lorsqu'il s'agit d'une relation synonymiques (*l'enseignant... le professeur*) ou d'une relation hyperonymique (*le président... le chef d'état*).

Dans le cas des pronoms, la résolution d'anaphore peut même devenir très complexe :

(24a) *Pierre a dit au serveur qu'il voulait un bœuf bourguignon.*

(24b) *Pierre a dit au serveur qu'il lui avait recommandé un excellent vin.*

Malgré l'identité de la proposition principale où se trouve l'antécédent du pronom *il*, on a tendance à interpréter comme antécédent *Pierre* dans (24a) et *serveur* dans (24b). De nouveau, cette interprétation fait appel à des connaissances qui ne sont pas encodées de façon explicite et directe dans ces énoncés.

La présentation de ces trois problèmes d'interprétation termine ce chapitre où nous avons examiné de près la notion d'interprétation et celle d'énoncé. Le chapitre suivant sera entièrement consacré à un autre problème d'interprétation des énoncés : l'interprétation du défini.

5 Une étude de cas : les expressions définies

Ce chapitre est consacré à l'interprétation des expressions définies. Son objectif n'est pas tant d'en faire une étude complète mais plutôt d'en examiner les aspects qui sont pertinents pour notre but général : déterminer le rôle des connaissances lexicales face aux connaissances du monde lors du processus d'interprétation des énoncés. En effet, certaines connaissances sont nécessaires à l'interprétation du défini. La question reste à savoir lesquelles exactement ?

De nombreuses recherches ont été effectuées sur le défini ; nous en présentons les plus significatives dans (5.1), selon trois approches différentes : grammairienne, philosophique et linguistique. Ensuite, nous proposons notre propre étude du défini dans (5.2). Dans (5.3), nous proposons une introduction aux analyses présentées dans (5.4), alors que dans (5.5), nous faisons une évaluation de ces analyses.

5.1 Trois approches concernant le défini

Dans certains énoncés, on rencontre des expressions définies qui ne sont ni anaphoriques, c'est-à-dire déjà apparues dans le texte, ni déictiques, c'est-à-dire présentes dans la situation de l'énonciation (*Georges va faire un voyage à Boston parce qu'il trouve la côte est des États-Unis plus intéressante que la côte ouest*). Or l'utilisation d'une expression définie suggère que son référent doit déjà être présent, identifié

dans l'esprit des interlocuteurs. Comment cela se fait-il ? La question n'est pas nouvelle et même des pistes intéressantes étaient proposées aussi bien par des grammairiens que par des philosophes et par des linguistes. Nous étudions cette question d'abord sous l'angle de la philosophie analytique du langage (5.1.1), ensuite sous celui de la grammaire (5.1.2) et finalement sous celui de la linguistique (5.1.3).

5.1.1 Études du défini en philosophie analytique

Le défini représentait un sujet intéressant pour les philosophes analytiques du langage. Un des plus anciens écrits consacrés au problème de l'interprétation des expressions définies est celui de Russell (1905). Cet article est, en quelque sorte, une réponse à Frege (1892).

En effet, Frege propose ce que certains appellent une sémantique à double niveau : les noms (qu'ils soient des noms propres ou des noms communs) ont chez lui à la fois un référent et un sens. Le référent d'un nom est l'entité du monde à laquelle le nom renvoie, alors que son sens est la manière dont ce référent est présenté. Par exemple, les expressions *le vainqueur de Iéna* et *le vaincu de Waterloo* ont tous les deux le même référent (à savoir Napoléon Bonaparte) mais ce référent est présenté de deux manières différentes.

En opposition à Frege, Russell propose une sémantique à un seul niveau : les signes n'ont chez lui qu'une dénotation. Il s'intéresse particulièrement à ce qu'il appelle les

descriptions définies sans en donner une définition précise ; il semble considérer que ce sont des expressions dont l'élément principal est un nom commun au singulier précédé par l'article défini. Il pense que, contrairement aux noms propres qui ont une véritable dénotation, les descriptions définies ne sont des expressions dénotantes qu'en apparence : en réalité, elles seraient des quantifications avec une double affirmation. Par exemple, *le roi de France* ne dénote rien dans (1) car (1) doit être analysé comme (2) :

(1) *Le roi de France est chauve.*

(2) *Il existe un et un seul individu qui soit roi de France et cet individu est chauve.*

On remarque que le syntagme nominal *le roi de France* est disparu au cours de l'analyse : (2) ne le contient plus. Ce qui est le plus intéressant pour nous dans cette analyse, c'est l'idée d'unicité que Russell attribue à ces expressions : si on utilise une description définie, c'est parce qu'il n'existe qu'une seule entité au monde qui puisse en être le référent. Il est conscient que ce critère d'unicité n'est pas conforme à l'usage : on peut dire par exemple *le frère de Untel* même si Untel a plusieurs frères. Il soutient toutefois qu'il serait plus correct de dire dans un cas pareil *un frère de Untel*.²⁰ ; apparemment, il ne prend pas en considération le contexte.

²⁰ Russell se situe dans une optique normative ici et non pas dans une optique descriptive.

D'autres philosophes du langage ont poursuivi la réflexion de Russell. Quine²¹ considère que, la plupart du temps, les expressions définies n'impliquent pas l'unicité du référent mais qu'il faut y ajouter des informations supplémentaires. Il y voit tout simplement des contraintes contextuelles pour résoudre les ambiguïtés de la langue.

Linsky (1974 : 162) écrit en parlant du pronom IL : « De toute évidence « à quoi IL fait-il référence ? » est une question dépourvue de sens tant que l'on ne précise pas le contexte ». Il défend la même position concernant les descriptions définies. En effet, on peut très bien faire référence à quelqu'un par l'expression *le vieux fou* même si cet homme n'est pas fou et, par ailleurs, il existe de nombreux vieux fous. Linsky soutient que c'est avant tout le contexte d'utilisation d'une expression que confère l'unicité du référent.

De nombreux autres auteurs ont enrichi le débat qu'a lancé Russell sur les DD : Carnap, Strawson, Grice, Kripke, Neale, etc. Nous ne pouvons pas étudier leurs œuvres ici mais le lecteur intéressé peut en trouver plusieurs réunis dans Oster-tag(1998).

²¹ Cf. Quine (1953) et Quine (1960).

5.1.2 Études du défini en grammaire

Les grammairiens ont également abordé la question du défini. Grevisse (1986 : 906 et ssq) considère que l'article défini s'emploie devant le nom qui désigne un être ou une chose connu du locuteur et de l'interlocuteur. Il illustre cette position par des exemples suivants :

Le soleil luit pour tout le monde. (réalité faisant partie de l'expérience commune)

*Donnez-moi **la** clé.* (réalité que la situation identifie parfaitement)

L'archevêque de Paris (il y en a un seul)

*J'ai pris **la** route qui conduit à Reims.* (le complément du nom, ici une relative, permet d'identifier la réalité)

*Aimer **les** livres* (tous les représentants de la catégorie)

*Marinette chantait une petite chanson à un poussin jaune qu'elle tenait sur les genoux. Tiens, dit **le** poussin en regardant du côté de la route, voilà un bœuf.* (le personnage a été présenté avant que l'on ne le désigne au moyen de l'article défini)

Grevisse remarque toutefois que l'article défini peut aussi s'employer quand on envisage une espèce, une catégorie et non seulement un individu : ***Le** chien est l'ami de l'homme.* Quant à l'article indéfini, il s'emploie devant un nom désignant un être ou une chose (ou des êtres et des choses) dont il n'a pas encore été question, qui ne sont pas présentés comme connus, comme identifiés.

Dans une autre grammaire, rédigée par Riegel *et al.* (1994 : 154 et ssq), les auteurs proposent une procédure afin d'identifier le référent d'une expression définie :

« L'article défini sert à référer à une entité identifiable à partir du seul contenu descriptif du reste du groupe nominal. On peut se représenter l'opération sémantique qu'il réalise de la façon suivante :

a) Le récepteur doit prendre en considération le signifié de l'ensemble formé par le nom et son expansion éventuelle ;

b) puis, dans l'espace référentiel constitué par la situation de discours, le contexte ou le savoir qu'il partage avec l'émetteur,

c) il doit effectuer la saisie (selon que l'article est au singulier ou au pluriel) du ou des référents qui sont les seuls à correspondre au signalement donné en a).

Autrement dit, l'article défini présuppose l'existence et l'unicité : il n'y a pas d'autre(s) référent(s) possible(s) qui vérifie(nt) la description de la réalité désignée par le groupe nominal.

La référence ainsi établie peut être spécifique, c'est-à-dire concerner un ou des individus particuliers, ou générique, c'est-à-dire concerner l'ensemble d'une classe ou d'une sous-classe d'individus. (...) De ce point de vue, certains énoncés peuvent être ambigus : *Le chien aboie* peut signifier qu'il y a actuellement un chien particulier en train d'aboyer, ou bien attribuer un type de cri à la classe entière des chiens. (...) La plupart du temps, le contexte ou la situation permet de lever l'ambiguïté. »

Nous croyons qu'une telle procédure, tout comme la position de Grevisse, peut permettre l'identification du référent de nombreux définis. Deux remarques s'imposent cependant. Premièrement, comme nous le verrons dans la section suivante, certaines expressions définies n'entrent pas dans ces catégories. Deuxièmement, ces considérations sont dépendantes du contexte. Malheureusement, en quoi ce contexte consiste n'est pas décrit de façon précise.

5.1.3 Études du défini en linguistique

Un des apports des linguistes par rapport aux philosophes du langage est l'introduction d'une distinction entre le **posé** et le **présupposé** : nous l'avons déjà présentée dans (4.2.3). Bien que cette opposition s'applique aux énoncés en général²², elle peut être très utile dans le cas des expressions définies. On peut ainsi admettre qu'un énoncé comme (1) mentionné dans (5.1.1) contient deux affirmations, une posée et une présupposée :

Présupposé de (1) : il existe un individu qui est roi de France

Posé de (1) : cet individu est chauve

Dans cette optique, le défini pourrait s'expliquer par le fait que l'existence d'un individu qui est roi de France fait partie du présupposé de l'énoncé.

En nous inspirant de St-Germain (1995 : 48), nous pouvons définir deux oppositions semblables : Donné/Nouveau et Connu/Inconnu.

Donné par rapport à un énoncé T : ce qui fait partie explicitement du contexte de T.

²² L'opposition posé ~ présupposé peut s'appliquer même aux définitions lexicographiques. Cf. Mel'čuk *et al.* (1995 : 102).

Nouveau par rapport à un énoncé T : ce qui ne fait pas partie explicitement du contexte de T.

Connu par l'auditeur A : ce qui fait partie des connaissances de A.

Inconnu pour l'auditeur A : ce qui ne fait pas partie des connaissances de A.

Dans un tel cadre, il ne serait pas nécessaire que le défini fasse partie du donné : il pourrait aussi faire partie du nouveau pour peu qu'il soit du domaine de connu.

Chafe (1976) donne un assez bon aperçu de la question de la définitude. L'auteur considère que les objets particuliers sont, le plus souvent, exprimés à travers une catégorisation. Or, il existe un très grand nombre d'objets qui peuvent entrer dans la même catégorie. Lorsqu'un locuteur utilise une expression définie, il croit que son auditeur peut trouver, identifier parmi tous les référents qui pourraient être catégorisés de la même façon, celui qu'il a en tête. Il ajoute même qu'il conviendrait mieux de parler d'identifiabilité plutôt que de définitude.

Chafe mentionne cinq considérations qui permettent au locuteur de croire que son auditeur est capable de trouver le référent de ce dont il parle.

- 1) Il existe **soit un référent unique, soit un seul référent saillant**, comme c'est le cas de la terre, de la lune, du ciel, etc.
- 2) Même s'il n'existe pas un seul référent saillant à l'intérieur d'une catégorie comme telle, **il existe un référent saillant à l'intérieur d'un contexte précis**. Dans

une salle de classe, on peut dire *le tableau* ; le référent en sera facilement identifiable. Un membre d'une famille précise peut dire à un autre membre *As-tu promené le chien ?* et on comprendra qu'il s'agit du chien de la famille.

3) Le référent d'une expression peut être établi à travers **une mention antérieure dans le discours**. On peut introduire un nouveau référent en disant *J'ai reçu une lettre ce matin* et ensuite dire tout simplement *la lettre* en considérant que dans ce contexte l'auditeur sera capable de l'identifier.

4) Le locuteur peut utiliser des **spécificateurs afin de créer une catégorie ad hoc** à un seul élément – même si le nom sans spécificateur ne permet pas de créer une telle catégorie. Par exemple, même si le référent de la description définie dans *J'ai rencontré le mécanicien* n'est pas identifiable, il peut l'être dans *J'ai rencontré le mécanicien qui a réparé notre voiture la semaine passée*.

5) Il existe des cas où **une entité entraîne (entail) une autre**. Après avoir dit *Nous avons visité une nouvelle maison hier*, le locuteur peut continuer par *La cuisine était très grande* puisque l'on peut s'attendre à ce qu'il y ait une cuisine dans une maison. Après avoir dit *J'ai vendu mon vélo hier*, on peut continuer par *Je me demande ce que je vais faire avec l'argent*.

Mel'čuk (1993, vol.II : 117ssq) présente le défini à partir de la catégorie de détermination qu'il définit ainsi : c'est une catégorie dont les éléments spécifient le référent de l'expression nominale en question ou la façon dont ce référent peut être identifié. Elle comprend, le plus souvent, le défini et l'indéfini – bien qu'il puisse y exister des déterminations plus complexes.

Le défini et l'indéfini peuvent tous les deux avoir une fonction individualisante et une fonction générique. La fonction individualisante du défini et celle de l'indéfini sont fondamentalement différentes l'une de l'autre : celle du défini signale un objet qui est déjà identifié dans l'esprit des locuteurs alors que celle de l'indéfini signale un objet qui n'est pas encore identifié. Par contre, la fonction générique rapproche le défini et l'indéfini, qui peuvent même être mutuellement substituables dans certains contextes, sans affecter le sens de l'énoncé : *Un/L'étudiant ordinaire n'aime pas se fatiguer.*

Mel'čuk (2001) examine le problème du défini à travers l'opposition entre **donné** et **nouveau**²³. Cette opposition est une des huit oppositions qui font partie de la Structure Sémantico-Communicative de la phrase. La Structure Sémantico-Communicative est une des quatre composantes de la Représentation Sémantique d'un énoncé qui décrit le sens propositionnel de cet énoncé.

Un segment de l'énoncé est considéré comme **donné** lorsque le locuteur l'expose comme étant déjà présent dans le conscient de son interlocuteur ou au moins comme étant facilement accessible à son conscient. Dans les langues qui ont une opposition entre le défini et l'indéfini, *grosso modo*, le donné correspond en général (mais pas

²³ La catégorie de Donné contient trois valeurs : donné, nouveau et non-applicable. Nous laissons de côté ici cette troisième valeur.

toujours) au défini et le nouveau à l'indéfini. Par exemple, dans (3) *livre* et *table* sont considérés comme donnés, alors que dans (4), seulement *table* est donné alors que *livre* est nouveau.

(3) *Le livre est sur la table.*

(4) *Il y a un livre sur la table.*

Contrairement à l'opposition thème-rhème qui est orientée vers l'énonciateur, l'opposition donné-nouveau est orientée vers l'auditeur. L'énonciateur choisit entre le donné ou le nouveau en fonction de ce qu'il croit être ou bien déjà présent dans le conscient de son auditeur ou bien facilement accessible à son conscient. Cela signifie que l'on a besoin d'un modèle qui décrit le conscient humain pour rendre compte clairement de l'opposition donné-nouveau. Étant donné qu'il n'existe pas encore un tel modèle tout ce que l'on peut dire concernant cette opposition reste provisoire et quelque peu intuitif.

L'énonciateur choisit d'attribuer le trait défini à une expression E en fonction de la prévisibilité de E par son auditeur A. La **prévisibilité** de l'expression E pour l'auditeur A signifie que A est capable, d'une part, de prévoir l'apparition de E dans le discours, d'autre part, d'identifier le référent de E sans ambiguïté. On peut distinguer deux types de prévisibilité : prévisibilité à partir du contexte linguistique et prévisibilité à partir du contexte extralinguistique.

L'expression E est **prévisible à partir du contexte linguistique** si le locuteur peut considérer que, à partir de ce qui précède E dans le discours, l'auditeur peut prévoir l'apparition de E et peut trouver son référent. Cette prévisibilité peut être basée aussi bien sur les connaissances linguistiques que sur les connaissances extralinguistiques. Lorsqu'elle est basée sur les connaissances linguistiques, elle peut être directe, c'est-à-dire qu'elle peut consister à répéter le même nominal (*J'ai vu un livre hier... Le livre était très intéressant.*) ou bien elle peut être indirecte, c'est-à-dire qu'elle peut faire appel à des relations lexicales. On peut illustrer ce dernier cas par l'exemple suivant :

(5) *Un peloton américain a été attaqué ce matin à l'Est de Bagdad. Les soldats ont dû ouvrir le feu sur les assaillants.*

Même si les soldats et les assaillants n'étaient pas littéralement mentionnés dans le discours auparavant, ils sont prévisibles à cause des relations lexicales qu'ils ont avec d'autres lexies déjà mentionnées : SOLDAT est le premier actant typique de PELOTON et ASSAILLANT est le premier actant typique de ATTAQUER.

La prévisibilité à partir du contexte linguistique peut être basée sur les connaissances extralinguistiques aussi. Par exemple, dans un discours qui parle de Paris, il est prévisible de rencontrer l'occurrence de l'expression *la capitale de la France* : nos connaissances nous indiquent en effet que les deux ont le même référent.

Une expression E est dite **prévisible à partir du contexte extralinguistique** si, sans être mentionnée auparavant, l'auditeur peut prévoir son apparition et peut trouver son référent à l'aide du contexte extralinguistique. Cela peut être le cas lorsque le référent de l'expression est présent dans la situation de l'énonciation (*Pose-le sur la table !*) ou lorsque l'auditeur possède des connaissances extralinguistiques appropriées :

(6) *Ce midi, j'ai rencontré le doyen de la faculté.*

Par défaut, il s'agit de la faculté à laquelle l'énonciateur est affilié : pour l'identifier, l'auditeur a besoin des connaissances sur l'énonciateur.

5.2 Caractéristiques du défini

Dans cette section, nous examinerons d'abord les trois types de rapport possibles entre le défini et l'indéfini : opposition effective, opposition neutralisée et absence de l'opposition (5.2.1). Dans le cas des locutions, collocations et expressions génériques, l'opposition a tendance à être soit absente, soit neutralisée (5.2.2). Ce qui a le plus d'intérêt pour nous, ce sont des expressions libres où il existe une opposition effective entre le défini et l'indéfini (5.2.3). Pour conclure, nous soulignons l'importance du contexte dans l'identification du défini (5.2.4).

5.2.1 Trois types de rapport entre le défini et l'indéfini

En matière d'opposition, nous pouvons distinguer trois types de rapport entre le défini et l'indéfini : opposition effective, opposition neutralisée ou absence d'opposition.

5.2.1.1 Opposition effective

Il y a **opposition effective** entre le défini et l'indéfini dans les énoncés E_1 et E_2 si ces énoncés remplissent les deux critères suivants :

- (C1) la seule différence entre les signifiants de ces énoncés est que E_1 contient l'article défini là où E_2 contient l'article indéfini ;
- (C2) le signifié de E_1 est différent du signifié de E_2 – et cette différence est liée au changement d'article.

Exemple :

(7a) *Pierre a visité un appartement.*

(7b) *Pierre a visité l'appartement.*

Il est clair que le sens de ces deux énoncés n'est pas le même : dans le premier, on parle d'un appartement non encore identifié par le destinataire alors que dans le se-

cond, l'appartement doit déjà être identifié. Il est également clair que cette différence de sens est liée au changement de l'article seulement.

5.2.1.2 Opposition neutralisée

Il y a **opposition neutralisée** entre le défini et l'indéfini dans les énoncés E_1 et E_2 si ces énoncés remplissent le critère (C1) sans remplir (C2). Exemple :

(8a) *Jean a **le** profond désir de partir.*

(8b) *Jean a **un** profond désir de partir.*

Ces énoncés remplissent le critère (C1) : la seule différence entre leur signifiant est que (8a) contient le défini là où (8b) contient l'article indéfini. Toutefois, le sens de ces énoncés est le même : on ne peut discerner de différence ni dans le degré du désir de partir de Jean, ni dans le caractère connu ou inconnu de ce désir.

5.2.1.3 Absence d'opposition

Il y a **absence d'opposition** entre le défini et l'indéfini si, pour un énoncé E_1 , on ne peut pas trouver un énoncé E_2 qui remplisse le critère (C1). Exemple :

(9a) *Il est souvent dans la lune.*

(9b) **Il est souvent dans une lune.*

Nous devons toutefois apporter une petite précision ici. Considérons les exemples suivants :

(10a) *Pierre a regardé la télévision hier soir.*

(10b) *Pierre a regardé une télévision hier soir.*

L'exemple (10a) est – potentiellement – ambigu. L'interprétation la plus évidente est qu'il contient une collocation²⁴ : *regarder la télévision*. Toutefois, il peut avoir aussi une deuxième interprétation, sans doute moins fréquente, où on aurait une expression libre à la place de la collocation. Dans (10b), il ne peut s'agir que d'une expression libre, comme c'est le cas par exemple dans (11).

(11) *Pierre a regardé le/un frigo hier soir.*

Tout cela signifie que l'opposition entre le défini et l'indéfini est absente lorsque l'on comprend (10a) comme une collocation ; et qu'elle est effective si on le comprend comme une expression libre.

²⁴ Cette collocation correspond aussi à une fonction lexicale : $\text{Real}_3(\text{télévision}) = \text{regarder}$.

5.2.2 Défini et indéfini dans les locutions, les collocations et les expressions génériques

Les locutions, les collocations et les expressions génériques présentent des caractéristiques particulières pour ce qui a trait au défini et à l'indéfini : elles ont tendance à limiter, à contrôler, à des degrés différents, le choix de l'article. Nous examinons d'abord le comportement des locutions et des collocations face aux choix de l'article, ensuite celui des expressions génériques.

5.2.2.1 Choix du défini et de l'indéfini dans les locutions et les collocations

Les locutions ont déjà été présentées dans (3.1.1) ; ce sont des expressions entièrement idiomatiques. Comme rappel, nous proposons deux séries d'exemples, la première contient des locutions avec l'article défini, la seconde des locutions avec l'article indéfini :

(12a) *tomber dans **les** pommes, jeter l'éponge, s'arracher **les** cheveux, c'est **la** croix et **la** bannière, mener **la** barque, sauver/perdre **la** face, le jeu n'en vaut pas **la** chandelle, être dans **la** lune, mettre quelqu'un à l'ombre, clouer quelqu'un **au** pilori, briser **la** glace, faire **la** fête, faire l'amour...*

(12b) *se noyer dans **un** verre d'eau, avoir **un** poil dans la main, monter **un** bateau à quelqu'un...*

Les collocations ont aussi été présentées, dans (3.2.4.4) ; ce sont des expressions semi idiomatiques. Nous en proposons également deux séries d'exemples, d'abord avec l'article défini, ensuite avec l'article indéfini :

(13a) *avoir **le** désir de (+Infinitif), perdre **la** tête, trouver **la** mort, prendre **le** départ, se brûler **la** cervelle, etc.*

(13b) *faire **une** promenade, prendre **une** marche, dormir comme **un** loir, pleurer comme **une** madeleine, être bavard comme **une** pie, se battre comme **un** lion, etc.*

Avant d'examiner le comportement des locutions et des collocations face au choix de l'article, signalons qu'en principe, il existe des critères précis pour les distinguer les unes des autres. Cependant, en pratique, on peut aussi rencontrer des cas litigieux :

(14) *poser **un** lapin*

En effet, certains locuteurs jugent que les énoncés *Encore un lapin !* ou *Je ne veux pas de lapin cette fois-ci !* sont acceptables. Cela suggère qu'il existe une lexie LAPIN qui signifie approximativement 'rendez-vous manqué' et par conséquent (14) est une collocation. D'autres locuteurs jugent que ces énoncés ne sont pas acceptables : dans ce cas, (14) serait plutôt une locution.

Revenons maintenant à la question de l'article défini vs indéfini. Les locutions ont un comportement rigide dans la phrase, notamment, elles ont tendance à contrôler le choix de l'article (défini ou indéfini). On peut dire par exemple :

(15a) *Susanne est tombée dans les pommes.*

(16a) *Jacques a jeté l'éponge.*

(17a) *Yves a sauvé la face.*

Par contre, (15b), (16b) et (17b) ne sont pas des phrases naturelles contenant des collocations :

(15b) **Susanne est tombée dans des pommes.*

(16b) **Jacques a jeté une éponge.*

(17b) **Yves a sauvé une face.*

Certes, une phrase comme (18) est tout à fait naturelle mais elle contient une expression libre qui n'a strictement rien à voir avec l'expression idiomatique *jeter l'éponge* :

(18) *J'ai jeté une éponge dans la poubelle ce matin, elle était trop vieille.*

Contrairement aux locutions, les collocations ont un comportement un peu plus libre concernant le choix de l'article. Il est vrai que sans l'ajout d'un modificateur, on ne

pourrait pas avoir l'indéfini dans les collocations de (13a) ni le défini dans celles dans (13b) ; c'est le point commun entre elles. Toutefois, nous pouvons encore les diviser en trois sous-catégories en fonction de leur comportement avec un modificateur (par exemple, un adjectif qualificatif ou une proposition subordonnée). Ces trois sous-catégories de collocations correspondent aux trois types de rapport entre le défini et l'indéfini, mentionnés dans (5.2.1) : collocations avec absence d'opposition, collocations avec opposition neutralisée et collocations avec opposition effective.

Absence d'opposition dans les collocations

Il existe des collocations qui, avec ou sans modificateur, gardent toujours le même article :

(19) *perdre **la** tête, se brûler **la** cervelle, prendre **une** marche*

Certes, on peut dire, en plaisantant, *Il s'est brûlé sa petite cervelle de moineau*, mais cette expression peut être considérée comme une forme de défini – l'indéfini ne serait pas possible ici.

Il existe d'autres collocations où la présence d'un modificateur entraîne un changement d'article obligatoire :

(20a) *Deux passagers ont trouvé **la** mort dans cet accident.*

(20b) *Deux passagers ont trouvé **une** mort atroce dans cet accident.*

(21a) *Pierre a fait **une** promenade.*

(21b) *Pierre a fait **la** promenade que Marie lui a conseillée.*

Opposition neutralisée dans les collocations

Il existe des collocations où la présence d'un modificateur permet un changement d'article sans toutefois le rendre obligatoire – autrement dit, le changement d'article est optionnel :

(22a) *Jean a l'intense désir de partir.*

(22b) *Jean a **un** intense désir de partir.*

Comme nous l'avons mentionné plus haut, le contenu de ces deux énoncés est identique : il n'y a pas de différence quant aux informations que les interlocuteurs doivent posséder sur le désir de partir de Jean.

Opposition effective dans les collocations

Les cas où le défini et l'indéfini s'opposent à l'intérieur d'une locution sont certainement rares et les exemples que nous avons trouvés ne sont pas jugés de façon unanime :

(23a) *Jean fait **la** sieste dans sa chambre.*

(23b) *Jean fait **une** sieste dans sa chambre.*

(24a) *Marie prend **le** café sur la terrasse.*

(24b) *Marie prend **un** café sur la terrasse.*

Selon le jugement de certains locuteurs, dans ces exemples, le défini semble suggérer une activité régulière alors que l'indéfini une activité ponctuelle. D'autres locuteurs estiment que ces exemples ont le même sens et, par conséquent, ne représentent pas d'opposition.

5.2.2.2 Choix du défini et de l'indéfini dans les expressions génériques

On peut rencontrer des occurrences de groupes nominaux définis qui ont une valeur générique.

(25) *Pierre aime **le** fromage.*

Affirmer (25) revient à dire que Pierre aime, éventuellement à quelques exceptions près, tous les fromages. Dans cet exemple, seul le défini est possible avec la valeur générique : si on remplaçait l'article défini par l'article indéfini, on obtiendrait une expression à valeur particularisante. Dans l'exemple ci-dessous, le défini et

l'indéfini sont également possibles avec une interprétation générique et ont le même sens :

(26a) *Un Alsacien (en général) mange de la choucroute.*

(26b) *L'Alsacien (en général) mange de la choucroute.*

Dans le cas des expressions à valeur générique, comme dans le cas des locutions, le défini et l'indéfini ont tendance à ne pas s'opposer. La tâche principale concernant l'interprétation de ces expressions consiste principalement à les reconnaître comme des expressions génériques.

5.2.3 Le défini dans les expressions libres

Nous venons de voir que dans le cas des locutions, collocations et expressions génériques, le défini et l'indéfini ont tendance à ne pas s'opposer. Il en va tout autrement dans le cas des expressions libres à valeur particularisante : elles représentent une opposition effective entre le défini et l'indéfini. Nous examinons d'abord cette opposition et présentons ensuite les descriptions définies : ces dernières sont une sous-catégorie des expressions définies et c'est à elles que se sont intéressés avant tout les grammairiens, les philosophes et même une partie des linguistes.

5.2.3.1 Opposition entre défini et indéfini dans les expressions libres

Lorsque l'on trouve une opposition effective entre le défini et l'indéfini, alors le défini est signe des connaissances suffisantes que les interlocuteurs doivent posséder pour identifier l'entité dénotée par l'expression définie, et l'indéfini marquera l'absence de connaissances suffisantes sur l'entité. Pour illustrer cette opposition, examinons les énoncés suivants :

(27a) *J'ai mangé **la** pomme.*

(27b) *J'ai mangé **une** pomme.*

Dans ces deux exemples, on parle d'une pomme bien spécifique, mais le choix du locuteur montre qu'il considère que les informations que les interlocuteurs possèdent sur cette pomme sont différentes. Dans (27a), l'article défini signale qu'il ne doit y avoir qu'une seule pomme qui puisse être considérée : elle doit être identifiée parmi toutes les pommes possibles. Dans (27b), l'article indéfini indique que la pomme en question n'est pas encore identifiable : l'auditeur sait qu'il s'agit d'une pomme mais il ne sait pas précisément de laquelle. L'énonciateur ne produirait pas (27a) s'il estimait que son auditeur ne pourrait pas encore identifier la pomme dont il parle. De la même façon, il ne produirait pas (27b) si la pomme était déjà identifiée.

On pourrait donner une infinité d'exemples où l'article défini indique que le référent de l'expression doit être identifiable. Nous en proposons trois autres :

(28) *Marie a acheté **la** voiture.*

(29) ***Le** soldat est entré dans **la** maison.*

(30) ***Le** camion a renversé **la** vieille dame.*

Dans les énoncés (27a) et (28-30), les référents des six expressions (*la pomme, la voiture, le soldat, la maison, le camion et la vieille dame*) doivent être identifiés dans l'esprit des interlocuteurs de façon univoque. Ils doivent savoir de quelle pomme, voiture, etc. il est question dans l'énoncé. En remplaçant l'article défini dans ces exemples par l'article indéfini, les phrases obtenues seraient tout à fait grammaticales mais les énoncés ne pourraient pas apparaître dans les mêmes contextes : à savoir dans des contextes où le référent est identifiable.

En résumé : lorsqu'il existe une opposition effective entre le défini et l'indéfini, le défini indique que les interlocuteurs possèdent suffisamment d'informations sur le référent de l'occurrence de l'expression pour l'identifier alors que l'indéfini indique l'absence de telles informations. Le destinataire a un travail supplémentaire à faire avec une expression définie de ce type : trouver les informations nécessaires à l'identification du référent.

5.2.3.2 Descriptions définies

Les descriptions définies (DD) peuvent être considérées comme un cas particulier d'expressions définies. Ce sont elles qui intéressent le plus les philosophes et même les grammairiens. Toutefois, elles étaient traitées soit sans une définition précise, soit avec une définition précise mais inexacte. Nous avons déjà mentionné que Russell en parle mais il n'en donne pas une véritable définition. Ducrot et Schaeffer (1995 : 365), en s'inspirant de Russell, les définissent d'un point de vue syntaxique seulement. Cette définition peut être notre point de départ mais elle doit être précisée encore. En effet, une telle définition n'aurait pas vraiment d'utilité parce qu'elle inclut les expressions beaucoup trop hétérogènes, notamment les collocations (*Ils ont perdu la tête hier*), les locutions (*Il a jeté l'éponge*), les expressions génériques (*Marie adore le fromage*) et même le partitif (*Il a de la difficulté à partir*). Il est clair pourtant que ce que Russell et les autres philosophes ont en tête n'inclut pas ces quatre types d'expressions. Nous proposons de définir les descriptions définies à la fois d'un point de vue syntaxique et sémantique.

D'un point de vue syntaxique, une **description définie** est une occurrence d'un groupe nominal dont la tête, c'est-à-dire le nominal principal, est déterminée par un article défini (*la maison, la maison où j'habite*, etc.). On étend généralement cette définition en exigeant seulement qu'il y ait une paraphrase au moyen d'un groupe nominal dont la tête est déterminée par un article défini. On peut considérer comme une description définie un nominal introduit par un possessif (en interprétant *ma*

maison par *la maison qui est à moi* ou même par *la maison où j'habite*) ou par un démonstratif (*cette maison* par *la maison que voici*). D'un point de vue sémantique, c'est une expression libre à valeur particularisante : autrement dit, elle n'est ni une locution, ni une collocation, ni une expression à valeur générique et ni un nom partitif.

5.2.4 Identification contextuelle du référent d'un défini

Dans les écrits qui traitent du défini, on rencontre souvent des exemples qui illustrent comment son référent peut être identifié. Toutefois, ces exemples sont donnés surtout à titre d'illustration et ne constituent pas une méthode exhaustive d'identification. Nous soutenons que cette identification ne peut se faire qu'à l'aide d'un contexte approprié tel qu'il est défini et illustré par les exemples dans (4.4).

Nous pouvons être d'accord avec Chafe (1976) pour dire que lorsqu'un locuteur utilise une DD, il croit que son auditeur peut trouver, identifier parmi tous les référents qui pourraient être catégorisés de la même façon, celui qu'il a en tête. Les cinq considérations qu'il mentionne pour identifier le référent ne posent guère problème en soi – ce qui est problématique, c'est de déterminer de façon précise comment cette identification se fait.

C'est ici que la notion de contexte entre en jeu. Les considérations proposées par Chafe attribuent, dans le processus d'identification des DD, un rôle essentiel au

contexte, quoique à des degrés différents. Dans les deux premiers cas (référent saillant), l'identification est clairement contextuelle, la différence réside simplement dans le degré de généralité du contexte : il suffit d'aller d'une salle de classe à l'autre pour que le référent de la DD *le tableau* change alors qu'il faudrait aller sur une autre planète pour changer celle de *la lune*²⁵. Le troisième cas (mention antérieure) relève tout simplement du contexte linguistique. Dans le quatrième cas (spécification), l'identification ne peut se faire que lorsqu'il y a exactement une entité dans le contexte qui corresponde à la DD spécifiée. Nous pouvons même considérer le cinquième cas (implication) comme un aspect particulier du contexte : la nomination d'une première entité dans un énoncé introduit dans le contexte une deuxième entité qui sera le référent de la DD.

Tout cela revient à dire que l'utilisation (la production aussi bien que l'interprétation) d'une DD présuppose un contexte qui contient exactement une seule entité susceptible d'être le référent de cette DD. Toutefois, avec cette affirmation – assez triviale, admettons-le – nous sommes encore loin de régler la question d'identification des DD. À vrai dire, notre travail ne fait que commencer car il nous faut déterminer ce contexte.

²⁵ On pourrait remarquer qu'il faut distinguer deux lexies : LUNE ('satellite de la terre') et LUNE [de X] ('satellite de la planète X').

Ce que nous tenons à souligner ici est le fait suivant : si le référent d'une DD – ou même d'une expression définie en général – est unique non pas dans le monde mais dans le contexte de l'énonciation, alors l'identification de ce référent doit se faire non pas par rapport à nos connaissances du monde (par exemple, par rapport à un modèle, une représentation de ce monde) mais selon le contexte de l'énonciation.

5.3 Introduction aux analyses de l'interprétation du défini

Après avoir examiné ce que les grammairiens, les philosophes et les linguistes ont écrit sur le défini et après avoir présenté notre propre classification du défini, il convient de mettre toutes ces considérations en pratique et analyser ce type d'expression. En quoi les analyses des expressions définies que nous menons contribuent à mieux comprendre la place des CL dans l'interprétation des énoncés ? Sur quoi portent ces analyses ? Selon quelle méthodologie sont-elles faites ? Telles sont les questions que nous tenons à examiner au préalable concernant les analyses.

5.3.1 Intérêt des analyses du défini pour notre recherche

Nous avons vu que dans le cas des locutions et des collocations, les contraintes lexicales sont, en très grande partie, responsables du choix de l'article défini ou indéfini. Dans le cas du générique, il n'y a pas de véritable opposition entre le défini et l'indéfini. La situation est tout à fait différente avec les DD. Lors de la génération

d'un énoncé contenant une DD, l'énonciateur doit prendre en compte – entre autres – les CL et les CM de son interlocuteur pour avoir choisi le défini plutôt que l'indéfini. Pareillement, lors de l'interprétation de cet énoncé, l'auditeur doit utiliser ces connaissances pour identifier le référent de la DD. Le véritable intérêt des DD commence pour nous lorsque nous tentons de répondre à la question suivante : quelles sont les connaissances nécessaires pour créer un contexte propice à identifier le référent d'une DD ?

En effet, c'est ici que l'on peut constater les divergences d'opinion entre des chercheurs de différentes disciplines, et même parfois entre des chercheurs travaillant à l'intérieur de la même discipline. Devons-nous avoir recours de façon massive aux connaissances du monde, comme le proposent des spécialistes en Intelligence Artificielle ? Ou bien, disposons-nous d'un modèle linguistique qui pourrait prendre en charge une bonne partie des connaissances nécessaires ? Plus exactement, quelle est la part respective de ces deux types de connaissances ?

Nous pouvons illustrer ces questions un peu abstraites par des exemples concrets :

(31) *Pierre a visité l'Angleterre. Il a bien aimé Oxford mais il a trouvé **la capitale** trop bruyante.*

(32) *Véronique est banlieusarde mais **son père** est né à Paris.*

(33) *Nicole est allée à la gare de Fontainebleau ce matin et elle a rencontré **le chef de gare**.*

Même s'il n'en a jamais été question auparavant, le locuteur considère que le référent des expressions *la capitale, son père, le chef de gare* apparaissant dans (31), (32) et (33) est déjà identifiable par le destinataire. Comment rendre compte de ces données ? Plus précisément : comment se fait-il que ces éléments soient déjà identifiables ? Les chercheurs en IA auraient tendance à dire que ce sont nos CM qui nous indiquent que dans une gare, il y a un chef de gare, qu'un pays a une capitale, une personne a un père, etc. Toutefois, bien que cela soit l'analyse la plus répandue, il existe une analyse alternative à l'aide des connaissances lexicales qui pourraient couvrir une bonne partie des connaissances du monde. Ainsi, l'intérêt de se pencher sur l'interprétation des DD est de pouvoir examiner le rôle de ces deux types de connaissances.

Nous devons faire deux grands choix concernant les définis à analyser : seulement les descriptions définies proprement dites ou les autres définis aussi ? Seulement les définis qui ont un lien (lexical ou conceptuel) avec le lieu étudié ou les autres définis aussi ? Nous avons opté pour l'analyse de tous les définis. Ce choix se justifie en premier lieu par son utilité : lorsque l'on veut faire du TAL, on doit – comme but ultime – rendre les machines capables de traiter l'intégralité des textes. Cela implique, entre autre, qu'elles doivent distinguer entre les expressions libres et les expressions idiomatiques. Par exemple, la machine doit savoir que *brouiller les cartes* est une locution (bien qu'il existe aussi une expression libre avec le même signifiant).

5.3.2 But et portée des analyses

Le but de cette analyse est d'évaluer le rôle, l'importance des CL dans l'interprétation du défini. Telle que nous l'avons définie plus haut, l'interprétation d'un énoncé inclut aussi bien l'identification de son sens que l'identification de son référent. Nous considérons comme évident que, dans l'identification du sens d'un énoncé, les CL jouent un rôle primordial ; il reste à déterminer leur rôle exact et l'éventuel rôle des CM. Nous nous intéressons surtout à l'identification du référent des expressions définies. Bien entendu, cette identification ne se fait pas ici par rapport à une base de connaissances représentant un modèle de monde (faute de disposer d'une telle base) mais en (re)construisant un contexte (voir 4.4) où l'occurrence de l'expression peut être considéré comme n'ayant qu'un seul référent.

Nous ne considérons pas les noms propres précédés d'un article défini (*la France, la SNCF*) car – bien qu'ils soient, eux aussi, des lexies de la langue – leur rapport à leur référent est fondamentalement différent de celui des noms communs. Les grammairiens considèrent en effet que les noms propres ne conviennent qu'à une seule entité²⁶. On pourrait objecter à cette affirmation que les noms propres comme *Pierre* peuvent désigner de nombreuses personnes. La Grammaire de Port-Royal (cf. Pariente (1985)) répond à une telle objection en disant que la pluralité de référent dans le cas des noms propres est accidentelle, alors qu'elle est essentielle pour les

²⁶ Cf. Ducrot et Schaeffer (1995 : 367).

noms communs. En effet, les noms propres font appel à une connaissance directe de l'entité (personne, lieu, organisation, etc.) dénotée. Ils ne pourraient s'employer avec l'indéfini que s'ils ont un modificateur (*la/*une France* vs. *la/une France ravagée par la guerre*).

5.3.3 Méthodologie des analyses

Nous faisons d'abord une présentation générale de notre méthodologie, incluant la notion d'ancre, les parties qui composent les analyses et quelques questions de formalisme (5.3.3.1). Ensuite, nous décrivons les huit types d'interprétation du défini (5.3.3.2).

5.3.3.1 Présentation générale de la méthodologie

Notre corpus est constitué de douze textes, trouvés sur Internet, la plupart du temps des faits divers. Dans chaque texte, nous analysons les expressions définies. Les textes et les expressions définies sont numérotés. Par exemple, [2.7] renvoie au septième défini du deuxième texte.

Le but de ces analyses est d'identifier, pour chaque occurrence de l'expression *le X*, le référent de *X*. Il arrive, notamment dans le cas des expressions idiomatiques, que l'on ne puisse pas vraiment parler de référent : il serait insolite de demander de quel-

les pommes il s'agit dans la locution *tomber dans les pommes*. Dans un tel cas, nous tenons tout simplement à signaler qu'il s'agit d'une locution.

Nous avons déjà mentionné que l'interprétation du défini se fait par rapport au contexte d'énonciation. Pour rendre le défini unique dans le contexte, nous proposons d'introduire la notion d'ancre : l'**ancre** A d'un défini D est une partie du contexte de D qui, en vertu de ses liens avec D, le rend unique dans le contexte de l'énonciation. Les liens entre A et D doivent remplir au moins une des deux conditions suivantes :

- 1) A et D se trouvent dans le même groupe nominal et ils sont liés par un lien syntaxique. Ce lien est obligatoire pour les définis du type V (ces types sont présentés un peu plus bas) mais il peut également exister pour les types II et III.
- 2) A et D sont liés par une relation anaphorique (type I), par une relation actancielle (type II) ou par une fonction lexicale autre que **Syn**, **Gener** ou **S_i** (type III).

Le contexte du défini, suivant notre définition donnée en (4.4.1), a une partie linguistique et une partie non linguistique mais nous allons nous concentrer sur la partie linguistique, étant donné que nous nous intéressons principalement à la place des connaissances lexicales dans l'interprétation du défini.

Dans les analyses, le nominal principal des expressions définies et son déterminant sont soulignés alors que l'ancre est en gras. Par exemple, dans *la gare **d'Ottawa***

[4.9] *gare* est le nominal principal et *Ottawa* est l'ancre : dans ce contexte, il n'y a qu'une seule gare d'Ottawa.

Deux remarques s'imposent concernant l'ancre :

- 1) Un défini peut avoir plusieurs ancres. Par exemple, dans *les adieux de M. Chrétien à M. Trudeau* [4.8], nous avons deux ancres. Certes, chacune d'entre elles contient des informations importantes sur le défini mais une seule pourrait le rendre unique dans le contexte.
- 2) Les ancres ne contiennent pas nécessairement toutes nos connaissances sur le référent des définies. Le lecteur peut avoir toutes sortes de CM sur telle ou telle gare par exemple ou il peut y avoir des informations supplémentaires dans le contexte.

Une fois la notion d'ancre décrite, nous pouvons présenter les trois parties dont les analyses sont constituées : A), B) et C).

La partie A) est composée de deux sous-parties dont la deuxième peut être nulle :

A1) Le syntagme où se trouve le défini.

A2) Une partie du cotexte (à l'extérieur du syntagme du défini) qui contient l'ancre du défini, si une telle partie existe.

La partie B) contient des informations linguistiques sur le défini et sur l'ancre que nous jugeons pertinentes pour l'interprétation du défini, suivant la stratégie de recherche donnée pour l'interprétation de chaque défini.

La partie C) contient la classification du défini. Nous distinguons huit types principaux dans l'interprétation du défini : défini anaphorique, défini avec lien actanciel, défini par lien de FL, défini idiomatique, défini avec spécification, défini déictique, défini interprété de façon générique et défini avec identification cotextuelle plus complexe. Un peu plus loin, nous proposons une présentation détaillée de chacun de ces types.

Nous pouvons illustrer ces trois parties par l'extrait suivant :

[2.8]

A1) *les voyageurs*

A2) *les 650 passagers du TGV* [2.3]

B) VOYAGEUR et PASSAGER sont des quasi-synonymes.

C) Type Ib.

Dans (3.1.2), nous avons mentionné que la règle de la forme propositionnelle dit que pour définir une lexie prédicative L, il faut utiliser une forme propositionnelle où les variables représentent les actants sémantiques de L. Par exemple, pour définir REPROCHER, il faut définir *X reproche Y à Z*. Cette règle nous sera très utile pour le type Iib et Iic. Dans le cas des lexies qui sont déjà dans le DiCo, nous reprenons

cette forme propositionnelle telle qu'elle se trouve dans le champ de caractérisation sémantique.

5.3.3.2 Huit types d'identification du défini

Nous présentons ici les huit types d'identification du défini que nous distinguons dans nos analyses. La moitié de ces types est divisée en sous-types. La présentation de chaque type et sous-type contient trois parties : la description générale, un exemple et une stratégie de recherche sommaire qui décrit l'identification du défini de ce type.

Malheureusement, nous n'avons pas de méta-algorithme qui permettrait de classer chaque défini dans un de ces types. Certes, nos connaissances actuelles nous pourraient nous permettre de reconnaître certains types (surtout les types I, II, III et IV) mais pas tous.

I : Défini anaphorique

Lorsque le défini est anaphorique, il existe une ancre, mentionnée dans A2), qui précède le défini et qui a le même référent que ce défini. On peut distinguer trois sous-types ici, en fonction de la relation entre le défini et l'ancre : identique, synonymique ou générique.

Ia : Nous disons qu'il y a une **anaphore identique** lorsque le nominal principal du défini est identique au nominal principal de l'ancre.

Exemple : *le TGV Paris-Bourg-Saint-Maurice... le TGV* [2.10]. La même lexie (TGV) est répétée et le référent de la deuxième occurrence de *TGV* est le même que celui de la première.

Stratégie de recherche : chercher dans le texte qui précède une expression dont la tête est le même nominal que le nominal principal du défini. Si l'expression trouvée n'est pas identique au défini, il faut vérifier si leurs modificateurs sont compatibles : par exemple, les expressions *le président Bush* et *le président Chirac* ne peuvent pas être coréférentielles – bien qu'elles aient le même nominal principal.

Ib : Nous disons qu'il y a une **anaphore synonymique** lorsque le nominal principal du défini est un (quasi-)synonyme du nominal principal de l'ancre. Dans le cas d'un tel type de défini, il existe une relation lexicale entre le défini et l'ancre qui est marquée dans le dictionnaire.

Exemple : *les passagers... les voyageurs* [2.9]. VOYAGEUR et PASSAGER sont des lexies quasi-synonymiques et, dans cet exemple, leur occurrence a le même référent.

Stratégie de recherche : chercher dans le texte qui précède une expression dont la tête est un synonyme du nominal principal du défini et vérifier si les deux peuvent être coréférentiels.

Ic : Nous disons qu'il y a une **anaphore hyperonymique** lorsque le nominal principal du défini est un hyperonyme du nominal principal de l'ancre ou l'inverse – et les deux ont le même référent.

Exemple : *le collègue... l'établissement* [6.16]. La lexie ÉTABLISSEMENT (D'ENSEIGNEMENT) est un hyperonyme de COLLÈGE mais les occurrences de ces deux lexies ont le même référent dans notre exemple.

Stratégie de recherche : chercher dans le texte qui précède une expression dont la tête est un hyperonyme (ou hyponyme) du nominal principal du défini et vérifier si les deux peuvent être coréférentiels.

II : Défini et ancre ayant un lien actanciel

Nous distinguons trois types de lien actanciel entre le défini et l'ancre : soit l'ancre est une valeur d'une FL actancielle du défini, soit l'ancre occupe à la fois une position d'actant sémantique et d'actant syntaxique du défini, soit l'ancre occupe une position d'actant sémantique du défini sans en être un actant syntaxique.

IIa : Le défini et l'ancre sont liés par une FL actancielle : en général, c'est l'ancre qui est le nom de l'actant typique du défini. L'information nécessaire à l'interprétation du défini se trouve directement dans un article de dictionnaire. Contrairement aux types Ib et Ic, le défini et l'ancre n'ont pas le même référent.

Exemple : *les trains en gare* [1.1]. En effet, $S_1(gare) = train$, autrement dit, TRAIN est le nom typique du premier actant de GARE. Cette relation est incluse dans l'article de dictionnaire GARE.

Stratégie de recherche : chercher dans le texte, aussi bien parmi les actants syntaxiques de surface (ASynS) qu'ailleurs, les noms typiques des actants du défini et vérifier si l'expression ainsi trouvée et le défini peuvent être coréférentiels.

I**b** : L'ancre occupe à la fois une position d'actant sémantique du défini et une position d'actant syntaxique profond. Contrairement au sous-type précédent (qui pourraient être vu comme un type plus spécifique de celui-ci), le lien entre le défini et l'ancre n'est pas mentionné directement dans le dictionnaire.

Exemple : *la gare d'Ottawa* [4.9]. Suivant le DiCo, la forme propositionnelle (FP) de la définition de GARE est :

~ du lieu Y pour le train X.

Dans notre exemple, *Ottawa* (l'ancre) occupe la deuxième position actancielle (lieu Y) de *la gare* (le défini). Bien entendu, dans l'article GARE, on ne trouvera pas OTTAWA, seulement la position actancielle Y et son étiquette sémantique.

Stratégie de recherche : chercher parmi les dépendances syntaxiques du défini celles qui pourraient être des actants sémantiques. Pour ce faire, il faut regarder dans le tableau de régime du défini comment ses ASynS sont réalisés. Éventuellement, on peut vérifier que les ASynS ainsi trouvés sont conformes à l'étiquette sémantique de l'actant ; toutefois, cette vérification n'est pas nécessaire, étant donné que les ASynS trouvés à l'aide du tableau de régime indiquent sans ambiguïté les ASém. Dans no-

tre exemple (*la gare d'Ottawa*), il est clair qu'*Ottawa* occupe la position du deuxième ASynS de *gare* : on peut en conclure qu'elle correspond aussi au deuxième ASém. Cette conclusion est confirmée par l'observation que parmi les hyperonymes de OTTAWA, on trouvera VILLE, MUNICIPALITÉ et LIEU. Ce dernier correspond à l'étiquette sémantique de l'actant Y de GARE.

Remarque : les noms propres sont aussi des lexies d'une langue et ils doivent avoir leur entrée de dictionnaire. Parmi les informations sémantiques encodées à leur sujet, on trouvera notamment leur étiquette sémantique (lieu, individu, etc.) et leurs hyperonymes.

IIC : L'ancre occupe une position d'actant sémantique du défini sans en être l'actant syntaxique profond. Cela équivaut à dire que, contrairement à IIB, l'ancre ne fait pas partie du groupe nominal du défini.

Exemple : *la SNCF... a fait un beau cadeau à ses clients* [3.13]. La FP de CLIENT dans le DiCo est :

[X EST ~] de Y

la SNCF occupe ici la position de l'actant sémantique Y.

Stratégie de recherche : trouver un actant sémantique du défini à l'extérieur de son groupe nominal. Si le déterminant du défini est un adjectif possessif, comme c'est le cas dans notre exemple, il faut vérifier si son antécédent peut être un actant sémantique du défini ; c'est-à-dire vérifier si un des hyperonymes de l'antécédent de l'adjectif possessif correspond à l'étiquette sémantique (ou au nom typique) d'un actant du défini – c'est effectivement le cas ici. Si le défini n'est pas déterminé par un adjectif possessif, il faut chercher dans le texte les lexies dont un hyperonyme

pourrait correspondre à l'étiquette sémantique (ou au nom typique) d'un actant du défini. Par exemple, nous avons dans le dixième texte : *Lyon... le conseil municipal* [10.5]. *Lyon*, qui n'est pas un actant syntaxique *du conseil municipal*, peut en être un actant sémantique puisqu'un de ses hyperonymes (LIEU) correspond à l'étiquette sémantique de l'actant de CONSEIL MUNICIPAL.

Remarque : il est important de bien souligner la différence entre l'étiquette sémantique d'un actant et son nom typique. La lexie VENDRE en est une bonne illustration. Sa forme propositionnelle (minimale) est : *X vend Y à Z pour W*. Nos CM peuvent nous indiquer que, typiquement, Y dénote une marchandise et W dénote une somme d'argent. Toutefois, d'un point de vue linguistique, ces deux connaissances sont de nature différente et sont encodées différemment. *Somme d'argent* est une étiquette sémantique de W : pour qu'une situation linguistique puisse être désignée par le verbe VENDRE, il est obligatoire que ce que son actant W désigne soit une somme d'argent (indépendamment de cette situation). Par contre, il n'y a pas de contrainte sur l'actant Y : tout peut devenir une marchandise dans la situation linguistique de VENDRE et ne pas l'être en dehors de cette situation. Par exemple, on peut vendre des pommes : dans la situation de vente, elles sont des marchandises mais elles ne le sont pas en dehors de cette situation. Par contre, une somme d'argent, par exemple 100 dollars, est une somme d'argent indépendamment de toute situation. MARCHANDISE est tout simplement un nom typique de l'actant Y de VENDRE.

III : Défini et ancre liés par une FL autre que Syn, Gener et S_i

Le défini et l'ancre sont liés par une FL autre que synonymique (Ib), générique (Ic) ou actancielle (IIa). Cette relation se trouve également dans le dictionnaire de façon explicite.

Exemple : *gare de Montpellier... les voies* [1.7]. VOIE, le nominal principal du défini, est liée par une FL non standard à GARE (qui est le nominal principal de l'ancre).

Stratégie de recherche : chercher dans le texte toutes les lexies qui sont des valeurs pour les fonctions lexicales du défini ; ensuite, établir une relation entre le défini et ces lexies en fonction du contenu de la FL.

IV : Défini dans certaines expressions idiomatiques

Le défini fait partie d'une expression idiomatique (ou semi-idiomatique) qui contrôle le choix de l'article : l'indéfini serait impossible dans cette expression.

Exemple : *à la suite de* [2.7]. Cette expression est une locution à laquelle un article de dictionnaire est consacré. Cette locution est figée, seul le défini y est possible.

Stratégie de recherche : vérifier dans le dictionnaire si le défini correspond à une locution qui contrôle le choix de l'article ; le défini s'interprète suivant la définition de cette locution.

V : Défini spécifié

Un défini peut être unique dans son contexte en vertu d'une partie de son groupe nominal qui le spécifie. Deux sous-types peuvent être distingués : le défini avec nom propre coréférentiel et le défini avec spécification directe.

Va : Un défini D fait partie du type du défini avec nom propre coréférentiel si un nom propre ayant le même référent que D est apposé à D (ou, dans des cas plus rares, rattaché à D à l'aide d'une préposition) sans qu'il y ait un lien actanciel entre les deux. Ainsi, ce type se distingue du type IIb (par exemple, *le président Bush*) où un nom propre en apposition peut occuper une position actancielle de la définition de l'expression définie.

Exemple : *le fleuve St-Laurent* [8.7]. Le nom propre *St-Laurent* n'occupe pas une position actancielle du défini. Les expressions *le fleuve* et *St-Laurent* pourraient être utilisées indépendamment l'une de l'autre tout en désignant le même référent.

Stratégie de recherche : vérifier si le défini contient un nom propre ; si le nom propre est introduit par une préposition (*la ville de New York* [12.12]), vérifier s'ils peuvent être coréférentiels. Pour cette dernière vérification, il faut comparer l'hyperonyme du nom propre et le défini.

Vb : Le défini est spécifié directement par ce qui est à l'intérieur de son groupe nominal, sans avoir des liens particuliers, comme c'est le cas dans les types I, II et III.

Exemple : *les salariés chargés de l'entretien des jardins publics* [12.15]. De quels salariés est-il question ? La réponse est dans l'expression même : il est question des

salariés chargés de l'entretien des jardins publics²⁷. L'information nécessaire à l'identification du référent du défini est à l'intérieur du groupe nominal du défini.

Stratégie de recherche : identifier les limites du groupe nominal du défini, analyser les relations syntaxiques et sémantiques entre les différentes parties de ce groupe nominal.

VI : Défini déictique

Le défini déictique se comprend par rapport à la situation de l'énonciation²⁸ dont les principaux paramètres sont, rappelons-le, l'énonciateur, le locuteur, le lieu et le moment de l'énonciation.

Exemple : le référent de l'occurrence *le premier janvier* [3.6] se comprend par rapport au moment de l'énonciation, sauf spécification contraire explicite : il s'agit du dernier 1^{er} janvier si la phrase est au passé, du prochain 1^{er} janvier si la phrase est au futur et du 1^{er} janvier le plus proche si la phrase est au présent.

Stratégie de recherche : vérifier si le défini peut être lié à un des paramètres de l'énonciation. Typiquement, on peut s'attendre à ce que les définis qui ont un contenu temporel (par exemple, les événements, les dates) puissent être liés au moment de

²⁷ Certes, on pourrait encore rajouter des informations et dire par exemple qu'il s'agit des salariés chargés de l'entretien des jardins publics new-yorkais. Toutefois, l'adjectif *new-yorkais* qualifierait non pas *les salariés* mais *les jardins publics* et, par conséquent, serait nécessaire non pas pour la question *De quels salariés s'agit-il ?* mais pour celle *De quels jardins publics s'agit-il ?*

²⁸ Par conséquent, ce type de défini fait clairement appel à des connaissances non linguistiques.

l'énonciation et tout ceux qui ont un contenu spatial (les objets, les lieux, etc.) peuvent être liés au lieu de l'énonciation.

VII : Défini considéré dans sa généralité

Les définis appartenant à cette catégorie sont considérés dans leur généralité²⁹. On peut distinguer deux sous-catégories :

VIIa : *le(s) X* doit être considéré comme équivalent à *(presque) tous les X*. Ce type de défini a tendance à être au pluriel.

Exemple : *les Canadiens* [4.2]. Il s'agit des Canadiens en général, de presque tous les Canadiens.

Stratégie de recherche : interpréter le défini comme *tous les X* et essayer de trouver des exceptions.

VIIb : *le X* est une entité abstraite que l'on considère dans sa totalité, dans son ensemble, en général. (Il reste à étudier si certaines catégories sémantiques des noms sont plus facilement interprétables de façon générale que d'autres : *l'éducation* versus *la perturbation*.)

Exemple : *l'éducation* [7.17]. Il s'agit de l'éducation en général.

²⁹ A ne pas confondre avec le sous-type Ic – le défini hyperonymique.

Stratégie de recherche : interpréter le défini de la façon la plus générale et vérifier que le cotexte ne contient pas de restriction.

VIII : Défini identifié de manière complexe

C'est notre dernier type d'identification du défini. Le plus souvent, on y trouve une spécification contextuelle du défini dont le traitement est plus complexe que celui des types précédents. Autrement dit, l'information nécessaire à l'interprétation du défini se trouve dans le texte mais il n'y a pas d'ancre unique, le défini ne peut pas être mis en relation avec une lexie précise :

Exemple : dans [5.11], on a *l'affaire*. En lisant le texte, le lecteur interprète cette expression comme faisant référence à l'histoire d'Antoine, expulsé de l'école à cause d'une loi linguistique belge.

Stratégie de recherche : utiliser la représentation sémantique du texte entier et lier le défini à d'autres parties du cotexte.

Bien entendu, une expression définie peut faire partie de plusieurs types si elle a plusieurs ancres. Par exemple, dans le quatrième texte, nous avons *le premier ministre ... son chef de cabinet, Jean Pelletier* [4.6]. Le défini *son chef de cabinet* fait partie à la fois du type IIa (PREMIER MINISTRE est un nom typique du deuxième actant de CHEF DE CABINET) et du type IIb (*Jean Pelletier* occupe la première position actancielle de *son chef de cabinet*).

Ceci termine la présentation de notre méthodologie et nous pouvons maintenant passer à présenter les analyses de défini.

5.3.4 Connaissances lexicales nécessaires aux analyses

Les connaissances lexicales nécessaires aux analyses effectuées ici peuvent être divisées en deux catégories : celles qui sont déjà dans le DiCo et celles qui n'y sont pas encore.

5.3.4.1 Connaissances présentes dans le DiCo

Le DiCo que nous utilisons n'est pas encore un dictionnaire complet. Cela implique surtout que d'autres informations y seront ajoutées et non pas tant que les informations qui y sont déjà risquent d'être supprimées ou modifiées – même si les modifications sont également possibles. Rappelons également que les auteurs du DiCo distinguent quatre statuts dans l'élaboration d'un article de dictionnaire : 3, 2, 1 et 0. Les articles de statut 3 sont les moins élaborés (et par conséquent les plus susceptibles de subir des modifications et des ajouts) alors que les articles de statut 0 sont considérés comme terminés et risquent relativement peu d'être modifiés. Certes, même ces articles peuvent être incomplets mais l'information qu'ils contiennent peut être considérée comme cohérente et validée. Dans nos analyses, nous indiquons le statut 0, 1, 2 ou 3 de l'article utilisé en spécifiant que l'article est du Dico0, du Di-

Co1, du DiCo2 ou du DiCo3. Ce formalisme est propre à notre travail et nous l'utilisons dans le seul but d'indiquer l'information pertinente de manière concise.

Tout en étant conscient du fait que la désambiguïsation des lexies n'est pas une question résolue, nous nous limitons ici à utiliser les lexies (du DiCo) qui correspondent le mieux aux lexies que nous trouvons dans les analyses. Dans le cas des vocables homonymiques, un chiffre arabe entre parenthèses marque le numéro de vocable. Les lexies polysémiques sont distinguées les unes des autres à l'aide des chiffres romains (marquant une distance sémantique majeure), des chiffres arabes et des lettres (marquant une distance sémantique plus réduite).

5.3.4.2 Élaboration des CL utilisée dans l'analyse

Comme nous l'avons mentionné plus haut, le DiCo n'est pas encore un dictionnaire complet et cela entraîne quelques inconvénients par rapport à nos analyses. En effet, les interlocuteurs (les auteurs des textes analysés et leurs lecteurs) ont à leur disposition une grande quantité de CL, notamment pour l'interprétation des expressions définies, mais une partie de ces connaissances est, pour l'instant, absente du DiCo dont nous nous servons. Pour bien faire les analyses, nous nous trouvons dans l'obligation de « compléter » le DiCo, c'est-à-dire élaborer une petite partie des articles manquants du DiCo qui semble être utile lors de l'interprétation du défini. Nous présentons les analyses avec les CL que nous avons rajoutées et justifions ces modélisations lexicales additionnelles en nous servant de la méthodologie descriptive de

la LEC. Il s'agit de montrer que ces CL rajoutées ne sont pas des extensions *ad hoc* du DiCo.

Bien entendu, nous ne pouvons pas élaborer de façon formelle les articles complets qui manquent. Cela demanderait un travail énorme qui ne serait pas vraiment utile pour les analyses. Un tel travail inclurait en premier lieu la délimitation méticuleuse de la lexie par rapport aux autres lexies du même vocable, notamment en étudiant un corpus contenant un nombre important d'occurrences de ce vocable. Par exemple, trouve-t-on la même lexie ADIEU dans *Pierre a dit adieu à Marie* et dans *Pierre a dit adieu à la carrière d'enseignant* ? S'agit-il de la même lexie CIRCULATION dans *la circulation des trains* et dans *la circulation des voitures* ? Ensuite, il faudrait étudier toutes les positions actanciennes contrôlées par la lexie (et non seulement celles qui sont intéressantes pour nos analyses). Finalement, l'élaboration de la FP d'une lexie demanderait aussi de trouver les étiquettes sémantiques adéquates (par exemple, choisir entre 'individu' et 'personne' ou encore entre 'lieu', 'espace' et 'territoire').

Nous cherchons ici à démontrer surtout les liens actanciels. Une telle démonstration est basée sur l'examen de deux caractéristiques de la lexie : d'un côté, son sens fait appel à un ou plusieurs participant(s) de la situation linguistique que la lexie dénote, d'autre par ce(s) participant(s) est (sont) exprimable(s) auprès de la lexie. Rappelons que ces deux caractéristiques ont déjà été présentées dans (3.2.4.2).

Voici la façon dont nous procédons. L'examen des participants de la situation peut consister à établir le sens (au moins approximatif) de la lexie en nous inspirant d'un bon dictionnaire, le plus souvent le *Nouveau Petit Robert* (NPR), parfois du dictionnaire d'Antidote. Par exemple, selon le NPR, la définition de HABITANT est 'Personne qui réside habituellement en un lieu déterminé'. La situation décrite par cette définition indique deux participants : quelqu'un [X] est habitant d'un lieu [Y]. L'étude des collocations où se trouve la lexie peut également être révélatrice des actants. Par exemple, la définition de ÉPREUVE dans le NPR ('Souffrance, malheur, danger qui éprouve le courage, la résistance') ne permet pas facilement d'en établir les positions actanciennes mais les collocations que l'on trouve dans cet article (*essuyer, subir des épreuves; passer par de dures, de rudes épreuves*) le permettent. En effet, la situation linguistique évoquée par ces verbes (ESSUYER, SUBIR, PASSER [par]) nécessite deux participants : un premier participant qui essuie, qui subit quelque chose ou qui passe par quelque chose et un second participant qui est essuyé, subi par quelqu'un, par où quelqu'un passe. On ne pourrait pas utiliser ces verbes sans que la situation décrite ait ces deux participants. Nous utilisons également Beauchesne (2001) pour trouver des collocations. Ce dernier est un dictionnaire qui, pour chaque nom qu'il contient, donne une série de cooccurrents fréquents, dont la plupart sont des collocatifs au sein de collocations contrôlées par le mot-clé.

L'examen de la seconde caractéristique consiste à chercher les exemples où le participant de la situation qui nous intéresse est exprimé auprès de la lexie en question.

De façon complémentaire, nous vérifions si le DiCo contient des (quasi-) synonymes de la lexie. C'est également un bon point de départ pour établir des liens lexicaux parce que les quasi-synonymes ont souvent la même FP (ou une FP proche). Bien entendu, il peut y avoir aussi des différences, c'est notamment le cas entre VOYAGEUR et PASSAGER. Bien qu'il s'agisse des quasi-synonymes, on parlerait plus facilement des *passagers d'un train* que des *voyageurs d'un train*. En effet, un passager est forcément un passager d'un véhicule : la situation décrite par PASSAGER a donc un participant obligatoire qui est ce véhicule. Par contre, un voyageur peut soit emprunter un véhicule soit voyager à pied : le véhicule n'est pas un participant obligatoire de VOYAGEUR – il semble être plutôt un participant optionnel.

Nous pouvons récapituler la procédure utilisée pour établir des liens actanciels de la façon suivante : si une situation linguistique décrite par la lexie L a un participant X et ce X est exprimable auprès de L, alors L contrôle une position actancielle correspondant à ce X.

En plus des relations actanciels, nous devons parfois démontrer d'autres relations lexicales, telle la synonymie. C'est une relation assez intuitive que nous tentons d'appuyer par des dictionnaires.

Nous nous permettons cependant d'utiliser sans justification des CL triviales : par exemple, VILLE est un hyperonyme des noms de ville, alors que INDIVIDU est un

hyperonyme des noms propres personnels. Nous nous dispensons également de justification dans le cas des lexies qui sont liées à la fois sémantiquement et morphologiquement de façon évidente : PROTESTATION et PROTESTATAIRE, GRÈVE et GRÉVISTE.

5.3.4.3 Vérification d'une partie des connaissances utilisées dans les analyses

Dans une première étape de notre travail, nous avons utilisé seulement les articles du DiCo qui étaient au statut zéro pour établir les liens lexicaux entre le défini et l'ancre. Dans le cas des lexies qui ne se trouvaient pas dans cette version du DiCo, nous avons échafaudé la FP. Nous sommes en mesure maintenant de vérifier une bonne partie de ces affirmations à l'aide d'une version plus récente du DiCo qui couvre un nombre de lexies plus élevé : en effet, cette deuxième version incluait aussi des lexies au statut 1, 2 et 3. Cette vérification concernant le lien actanciel entre les lexies a démontré deux choses : d'une part, nos FP étaient souvent approximatives, d'autre part, les liens actanciels qui nous intéressaient étaient confirmés dans la totalité des cas. Il est intéressant de s'attarder un peu sur chacun de ces deux résultats.

En premier lieu, voici quelques exemples des différences entre nos postulats et la FP du DiCo. Nous avons identifié dans [7.5] (*l'explosion de la guerre civile espagnole*) la FP suivante pour EXPLOSION :

explosion de l'événement X.

Voici la FP proposée par le DiCo :

~ DU fait X

Dans ce cas, la différence réside seulement dans l'étiquette de l'actant X : 'événement' versus 'fait'.

Nous avons proposé dans [8.18] (*afin d'assurer la préservation du littoral et d'en faire profiter la population*) la FP suivante pour POPULATION :

population du lieu X.

Voici la FP proposée par le DiCo :

[ensemble-2->individu X] ~ DE territoire Y

On peut constater ici trois différences : la FP de DiCo contient un actant de plus qui dénote un ensemble d'individus, l'étiquette sémantique de l'actant qui nous intéresse est 'territoire' et non pas 'lieu' et cet actant est identifié par la lettre Y plutôt que par la lettre X.

Nous avons proposé dans [10.4] (*Raymond Barre, député-maire depuis 1995, préside ce lundi sa dernière séance du conseil municipal*) la FP suivante pour SÉANCE :

séance de l'assemblée X (pour faire le travail Y).

Voici la FP proposée par le DiCo :

~ DES individus X À PROPOS DE Y

L'étiquette de l'actant X dans le DiCo ('individu') est plus générale que ce que nous avons postulé ('assemblée') et il en est ainsi pour l'actant Y aussi ('à propos de' dans le DiCo est 'pour faire le travail' dans notre postulat).

Toutefois, le plus important est le deuxième résultat, à savoir que le lien actanciel que nous avons postulé pour nos analyses était confirmé par tous les articles que la deuxième version du DiCo contenait en plus. Dans les exemples donnés plus haut, l'important par rapport à nos analyses est que *la guerre* occupe une position actancielle de EXPLOSION dans [7.5], *les rives* occupe une position actancielle de POPULATION dans [8.18] et *conseil municipal* occupe une position actancielle de SCÉANCE dans [10.4].

Ces résultats cohérents avec les descriptions du DiCo viennent de l'application des principes et méthodes de la LEC. La raison en est qu'à partir du moment où on applique ces mêmes principes et méthodes, on tend à parvenir à des descriptions identiques ou, tout du moins, compatibles.

5.4 Analyse de l'interprétation du défini dans les textes

Le corpus analysé est composé de textes trouvés sur Internet. Chacun d'entre eux traite d'un des quatre lieux publics suivants : la gare, l'école, la mairie et le jardin public. Dans certains textes, le lieu public joue un rôle central, dans d'autres, un rôle secondaire. Ce sont des faits divers qui étaient faciles à trouver au moment de la constitution de notre corpus, début 2001. N'étant plus d'actualité, pour la plupart, ils ne sont malheureusement plus disponibles sur Internet maintenant. Ils sont intéres-

sants pour nous parce qu'ils s'adressent à un public général dans un français contemporain. Nous avons jugé intéressant d'inclure dans nos analyses des textes de longueur variable, notamment pour avoir une vision plus juste de l'importance des relations anaphoriques qui a tendance à s'accroître avec la longueur du texte.

Chaque analyse contient l'expression définie (et son ancre, s'il y en a), les informations utiles à son identification et, comme conclusion, le type d'identification. Les stratégies de recherche pour l'identification du défini données plus haut, que nous ne reprenons pas ici, peuvent être appliquées en fonction du contenu des analyses.

Ces analyses nécessitent de nombreuses informations lexicales. Celles qui sont déjà encodées dans le DiCo sont présentées comme telles, en utilisant la police `courrier` et suivant le formalisme du DiCo.

5.4.1 Premier lieu : la gare

Nous analysons ici quatre textes contenant 53 expressions définies.

Premier texte

6 février 2001

Des infirmiers grévistes bloquent les trains (1) en gare de Montpellier

MONTPELLIER (AP) -- Une cinquantaine d'infirmiers anesthésistes grévistes bloquent mardi depuis 11h30 la circulation (2) des trains (3) en gare de Montpellier (Hérault), a-t-on appris auprès de la direction (4) régionale de la SNCF.

Les protestataires (5) venus de Montpellier, Alès, Nîmes (Gard) et Perpignan (Pyrénées-Orientales), se sont assis au milieu (6) des voies (7), entraînant l'arrêt (8) de tous les trains (9) dès leur arrivée (10) en gare.

Un commissaire de police s'est rendu sur place à midi pour demander l'évacuation (11) des voies (12), mais a essuyé un refus.

[1.1]

A1) les trains en **gare** de Montpellier

B) GARE est le lieu typique où se trouvent les trains : elle est une valeur de la FL correspondante, ce que l'article TRAIN du DiCo3 encode de la façon suivante :

{S-loc} gare; station, arrêt

Autrement dit, GARE et TRAIN sont liés par une FL.

C) Type III.

[1.2]

A1) *la circulation des trains en gare de Montpellier (Hérault)*

B) La FP de CIRCULATION_{I.4} dans le DiCo3 est :

~ DES véhicules X ENTRE L'endroit Y ET L'endroit Z

On trouvera – entre autre – TRAIN parmi les noms typiques de l'actant sémantique X de CIRCULATION car son étiquette sémantique ('véhicule') est un hyperonyme de TRAIN. En effet, l'article DiCo3 de VÉHICULE contient la FL suivante :

{Types de véhicules}

Une partie de la valeur de cette FL est :

sur rail : autorail, draisine, locomotive, locomotrice, locotracteur, train, tramway, wagon

C) Type IIa.

[1.3]

A1) *les trains en gare de Montpellier*

B) Comme dans [1.1], GARE et TRAIN sont liés par une FL :

{S-loc} gare; station, arrêt

C) Type III.

[1.4]

A1) *la direction régionale de la SNCF*

B) Le NPR définit DIRECTION comme 'La personne ou l'équipe qui dirige une entreprise' et donne les exemples suivants : *S'adresser à la direction du journal. La direction d'un syndicat.* Cette définition indique que la situation évoquée par DIRECTION a deux participants : quelqu'un dirige quelque chose. Nous nous intéressons ici au deuxième participant, il est exprimable auprès de la lexie, comme en témoignent les exemples. Cela montre que la FP de DIRECTION contient une position actancielle : l'entité dirigée. *la SNCF*, qui est un ASynS de *la direction* occupe dans notre exemple cette position actancielle.

C) Type IIb.

[1.5]

A1) *les protestataires* venus de Montpellier, Alès, Nîmes (Gard) et Perpignan (Py-rénées-Orientales)

A2) *une cinquantaine d'infirmiers anesthésistes grévistes*

B) Le lecteur peut établir une relation coréférentielle entre les protestataires et les infirmiers grévistes, notamment à travers les lexies PROTESTATAIRE, PROTESTATION, GRÈVE et GRÉVISTE. Voici la chaîne des relations :
 $S_1(\text{PROTESTATION}) = \text{PROTESTATAIRE}$,

$S_{\text{med}}(\text{GRÈVE}) = \text{PROTESTATION}$ (selon le NPR, GRÈVE se définit comme ‘arrêt volontaire et collectif d'une activité, par revendication ou **protestation**’),

$A_1(\text{GRÈVE}) = \text{GRÉVISTE}$ (GRÉVISTE est un adjectif ici, pas un nom).

Autrement dit, les informations nécessaires à l'identification du référent du défini sont dans le cotexte ; leur traitement demande la prise en considération d'une série de trois liens lexicaux.

C) Type VIII.

[1.6]

A1) *le milieu des voies*

B) La FP de MILIEU(1)1 dans le DiCo3 est :

~ DE L' espace X

les voies, qui est un ASynS *du milieu*, occupe ici la position de X.

C) Type IIb.

[1.7]

A1) *les voies*

A2) *gare de Montpellier*

B) VOIE est une FL non standard de GARE, encodée dans le DiCo0 sous l'article de GARE de la façon suivante :

{Endroit dans une G. où roulent les trains} voie
Le défini est lié à l'ancre par une FL.

C) Type III.

[1.8]

A1) *l'arrêt de tous les trains*

B) La FP de ARRÊT(1)2 dans le DiCo est :

~ DE objet X DANS endroit Y

les trains, qui est un ASynS de *l'arrêt*, occupe ici la position actancielle de X.

En plus, S'ARRÊTER est une valeur de FL de TRAIN, encodée dans l'article TRAIN du DiCo de la façon suivante :

{FinFact0} s'arrêter

Or, S'ARRÊTER est une valeur de FL du nom ARRÊT(1)2 dont l'article DiCo3 contient le suivant :

{Oper12} faire [ART ~ Loc-in N=Y] //s'arrêter [Loc-in N=Y]

C) Type IIb.

[1.9]

A1) *tous les trains*

A2) *les trains* [1.4]

B) Il s'agit ici des mêmes trains que dans [1.4], le défini est anaphorique, il contient le même nominal que son antécédent.

C) Type Ia.

[1.10]

A1) *leur arrivée en gare*

A2) *tous les trains*

B) La FP de ARRIVÉE dans le DiCo3 est :

~ DE L'individu OU DU véhicule X DANS endroit Y

gare, qui est un ASynS de *leur arrivée*, occupe ici la place de Y.

tous les trains, qui est l'antécédent du déterminant du défini, occupe la place de X.

Remarquons qu'en plus de cette relation actancielle entre TRAIN et ARRIVÉE, dans l'article TRAIN, dont la FP dans le DiCo3 est :

~ [CONDUIT PAR individus X POUR individus/chargement Y] DE ville Z à ville W,

on trouve la FL suivante :

{FinFact0--actuel à Z} arriver < entrer en gare [à N=Z]; s'arrêter, s'immobiliser [à N=Z]

Or, $V_1(\text{ARRIVÉE}) = \text{ARRIVER}$.

C) Type IIb (pour Y) et IIc (pour X).

[1.11]

A1) *l'évacuation des voies*

B) Le NPR définit ÉVACUATION comme 'Action d'évacuer (un lieu)' et donne les exemples suivants : *Évacuation d'une place forte, d'un territoire, d'un pays par les troupes d'occupation*. Le sens 'un lieu' et les syntagmes *une place forte, un territoire*, etc. indique un participant de la situation qui est exprimable auprès de ÉVACUATION : un lieu qui est évacué. Qui plus est, ABANDON^{1.2a} (un quasi-synonyme de ÉVACUATION) contient également une position actancielle semblable (marquée par Y) dans sa FP, selon le DiCo0 :

~ PAR LA personne X DU lieu Y

La FP de ÉVACUATION contient donc une position actancielle : le lieu à évacuer. *les voies*, qui est un ASynS de *l'évacuation* occupe cette position.

C) Type IIb.

[1.12]

A1) *les voies*

A2) *les voies* [1.9]

B) *les voies* a le même référent ici que *les voies* [1.9]

C) Type Ia.

Deuxième texte

Les passagers (1) d'un TGV évacués après une alerte à la bombe (2)

10 février 2001

GRENOBLE (Reuters) - Les 650 passagers (3) du TGV (4) Paris-Bourg-Saint-Maurice ont du être évacués dans la matinée (5) en gare de Bourgoin-Jallieu (Isère) à la suite (6) d'une alerte à la bombe (7).

Les voyageurs (8) ont été transbordés dans un autre train afin qu'ils puissent rejoindre Bourg-Saint-Maurice (Savoie).

Le TGV (9) est examiné par les services (10) de déminage.

[2.1]

A1) *les passagers d'un TGV*

B) Antidote définit PASSAGER comme 'usager d'un moyen de transport' et donne les exemples comme *le passager d'un train, d'un navire, d'un avion*, etc. La définition indique que le moyen de transport est un participant de la situation évoquée par PASSAGER et les exemples montrent que ce participant est exprimable auprès de PASSAGER. Tout cela indique que la FP de PASSAGER contient une position actancielle : un moyen de transport.

un TGV, qui est un ASynS de l'expression *les passagers* occupe ici cette position actancielle.

Remarquons aussi qu'il existe un lien lexical indirect entre PASSAGER et TGV, à travers TRAIN. En effet, TRAIN, qui est clairement l'hyperonyme de TGV, contient dans son article du DiCo la ligne suivante :

{S1Real2 = S2Real1} passager [de ART ~]//voyageur

C) Type IIb.

[2.2]

A1) *une alerte à la bombe*

B) Nous avons ici une locution : ALERTE À LA BOMBE. L'article y est obligatoirement au défini.

C) Type IV.

[2.3]

A1) *les 650 passagers du TGV Paris-Bourg-Saint-Maurice*

B) Comme dans [2.1], *TGV* occupe une position actancielle de PASSAGER. En outre, il existe également un lien lexical indirect entre TGV et PASSAGER.

C) Type IIb.

[2.4]

A1) *le TGV Paris-Bourg-Saint-Maurice*

B) Étant donné que TGV est un acronyme de *Train à Grande Vitesse*, il aura la même FP que TRAIN dont il est un hyponyme. Cette FP a déjà été présentée dans [1.10]. L'ancre remplit ici la position actancielle de Z et de W.

En outre, l'article TRAIN contient la FL suivante :

{Si} ville, localité

Or, l'hyperonyme de l'ancre ici est VILLE et LOCALITÉ.

C) Type IIb.

[2.5]

A1) *la matinée*

B) Il n'y a pas d'ancre dans le texte. Le défini doit être interprété ici par rapport au moment de l'énonciation, comme un déictique.

C) Type VI.

[2.6]

A1) à la suite de

B) Nous avons ici une locution : À LA SUITE DE. Elle contrôle le choix de l'article.

C) Type IV.

[2.7]

A1) **alerte** à la bombe

B) Nous avons ici la même locution que dans [2.2] : ALERTE À LA BOMBE. Elle contrôle le choix de l'article.

C) Type IV.

[2.8]

A1) les voyageurs

A2) les 650 passagers du TGV [2.3]

B) Selon le NPR, VOYAGEUR et PASSAGER sont des quasi-synonymes.

C) Type Ib.

[2.9]

A1) le TGV

A2) le TGV Paris-Bourg-Saint-Maurice [2.4]

B) Le défini est anaphorique, sont antécédent étant [2.4].

C) Type Ia.

[2.10]

A1) les services de déminage

B) La FP de SERVICEII dans le DiCo3 est :

~ DE L'activité Y

déminage occupe ici la position actancielle de Y.

C) Type IIb.

Troisième texte

C'est reparti !

Par Alain-Marie Carron — *Canoë.fr*

Le démarrage (1) a été pénible — gueule de bois oblige — mais, en une semaine, les Français (2) ont retrouvé « la pêche » (3), c'est-à-dire l'énergie (4) et de bonnes habitudes toniques, comme on va le voir.

Pourtant, au matin (5) du 1er janvier (6), la situation (7) semblait glauque. Gare de Lyon, 7h. Tous les cafés, boulangeries et commerces (8) divers ont baissé leurs stores (9) de fer : la gare (10) est occupée par environ deux mille jeunes qui attendent les premiers trains (11) pour rentrer chez eux, c'est-à-dire, en l'occurrence (chacune des gares (12) parisiennes dessert une direction particulière) le Sud: Toulon, Béziers, Marseille.

La SNCF, compagnie nationale de chemins de fer, a fait un beau cadeau à ses clients (13) : tous les billets (14) sont à 100F aller-retour pour se rendre à Paris le soir (15) du 31 Décembre (16), d'où que l'on vienne, à condition de repartir au plus tard le 1er Janvier (17) avant-midi...

[3.1]

A1) *le démarrage*

A2) *C'est reparti !*

B) Il est à signaler qu'un texte qui commence par un défini sans modificateur a quelque chose de surprenant. Toutefois, le titre de ce texte contient une locution : *C'est reparti !* Or, l'article REPARTIR (quasi-synonyme de REDÉMARRER) mentionne l'expression *Allez, c'est reparti !* Cela suggère un lien lexical : la lexie

DÉMARRAGE aura une fonction lexicale non standard qui exprime ce que l'on dit (en occurrence, *C'est reparti !*) quand on constate le démarrage de quelque chose. Ainsi, le lecteur peut établir un lien lexical direct entre le titre et ce défini.

C) Type III.

[3.2]

A1) *les Français*

B) En l'absence d'autre information, le lecteur comprend qu'il s'agit des Français en général, de presque tous les Français.

C) Type VIIa.

[3.3]

A1) *la pêche*

B) Le NPR mentionne l'expression *avoir la pêche*. L'expression *retrouver la pêche* est équivalent à *(re-)commencer à avoir la pêche*. Le défini se trouve donc à l'intérieur d'une collocation (*retrouver la pêche*). L'article défini est obligatoire dans cette expression. (Toutefois, si l'expression avait un modificateur, on aurait plutôt l'article indéfini : *retrouver une pêche d'enfer*.)

C) Type IV.

[3.4]

A1) *l'énergie*

A2) *la pêche*

B) Le NPR définit l'expression *avoir la pêche* par 'avoir le moral, être en forme' alors que ÉNERGIE y est définit comme 'Force et fermeté dans l'action, qui rend capable de grands effets'. Ces définitions suggèrent que PÊCHE et ÉNERGIE sont des quasi synonymes. (Éventuellement, on pourrait aussi soutenir que PÊCHE est un hyponyme de ÉNERGIE.)

C) Type Ib.

[3.5]

A1) *le matin du 1^{er} janvier*B) La FP de MATIN_b dans le DiCo0 est :

~ DU jour X

Or, dans le NPR, on trouve une lexie QUANTIÈME qui se définit comme ‘Le jour du mois, désigné par un chiffre (de premier, deux... à trente ou trente et un)’. Il est synonyme de la lexie DATE, cette dernière étant définie comme ‘Indication du jour du mois’. Cela suggère qu’il existe une lexie LE N DU MOIS [M] (dont l’expression *le 1^{er} janvier* est une instantiation) et JOUR est son hyperonyme. *le 1^{er} janvier*, qui est un ASynS de l’expression *le matin* occupe donc la position actancielle de X dans la FP de MATIN_b.

C) Type Iib.

[3.6]

A1) *le 1^{er} janvier*

B) Cette date doit être interprétée comme le dernier 1^{er} janvier qui précède le moment de l’énonciation, l’expression est déictique.

C) Catégorie VI.

[3.7]

A1) *la situation*

B) Le NPR définit SITUATION comme ‘Ensemble des circonstances dans lesquelles une personne se trouve’. Cette personne peut être exprimée aussi bien par un adjectif possessif (*ma/ta/sa situation*) que par un verbe support (*être dans une situation*).

En plus, ÉTAT(1)**1a** et SITUATION sont des quasi-synonymes selon le DiCo3 et la FP du premier y est :

~ DE X,

Tout cela suggère que la définition de SITUATION contient une position actancielle : la personne dont c’est la situation. Bien qu’il ne soit pas spécifié de façon

explicite ce qui occupe cette position actancielle dans notre exemple, le lecteur peut comprendre en lisant le texte qu'il s'agit de la situation des Français après le 31 décembre en général (et en particulier à la Gare de Lyon).

C) Type VIII.

[3.8]

A1) *tous les cafés, croissanteries et commerces divers*

A2) ***Gare de Lyon***

B) Le lecteur comprend qu'il s'agit des cafés, commerces et croissanteries de la Gare de Lyon même si cela n'est pas spécifié de façon explicite et même s'il n'y pas de lien lexical entre l'acre et le défini.

C) Type VIII.

[3.9]

A1) *leurs stores de fer*

A2) *les **cafés, croissanteries et commerces divers*** [3.8]

B) Le NPR défini STORE comme 'Grand rideau, à la devanture d'un magasin'. Ce magasin peut être facilement exprimé : *le store d'un magasin, son store*.

En plus, selon le NPR, STORE et RIDEAU (comme dans *Les commerces avaient baissé leur rideau.*) sont synonymes. Or, dans le DiCo3, la FP de RIDEAU.1c est la suivante :

~ DU commerce X

Tout cela indique que la FP de STORE contient une position actancielle : un commerce. L'ancre (*les cafés, croissanteries et commerces divers*) est l'antécédent du déterminant (*leurs*) du défini : elle occupe cette position actancielle.

C) Type IIc.

[3.10]

A1) *la gare*

A2) ***Gare de Lyon***

B) GARE DE LYON est un nom propre dont GARE est un hyperonyme.

C) Type Ic.

[3.11]

A1) *les premiers trains*

A2) *Toulon, Béziers, Marseille*

B) Rappelons que dans la définition de TRAIN, il y a une position actancielle pour le lieu d'arrivée (W). Or, VILLE et LIEU sont des hyperonymes de chacun des noms qui se trouvent dans A2) – ce qui permet de les reconnaître comme des actants W de la définition de TRAIN.

Remarque : TRAIN a aussi une position actancielle : Z qui est le nom du lieu de départ. Grâce à l'expression *pour rentrer chez eux*, le lecteur interprète ces villes comme faisant référence à la destination et non pas au départ.

C) Type IIc.

[3.12]

A1) *les gares parisiennes*

B) La FP de GARE dans le DiCo0 est :

~ DU lieu Y [POUR LE train X]

Son tableau de régime contient l'information suivante :

Y = II = de N, A-poss, Adj | Adj = A0(N)

Cela signifie que la deuxième actant syntaxique de surface de GARE peut se réaliser soit comme un nom introduit par la préposition DE, soit par un adjectif possessif, soit par un adjectif qui est le dérivé adjectival de ce nom. L'ancre (*parisienne*) est un adjectif dérivé d'un nom de lieu : elle occupe la position de Y.

C) Type IIb.

[3.13]

A1) *ses clients*

A2) *la SNCF*

B) La FP de CLIENT dans le DiCo3 est :

[X EST ~] de Y

SNCF, qui est l'antécédent du déterminant dans *ses clients*, occupe la position de Y.

C) Type IIc.

[3.14]

A1) tous les billets aller-retour pour se rendre à **Paris**

B) La FP de BILLET^{I.2} dans le DiCo0 est :

~ DE L'individu X POUR ALLER AU lieu Y [À PARTIR DU lieu Z]

Paris, dont LIEU est un hyperonyme, occupe la position actancielle de Y.

C) Type IIb.

[3.15]

A1) le soir du **31 décembre**

B) L'interprétation de ce défini est semblable à celle du défini dans [3.5] (*le matin du 1^{er} janvier*). La FP de SOIR^I est :

~ DE LA journée X

le 31 décembre, qui est un ASynS du syntagme *le soir*, occupe la position de X.

C) Type IIb.

[3.16]

A1) *le 31 décembre*

B) Le défini est déictique.

C) Type VI.

[3.17]

A1) le 1^{er} janvier

B) Le défini est déictique.

C) Type VI.

Quatrième texte

Chrétien ébranlé

Le décès (1) de Pierre Trudeau jeudi dernier, a remué beaucoup d'émotions chez les Canadiens (2) et Québécois (3).

L'épreuve (4) a été particulièrement dure pour le premier ministre (5) Jean Chrétien qui a de la difficulté à s'en remettre, admettait mardi son chef de cabinet (6), Jean Pelletier.

Le moment (7) le plus difficile pour M. Chrétien a été de faire ses adieux (8) à M. Trudeau à la gare (9) d'Ottawa. Exposée en chapelle ardente au Parlement pendant deux jours, la dépouille (10) de M. Trudeau a quitté Ottawa lundi en direction de Montréal.

Sur le quai (11) de la gare (12), M. Chrétien a embrassé les fils (13) Trudeau, Justin et Sacha, avant de retourner, ébranlé, à son bureau (14).

[4.1]

A1) le décès de **Pierre Trudeau**

B) La FP de DÉCÈS dans le DiCo3 est :

~ DE L'individu X

Pierre Trudeau, qui est un ASynS du syntagme *le décès*, occupe la position de X.

C) Type IIb.

[4.2]

A1) *les Canadiens*

B) En l'absence d'autre information, le lecteur comprend qu'il s'agit des Canadiens en général.

C) Type VIIa.

[4.3]

A1) *les Québécois*

B) Comme dans le cas du défini précédent, il s'agit des Québécois en général.

C) Type VIIa.

[4.4]

A1) *l'épreuve*

A2) *le décès* [4.1]

pour le **premier ministre** [4.5]

B) L'article ÉPREUVE du NPR contient les exemples avec des verbes supports : *Essuyer, subir des épreuves. Passer par de dures, de rudes épreuves.* Ces verbes évoquent des situations avec (au moins) deux participants : quelqu'un essuie quelque chose, quelqu'un subit quelque chose, quelqu'un passe par quelque chose. Ces participants sont facilement exprimables : *Échouer à cet examen serait une épreuve terrible pour Pierre.* Nous en concluons que la FP de ÉPREUVE contient deux positions actanciennes : l'événement X est une épreuve pour Y. La position de X est occupée dans notre exemple par *le décès*, dont ÉVÉNEMENT est un hyperonyme, et la position de Y par *le premier ministre*, dont INDIVIDU est un hyperonyme.

C) Type IIc.

[4.5]

A1) *le premier ministre Jean Chrétien*

B) Le NPR définit PREMIER MINISTRE comme 'chef du gouvernement (en régime parlementaire), nommé par le président de la République'. Cette définition suggère deux participants de la situation : quelqu'un est le premier ministre d'un gouvernement. Bien entendu, ces deux participants sont facilement exprimables :

Édouard Balladur était premier ministre de France de 1993 à 1995. Nous pouvons en conclure que la FP de PREMIER MINISTRE contient deux positions actanciellles : une personne X est premier ministre d'un gouvernement Y

Par ailleurs, suivant le DiCo, la FP de MINISTREI est :

[L'individu X EST] ~ EN CHARGE DE Y DANS LE gouvernement Z1 DE L'entité politique Z2

Cette définition contient également les deux positions actanciellles qui nous intéressent (en plus de contenir aussi une troisième qui exprime ce dont un ministre est en charge).

Jean Chrétien occupe ici la position actanciellle de X.

C) Type IIb.

[4.6]

A1) *son chef de cabinet, Jean Pelletier*

A2) *le premier ministre* [4.5]

B) CHEF DE CABINET est clairement une sorte de CHEF, une sorte de RESPONSABLE. La FP de RESPONSABLE2 (considérée par le DiCo3 comme un quasi-synonyme de CHEF) est la suivante :

[individu X] ~ DE L'organisme/DE L'activité Y

L'individu et l'organisme/l'activité sont clairement exprimables auprès de la lexie CHEF DE CABINET : *Pierre Dupont est le chef du cabinet du ministre de l'intérieur.* La FP de CHEF DE CABINET contient donc deux positions actanciellles : un individu X est le chef de cabinet de l'organisme (ministère) Y.

D'une part, *Jean Pelletier*, qui est un ASynS de *son chef de cabinet*, occupe ici la position de X. D'autre part, *le premier ministre*, qui est l'antécédent du déterminant de *son chef de cabinet* et qui est un nom typique de l'actant Y de CHEF DE CABINET, occupe la position de Y.

C) Type IIa (pour l'actant Y), type IIc (pour l'actant X).

[4.7]

A1) *le moment le plus difficile pour M. Chrétien*A2) *de faire ses adieux*

B) Le NPR définit MOMENT comme ‘Espace de temps limité (relativement à une durée totale) considéré le plus souvent par rapport aux faits qui le caractérisent’ et donne les exemples suivants : *moments de la vie, de l'existence; moment de l'année; moment où un événement s'est produit*. Le sens ‘les faits qui le caractérisent’ indique un participant de la situation et les exemples cités montrent que ce participant est exprimable auprès de MOMENT.

En plus, le DiCo3 considère POINTII.2 comme un quasi-synonyme de MOMENT, notamment dans les exemples suivants : *Arrivé à ce point de son raisonnement, le roi se sentit plein de honte et de remords. C'étaient là les points culminants de la journée de Poirot*. Or, la FP de POINTII.2 dans le DiCo3 contient une position actancielle similaire :

~ DE L'événement X

Tout cela démontre que la FP de MOMENT contient une position actancielle : l'événement (dont c'est le moment). L'ancre (*faire ses adieux*) qui dénote un événement, occupe dans notre exemple cette position actancielle. Elle peut être facile à trouver car, bien qu'elle ne fasse pas partie du groupe nominal du défini, elle en est un dépendant syntaxique : *Le moment le plus difficile pour M. Chrétien a été de faire ses adieux*.

Il est intéressant de remarquer aussi que le défini est au superlatif et le superlatif a tendance à aller de pair avec le défini.

C) Type IIc.

[4.8]

A1) *ses adieux à M. Trudeau*A2) *M. Chrétien*

B) Le NPR définit ADIEU comme ‘Fait de prendre congé, de se séparer de qqn’. Bien que seulement un participant de la situation (qqn) apparaisse de façon explicite

dans cette définition, il est clair que deux sont nécessaires pour comprendre la définition : quelqu'un prend congé de quelqu'un, quelqu'un se sépare de quelqu'un. Ces deux participants sont facilement exprimables auprès de ADIEU : *Lucille a dit adieu à Susanne.*

Par ailleurs, la FP de BONSOIRI (un quasi-synonyme de ADIEU) contient également deux positions actanciennes selon le DiCo3 :

~ [PAR L'individu X] À L'individu Y

La définition de ADIEU contient donc deux positions actanciennes : un individu X dit/fait adieu à Y. *M. Trudeau*, qui est un ASynS du syntagme *ses adieux*, occupe la position de Y. *M. Chrétien*, qui est l'antécédent du déterminant du défini et dont INDIVIDU est un hyperonyme, occupe la position de X.

C) Type IIb (pour l'actant Y) et type IIc (pour l'actant X).

[4.9]

A1) *la gare d'Ottawa*

B) Rappelons que la FP de GARE suivant le DiCo0 est :

~ DU lieu Y [POUR LE train X]

Ottawa, qui est un ASynS du défini, occupe la position de Y.

C) Type IIb.

[4.10]

A1) *la dépouille de M. Trudeau*

B) Le NPR définit DÉPOUILLE (MORTELLE) comme 'le corps humain après la mort' et la considère comme synonyme de CADAVRE. Il s'agit clairement de la dépouille de quelqu'un, du cadavre de quelqu'un (*la dépouille de Pierre, de Marie, d'un homme, d'une femme, du président, du charpentier, etc.*).

Qui plus est, le DiCo1 propose la FP suivante pour CORPS(1)III.2 (qu'il considère comme quasi-synonyme de CADAVRE et de DÉPOUILLE) :

~ de individu/animal mort X

Tout cela indique que la définition de DÉPOUILLE contient une position actancielle : l'individu (dont c'est la dépouille).

M. Trudeau, qui est un ASynS du défini occupe cette position actancielle.

C) Type IIb.

[4.11]

A1) *le quai de la gare*

B) L'article GARE du DiCo contient la FL suivante :

{Endroit dans une G. où les voyageurs montent dans
les trains} quai

Autrement dit, le défini et l'ancre sont liés par une FL non standard.

C) Type III.

[4.12]

A1) *la gare*

A2) *la gare d'Ottawa* [4.9]

B) Le défini est anaphorique, il contient le même nominal que son antécédent.

C) Type Ia.

[4.13]

A1) *les fils Trudeau, Justin et Sacha*

B) La FP de FILS dans DiCo3 est :

[X EST] ~ DE L'individu Y1/Y2

Trudeau, qui est un ASynS du syntagme *les fils*, occupe ici la position de Y. *Justin* et *Sacha*, qui sont apposés au défini occupent la position de X.

C) Type IIb.

[4.14]

A1) *son bureau*

A2) *le premier ministre* [4.5]

B) La FP de BUREAU^{II} dans le DiCo3 est :

~ UTILISÉ PAR L'individu X

le premier ministre, qui est l'antécédent du déterminant du défini, occupe ici la position de X.

C) Type IIc.

5.4.2 Deuxième lieu : l'école

Nous analysons ici trois textes contenant 50 expressions définies.

Cinquième texte

Mais où est donc passé Antoine ?

Dans nos éditions du 20 février (1), nous relations l'histoire (2) du petit Antoine, 12 ans, viré en janvier de l'école (3) néerlandophone qu'il fréquentait à Bruxelles. Ce ne sont pas des résultats médiocres ou une conduite répréhensible qui sont à la source (4) de cette sentence (5), mais bien l'inspection (6) linguistique exhibant une disposition des lois (7) linguistiques de 1963 : un élève résidant en dehors de Bruxelles ne peut s'inscrire dans une école bruxelloise de régime linguistique différent de celui de sa Région. Le tort (8) d'Antoine est d'être wallon, tout simplement. Il habite à Ottignies : il ne pouvait donc s'inscrire dans une école flamande de Bruxelles sauf si sa langue (9) maternelle était le néerlandais (10).

L'affaire (11) est déplorable : renvoyé en plein milieu de l'année (12) scolaire, le jeune garçon (13) a dû trouver en catastrophe une nouvelle école qui veuille bien l'accueillir. Lui et ses parents (14) sont extrêmement déçus qu'il n'ait pu continuer...

[5.1]

A1) le 20 février

B) Cette date doit être interprétée par rapport au moment de l'énonciation.

C) Type VI.

[5.2]

A1) *l'histoire du petit Antoine*

B) La FP de HISTOIREIII.1 dans le DiCo3 est :

~ DE L'individu X

Antoine, qui est un ASynS du défini, occupe ici la position de X.

C) Type IIb.

[5.3]

A1) *l'école néerlandophone qu'il fréquentait à Bruxelles*

B) Dans l'article ÉCOLEI.1a du DiCo1, on trouve la FL suivante :

{Real3} aller [à ART ~], fréquenter [ART ~], faire
ses études, se former [à < dans > ART ~]

Le défini et l'ancre sont ainsi liés par une FL standard.

C) Type III.

[5.4]

A1) *la source de cette sentence*

B) La FP de SOURCEII.1 dans le DiCo3 est :

[X EST] ~ DE Y POUR Z

cette sentence, qui est un ASynS du défini, occupe la position de Y.

C) Type IIb.

[5.5]

A1) *cette sentence*

A2) *Antoine, 12 ans, viré en janvier de l'école néerlandophone*

B) Le NPR définit SENTENCE comme 'décision rendue par un juge ou un arbitre'.

Cette définition indique directement un participant de la situation décrite par cette lexie : le juge ou l'arbitre qui rend la sentence. Il est clair cependant que la situation décrite par cette lexie doit avoir aussi d'autres participants car une sentence est for-

cément au sujet d'un événement impliquant une personne. Cette affirmation est confirmée par deux quasi-synonymes de SENTENCE : ARRÊT(2) et JUGEMENT.

L'article ARRÊT(2) du DiCo3 (qui est un QSyn de SENTENCE) a la FP suivante :

~ DE personne X SUR fait Z(Y) [IMPLIQUANT personne
Y PRESCRIVANT W]

La FP de JUGEMENTII.2 du DiCo3 (qui est également un QSyn de SENTENCE) est :

~ DE personne X SUR fait Y1 OU DE personne Y2
[STIPULANT W]

Dans notre exemple, *Antoine* occupe la position actancielle correspondant à Y de ces deux lexies, alors que son renvoi (*viré*) occupe la position correspondant à W.

C) Type IIc.

[5.6]

A1) *l'inspection linguistique* exhibant une disposition des lois linguistiques de 1963

B) Le NPR définit INSPECTION comme 'Examen attentif dans un but d'enquête, de contrôle, de surveillance, de vérification; travail, fonction d'inspecteur' et donne les exemples suivants : *faire, passer une inspection; inspection d'un navire; inspection de l'armée*. La définition indique de manière implicite un participant de la situation décrite par INSPECTION (examen de quelque chose) alors que les exemples (*un navire, l'armée*) l'indiquent de manière explicite et montre que ce participant est exprimable auprès de la lexie.

Qui plus est, selon le NPR, REVUE est synonyme de INSPECTION. Or, la FP de REVUE(1)I, suivant le DiCo3, est :

~ DE Y PAR L'individu X

Nous pouvons conclure que la FP de INSPECTION contient une position actancielle : ce qui est inspecté.

linguistique, qui est un ASynS de *l'inspection*, occupe cette position actancielle.

C) Type IIb.

[5.7]

A1) *les lois linguistiques de 1963*

B) Le NPR définit LOI comme ‘Règle ou ensemble de règles obligatoires établies par l'autorité souveraine d'une société et sanctionnées par la force publique.’ et donne les exemples suivants : *lois civiles, lois pénales, lois martiales, loi antitrust, loi d'amnistie*. Ces exemples indiquent clairement un domaine où la loi s'applique. Cette affirmation est confirmée par la FP de la lexie RÈGLEMENT(1), considérée comme un quasi-synonyme de LOI par le DiCo3 :

~ DE LA personne X SUR L'activité Y POUR LES individus/choses Z

La position actancielle Y de RÈGLEMENT(1) correspond à la position actancielle qu'occupe le domaine où la loi s'applique.

linguistique, qui est un ASynS de l'expression *les lois*, occupe cette position actancielle.

C) Type IIb.

[5.8]

A1) *le tort d'Antoine*

B) La FP de TORT2.a suivant le DiCo3 est :

~ DE LA personne X DE FAIRE Y

Antoine, qui est un ASynS du défini, occupe la position de X.

C) Type IIb.

[5.9]

A1) *sa langue*A2) *Antoine*

B) La FP de LANGUEII.1 suivant le DiCo3 est :

~ DES personnes X

Antoine, qui est l'antécédent du déterminant du défini et dont PERSONNE est un hyperonyme, occupe la position de X.

C) Type IIc.

[5.10]

A1) le néerlandais

B) En français, le nom des langues s'emploie surtout au défini : c'est clairement une information linguistique. Nous pouvons considérer ce type de défini comme une expression idiomatique.

C) Type IV.

[5.11]

A1) l'affaire

B) Le NPR défini AFFAIRE comme 'ensemble de faits créant une situation compliquée, où diverses personnes, divers intérêts sont aux prises'. Par conséquent, *affaire* ne peut pas être lié à un élément précis dans le texte mais renvoie plutôt à tout ce qui a été dit dans le paragraphe précédent.

C) Type VIII.

[5.12]

A1) l'année scolaire

B) *L'année scolaire* doit être interprétée par rapport au moment de l'énonciation – c'est une expression déictique.

C) Type VI.

[5.13]

A1) le jeune garçon

A2) *le petit Antoine, 12 ans*

B) Le lecteur peut établir une relation anaphorique entre *le jeune garçon* et *le petit Antoine, 12 ans* parce que tous les deux dénotent un être humain du sexe masculin de l'âge de l'enfance.

C) Type Ic.

[5.14]

A1) *ses parents*

A2) *Antoine*

B) La FP de PARENT2 dans le DiCo3 est :

[individu X] ~ DE individu Y

Antoine, qui est l'antécédent du déterminant du défini et dont INDIVIDU est un hyperonyme, occupe la position de Y.

C) Type IIc.

Sixième texte

5 février 2001

Deux collégiens réprimandés pour avoir menacé Bush par e-mail

LONDRES (Reuters) - Deux collégiens britanniques ont été sévèrement réprimandés pour avoir adressé par e-mail sur le site (1) internet de la Maison blanche des menaces de mort contre le président (2) George Bush.

Dès réception du message (3), qui prédisait l'assassinat (4) du président (5) américain, la CIA a été alertée et est rapidement remontée à la source (6), rapporte le Daily Mirror.

Leur enquête (7) les a menés jusqu'au collège (8) Blake, à Cannock, dans le centre (9) de l'Angleterre, d'où deux élèves âgés de 15 ans avaient envoyé leur sinistre blague (10) en utilisant l'ordinateur (11) de l'école (12).

« On ne les a pas suspendus mais ils ont subi une sévère mise en garde. Je pense qu'ils étaient assez stupéfaits des conséquences (13) de leurs actes » (14), a déclaré le proviseur (15) de l'établissement (16), Heather Bowman.

Un responsable de la CIA s'est de son côté contenté de dire: « Chaque menace dirigée contre le président (17) George W. Bush est prise très au sérieux ».

[6.1]

A1) le site internet de la Maison blanche

B) Antidote définit SITE [INTERNET] comme 'lieu où se trouve implanté un hôte Internet et qui est identifié par une adresse Internet; cet hôte' et donne les exemples

suivants : *site Internet d'un organisme, du gouvernement*. Or, un site Internet est forcément le site d'une personne (physique ou morale), comme le montrent les exemples. La FP de SITE [INTERNET] contient donc une position actancielle : le propriétaire du site.

la Maison Blanche, qui est un ASynS du syntagme *le site Internet*, occupe cette position actancielle.

C) Type IIb.

[6.2]

A1) *le président George Bush*

B) La FP de PRÉSIDENT2 dans le DiCo 3 est :

[X EST] ~ DE pays Y

George Bush, qui est un ASynS du syntagme *le président* occupe la position de X.

C) Type IIb.

[6.3]

A1) *le message*

A2) *avoir adressé par e-mail*

B) Antidote considère E-mail comme un anglicisme à remplacer entre autre par *message électronique*. Nous en concluons que E-MAIL est un quasi-synonyme de MESSAGE : le défini (*le message*) est anaphorique.

C) Type Ib.

[6.4]

A1) *l'assassinat du président américain*

B) La FP de ASSASSINAT dans le DiCo0 est :

~ PAR LA personne X DE L'individu Y

le président américain, qui est un ASynS du défini, occupe la position de Y.

C) Type IIb.

[6.5]

A1) *le président américain*

B) L'ancre *américain* (une dérivée adjectivale des Etats-Unis), qui est un ASynS du défini, occupe la position de Y dans la FP de PRÉSIDENT (voir [6.2]).

C) Type IIb.

[6.6]

A1) *la source*

A2) *le message* [6.3]

B) Rappelons que la FP de SOURCEII.1 est :

[X EST] ~ DE Y POUR Z

le message occupe la position de Y.

C) Type IIc.

[6.7]

A1) *leur enquête*

A2) *CIA* et *le message* [6.3]

B) Antidote définit ENQÊTE comme 'Ensemble de recherches effectuées à la demande d'une administration, d'une autorité judiciaire, pour éclairer une affaire litigieuse, notamment à l'aide de déclarations émanant d'experts, de personnes impliquées directement ou indirectement dans cette affaire' et donne les exemples suivants : *Ordonner, ouvrir, effectuer, clore une enquête. Enquête policière concernant une affaire de meurtre, de mœurs. Enquête judiciaire, publique, concernant une affaire de fraude impliquant des membres du gouvernement.* Tout cela implique deux participants de la situation : le premier est mentionné explicitement à la fois dans la définition ('une affaire litigieuse') et dans les exemples (*une affaire de meurtre, de mœurs; une affaire de fraude impliquant des membres du gouvernement*) alors que l'autre n'est explicite que dans les exemples (*policière, judiciaire* et aussi à travers les verbes supports dont il est le sujet comme *ouvrir* ou *effectuer*).

En plus, le DiCo3 considère que INVESTIGATION (un quasi-synonyme de ENQUÊTE) a la FP suivante :

~ PAR LA personne X DE Y

Nous pouvons en conclure que ENQUÊTE a deux positions actanciellées : une personne X fait l'enquête au sujet de Y.

la CIA, qui est l'antécédent du déterminant du défini et dont ORGANISATION (c'est-à-dire une PERSONNE MORALE) est un hyperonyme, occupe la position de X. *le message* occupe la position de Y.

C) Type IIc.

[6.8]

A1) *le collègue Blake, à Cannock*

B) Le référent du défini est identifié à l'aide de l'apposition d'un nom propre (et en spécifiant le nom du lieu où il se trouve).

C) Type Va.

[6.9]

A1) *le centre de l'Angleterre*

B) La FP de CENTREI.2 dans le DiCo3 est :

~ DE X

Angleterre occupe ici la position de X.

C) Type III.

[6.10]

A1) *leur sinistre blague*

A2) *deux élèves âgés de 15 ans*

B) Étant donné les verbes supports de BLAGUE que Beauchesne (2001 : 41) énumère – tels que COLLER, (RA)CONTER, DÉBITER, DIRE, FAIRE, (FAIRE) GOBER, LÂCHER, POSER et POUSSER – la situation décrite par BLAGUE implique un participant qui raconte, débite, dit, etc. une blague. Bien entendu, ce parti-

cipant est facilement exprimable auprès de la lexie : *la boulangère a raconté une blague, la blague de Pierre*, etc. Cela indique qu'il y a une position actancielle dans la FP de BLAGUE pour la personne qui raconte, dit etc. une blague.

On retrouve également cette position actancielle dans la FP de PLAISANTERIE qui, suivant le DiCo0, est un quasi-synonyme de BLAGUE :

~ DE LA personne X AU SUJET DE Y [ADRESSÉE À LA
personne Z]

deux élèves âgés de 15 ans, qui est l'antécédent du déterminant du défini (*leur sinistre blague*) est dont INDIVIDU est un hyperonyme, occupe cette position actancielle.

C) Type IIc.

[6.11]

A1) *l'ordinateur de l'école*

B) Beuchesne (2001 : 253) mentionne une série de verbes supports et des verbes de réalisation pour ORDINATEUR : *interagir avec un ~ ; tapoter à l'~ ; être devant son ~ ; allumer/éteindre son ~*. Ces verbes nécessitent quelqu'un (un utilisateur) qui interagit avec un ordinateur, qui « tapote » à l'ordinateur, etc.

Par ailleurs, la FP de la lexie OUTIL du DiCo0 (nous considérons un ordinateur comme un outil) contient également une position actancielle qui dénote l'utilisateur :

~ UTILISÉ PAR L'individu X POUR FAIRE Y

La définition de ORDINATEUR contient donc une position actancielle qui dénote son utilisateur.

l'école occupe cette position actancielle.

Rappelons que la LEC distingue souvent plusieurs lexies à l'intérieur d'un vocable. Nous avons ici une occurrence de ÉCOLEI.1a (qui est synonyme de ÉTABLISSEMENT). Or, les utilisateurs de l'ordinateur en question pourraient être dénotés par la lexie ÉCOLEI.1b (dont le sens approximatif est 'ensemble d'individus qui étudient ou travaillent à l'école I.1a'). Nous pouvons considérer ce type comme un cas particulier de lien actanciel.

C) Type IIb.

[6.12]

A1) *l'école*

A2) *le collège Blake* [6.8]

B) Selon le NPR, COLLÈGE et ÉCOLE sont des quasi synonymes. (Plus précisément, il sont des synonymes à intersection – cf. Mel'čuk *et al.* (1995 : 130)).

C) Type Ib.

[6.13]

A1) *les conséquences de leurs actes*

B) Le NPR définit CONSÉQUENCE comme 'Suite qu'une action, un fait entraîne' et donne les exemples suivants : *Entrevoir, prévoir les conséquences de qqch., évaluer les conséquences possibles de qqch. accepter, subir les conséquences de ses actes*. Un participant de la situation linguistique est mentionné à la fois dans la définition ('une action, un fait') et dans les exemples (*qqch., ses actes*).

On retrouve également une position actancielle similaire dans la FP de SUITE(1)I.4 que le DiCo3 considère comme un quasi-synonyme de CONSÉQUENCE :

[X EST] ~ DU fait Y

La FP de CONSÉQUENCE contient donc une position actantielle dénotant le fait qui a une conséquence.

leurs actes, qui est un ASynS du syntagme *les conséquences* occupe ici cette position.

C) Type IIb.

[6.14]

A1) *leurs actes*

A2) *deux élèves âgés de 15 ans*

B) La FP de ACTE(1)3 dans le DiCo3 est :

~ DE Y(X) [DE individu X]

deux *élèves âgés de 15 ans*, qui est l'antécédent du déterminant (*leurs*) du défini et dont INDIVIDU est un hyperonyme, occupe la position de X.

C) Type IIc.

[6.15]

A1) *le proviseur de l'établissement, Heather Bowman*

B) Le NPR défini PROVISEUR comme 'Fonctionnaire de l'enseignement qui dirige un lycée'. 'Fonctionnaire' et 'un lycée' sont deux participants de la situation que décrit PROVISEUR. Ces deux participants sont facilement exprimables auprès de la lexie : *Pierre est le proviseur du Lycée Danton*.

En plus, toujours selon le NPR, PROVISEUR et DIRECTEUR sont synonymes. La FP de DIRECTEUR dans le DiCo3 est :

[X EST] ~ DE établissement Y

Nous pouvons conclure que la FP de PROVISEUR contient deux positions actanciennes : l'individu X est proviseur dans l'établissement Y.

établissement, qui est un nom typique de Y, occupe dans cet exemple la position de Y. *Heather Bowman*, qui est apposé au défini et dont INDIVIDU est un hyperonyme, occupe la position de X.

C) Type IIa (pour Y) et type IIb (pour X).

[6.16]

A1) *l'établissement*

A2) *le collègue Blake* [6.8]

B) Le NPR défini COLLÈGE comme 'établissement d'enseignement' (et plus spécifiquement comme 'établissement du premier cycle du second degré'). Autrement dit, COLLÈGE dénote un type d'établissement d'enseignement.

Le défini (*l'établissement*) est un hyperonyme de l'ancre (*le collègue Blake*).

C) Type Ic.

[6.17]

A1) le président **George W. Bush**

A2) *le président* *George Bush* [6.2]

B) Nous avons déjà mentionné dans [6.2] que la FP de PRÉSIDENT2 est :

[X EST] ~ DE pays Y.

George W. Bush, qui est un ASynS du défini (*le président*) occupe la position de X.

En outre, ce défini est anaphorique, son antécédent étant *le président* dans [6.2].

C) Type IIb (pour l'ancre apposée) et type Ia (pour l'ancre dans [6.2]).

Septième texte

Les films (1) de la semaine (2)

12 mars 2001

- « La langue (3) des papillons » (4) (*La Lengua de las mariposas*) de José Luis Cuerda (Espagne - 1H35) - Hiver 1936 en Galice, quelques mois avant l'explosion (5) de la guerre civile (6) espagnole, le petit Moncho (Manuel Lozano Obispo), âgé de huit ans, doit aller à l'école (7) pour la première fois (8). Il est terrorisé, car il est persuadé que le maître (9), Don Gregorio (Fernando Fernan Gomez), bat les enfants (10). Le vieux professeur (11) va pourtant lui inculquer l'amour (12) de la connaissance (13) et la liberté (14) de penser... jusqu'à ce que la guerre (15) civile vienne brouiller les cartes (16). Sur une trame d'un grand classicisme -- l'éducation (17) contre la barbarie (18) -- Cuerda retrace l'émouvante relation (19) entre le vétéran Fernando Fernan Gomez et le jeune (et excellent) Manuel Lozano Obispo.

[7.1]

A1) les films de la semaine

B) Le défini (*les films*) est spécifié par une structure possessivale (*de la semaine*) sans avoir un lien lexical avec l'ancre.

C) Type Vb.

[7.2]

A1) *la semaine*

B) Le défini doit être interprété par rapport au moment de l'énonciation, il est anaphorique.

C) Type VI.

[7.3]

A1) la langue des papillons

B) LANGUE est un vocable fortement polysémique et il n'est pas évident dans quel sens il faut interpréter *langue* ici. Toutefois, qu'il s'agisse de la lexie dont le sens est 'Organe charnu, musculeux, allongé et mobile, placé dans la bouche' ou de celle dont le sens est 'Système d'expression et de communication commun à un groupe social (communauté linguistique)', deux définitions proposées par le NPR, il est clair que cette lexie évoque une situation avec un participant qui a une langue. Ce participant est exprimable auprès de la lexie, comme en témoigne notre exemple.

La définition de LANGUE (aussi bien dans le sens de 'organe' que dans le sens de 'système de communication') contient donc une position actantielle pour l'être dont c'est la langue : *les papillons*, qui est un ASynS de *la langue*, occupe ici cette position actancielle.

C) Type IIb.

[7.4]

A1) *les papillons*

B) Il s'agit des papillons en général.

C) Type VIIa.

[7.5]

A1) *l'explosion de la guerre civile espagnole*

B) La FP de EXPLOSIONII.1 dans le DiCo3 est :

~ DU fait X

la guerre civile, qui est un ASynS du défini *l'explosion*, occupe ici la position de X.

C) Type IIb.

[7.6]

A1) *la guerre civile espagnole*

B) La FP de GUERREI dans le DiCo3 est :

~ ENTRE L'ensemble-2->individu X ET

L'ensemble-2->individu Y POUR Z

Cet article contient également la FL suivante :

{X et Y = citoyens d'un même pays} civile, intes-
tine | postpos

Cela signifie que *civile*, qui est un ASynS du syntagme *la guerre*, occupe ici (simul-
tanément) la position de X et celle de Y.

espagnole, qui est un ASynS du défini, occupe ici la position de Z.

C) Type IIb.

[7.7]

A1) *aller à l'école*

B) Dans l'article ÉCOLEI.1a du DiCo, on trouve la FL suivante :

{Real3} aller [à ART ~], fréquenter [ART ~], faire
ses études, se former [à <dans> ART ~]

Autrement dit, *aller à l'école* est une fonction lexicale de ÉCOLE. Le défini est
obligatoire si *école* n'a pas de modificateur.

C) Type III.

[7.8]

A1) *la première fois*

B) L'expression *la N-ième fois* est une expression semi figée qui, dans sa forme non
marquée, est surtout au défini. L'expression *une première fois* ne serait possible que
si elle était accompagnée de l'expression *une deuxième fois*.

C) Type IV.

[7.9]

A1) *le maître, Don Gregorio*

A2) *l'école* [7.7]

B) D'une part, la FP de MAÎTREII du DiCo3 est :

[X EST] ~ DE L'individu Y DE LA matière Z

Don Gregorio, qui est un ASynS du défini *le maître*, occupe ici la position de X.

D'autre part, l'article ÉCOLEI.1a du DiCo1 contient la FL suivante :

{S1QReal1} maître [d'~] //enseignant#1; instituteur; professeur, "fam" prof

Autrement dit MAITRE est une valeur d'une FL de ÉCOLE.

C) Type IIb (pour *Don Gregorio*) et type III (pour *l'école*).

[7.10]

A1) les enfants

A2) *le maître* [7.9]

B) L'article MAITREII du DiCo3 contient la FL suivante :

{S2} disciple; élève

Or, selon le NPR, *enfant scolarisé* est synonyme de ÉLÈVE. Cela signifie que, à travers élève, on peut établir un lien lexical entre *les enfants* et *le maître*.

C) Type III.

[7.11]

A1) le vieux professeur

A2) *le maître* [7.9]

B) Dans l'article MAITREII du DiCo3, on trouve :

{QSyn} enseignant, instituteur, professeur; gourou, mentor.

Autrement dit, MAÎTRE et PROFESSEUR sont quasi synonymes.

C) Type Ib.

[7.12]

A1) l'amour de la connaissance

B) La FP de AMOURI.2 du DiCo2 est :

~ DE individu X POUR entité Y

la connaissance, qui est un ASynS du défini *l'amour*, occupe ici la position de Y.

C) Type IIb.

[7.13]

- A1) la connaissance
- B) Il s'agit de la connaissance en général.
- C) Type VIIb.

[7.14]

- A1) la liberté
- B) Il s'agit de la liberté en général.
- C) Type VIIb.

[7.15]

- A1) la guerre civile
- A2) **la guerre civile** espagnole [7.6]
- B) C'est un défini anaphorique.
- C) Type Ia.

[7.16]

- A1) brouiller les cartes
- B) Le NPR mentionne l'expression *brouiller les cartes* dont le sens est 'compliquer, obscurcir volontairement une affaire'. Nous en concluons qu'il existe une locution BROUILLER LES CARTES où seul l'article défini est possible.
- C) Type IV.

[7.17]

- A1) l'éducation
- B) Il s'agit de l'éducation en général.
- C) Type VIIb.

[7.18]

A1) *la barbarie*

B) Il s'agit de la barbarie en général.

C) Type VIIb.

[7.19]

A1) *l'émouvante relation entre ... Gomez et Obispo*

B) Le NPR définit RELATION comme 'Caractère de deux ou plusieurs choses entre lesquelles existe un lien' et l'illustre par les exemples suivants : *établir une relation entre deux phénomènes; mettre deux événements en relation, relation entre deux variables*. Le sens 'deux ou plusieurs choses' et les expressions *entre deux phénomènes, deux événements, deux variables* que l'on trouve dans les exemples indiquent que RELATION contrôle deux positions actanciennes.

Par ailleurs, AMITIÉ², dont RELATION est un terme générique suivant le DiCo3 contient également deux positions actanciennes dans sa FP :

~ ENTRE X ET Y

Les ancres remplissent ces positions actanciennes.

C) Type IIb.

5.4.3 Troisième lieu : la mairie

Nous analysons ici trois textes contenant 43 expressions définies.

Huitième texte

Le goût (1) de l'eau (2)

« L'eau (3) a beaucoup d'importance pour moi, c'est un élément qui a marqué mon imagination » (4), déclare Chantal Deschamps. La mairesse (5) de Repentigny a grandi entre les rives (6) du fleuve (7) St-Laurent et de la rivière (8) L'Assomption. Issue d'une des lignées (9) fondatrices du lieu (10), elle évoque avec émotion 330 ans d'histoire familiale sur la « presqueîle » (11) de Repentigny. «Jusqu'à tout récemment, ma famille (12) a possédé plusieurs îles dans le St-Laurent. Aussi, quand j'ai été élue à la mairie (13) de Repentigny, il était tout naturel que le fleuve (14) constitue une de mes préoccupations (15). J'avais à coeur de mettre en oeuvre des projets afin d'assurer la préservation (16) du littoral (17) et d'en faire profiter la population. » (18)

[8.1]

A1) le goût de l'eau

B) La FP de GOÛTI.1 dans le DiCo0 est :

~ DE Y DE LA substance X [PERÇU PAR L'être animé Z]

l'eau, qui est un ASynS du défini *le goût*, occupe la position de X.

C) Type IIb.

[8.2]

A1) l'eau

B) Il s'agit de l'eau en général.

C) Type VIIb.

[8.3]

A1) *l'eau*

B) Il s'agit toujours de l'eau en général.

C) Type VIIb.

[8.4]

A1) *mon imagination*

A2) *Chantal Deschamps*

B) La FP de IMAGINATION1 du DiCo0 est :

~ DE L'individu X [AU SUJET DE Y]

Chantal Deschamps, qui est l'antécédent du déterminant (*mon*) du défini et dont INDIVIDU est un hyperonyme, occupe la position de X.

C) Type IIc.

[8.5]

A1) *la mairesse de Repentigny*

B) Le NPR définit MAIRE comme 'En France, Premier officier municipal élu par le conseil municipal, parmi ses membres, et qui est à la fois une autorité locale et l'agent du pouvoir central' et donne les exemples suivants : *Le maire de Paris. Les maires d'arrondissement*. Le sens 'Premier officier municipal' indique un participant de la situation linguistique alors que les expressions *Paris* et *arrondissement* en indiquent un deuxième. Bien entendu, ces deux participants sont facilement exprimables : *Bertrand Delanoë est le maire de Paris*. Nous pouvons en conclure que la définition de MAIRE contient deux positions actanciennes qui correspondent à ces participants : un individu X est maire d'une localité Y.

On retrouve également deux positions actanciennes dans la FP de la lexie RESPONSABLE (MAIRE peut être considérée comme une sorte de RESPONSABLE) dont la FP suivant le DiCo3 est :

[individu X] ~ DE L'organisme/DE L'activité Y
Repentigny, qui est un ASynS du défini (*la mairesse*) occupe ici la deuxième position actancielle.

C) Type IIb.

[8.6]

A1) *les rives du fleuve St-Laurent et de la rivière L'Assomption*

B) La FP de RIVE du DiCo3 est :

~ DU cours d'eau X

le fleuve St-Laurent et la rivière L'Assomption, qui sont des ASynS du défini (*les rives*) et dont COURS D'EAU est un hyperonyme, occupent la position de X. En effet, l'article FLEUVEI du DiCo0 contient la FL suivante :

{Gener} _cours d'eau_

C) Type IIb.

[8.7]

A1) *le fleuve St-Laurent*

B) La FP de FLEUVEI dans le DiCo0 est :

~ [SE TERMINANT DANS L'étendue d'eau X]

Cela signifie que l'ancre *St-Laurent* n'occupe pas une position actancielle de FLEUVE, elle est un nom propre apposé au défini.

C) Type Va.

[8.8]

A1) *la rivière L'Assomption*

B) Comme dans le cas de FLEUVE, la forme propositionnelle de la définition de RIVIÈRE n'a pas de position actancielle que l'ancre pourrait occuper : elle est un nom propre apposé au défini.

C) Type Va.

[8.9]

A1) *les lignées fondatrices du lieu*

B) Le NPR définit LIGNÉE comme ‘Ensemble des descendants (d'une personne)’ – le sens ‘une personne’ indique clairement un participant de la situation évoqué par la lexie. Bien entendu, ce participant est facilement exprimable : *la lignée de Pierre, de ce prince, etc.* La FP de LIGNÉE contient donc une position actancielle : un individu (qui a des descendants).

On retrouve également cette position actancielle dans l'article NAISSANCEI.2 (qui est un quasi-synonyme de LIGNÉE, dans le sens que l'on trouve par exemple dans *Les gens de la région lui vouaient le respect et l'adulation réservés aux gens de haute naissance.*) du DiCo0 qui a la FP suivante :

~ DE L'individu X

fondatrice (dérivée sémantique adjectivale (A₁) de FONDATEUR, dont INDIVIDU est un hyperonyme) occupe cette position actancielle.

C) Type IIb.

[8.10]

A1) *le lieu*A2) **Repantigny**

B) REPANTIGNY, qui est un nom propre d'un lieu, est un hyponyme de LIEU. Autrement dit, le défini de notre exemple est un hyperonyme de l'ancre.

C) Type Ic.

[8.11]

A1) *la « presqu'île » de Repantigny*

B) Le NPR définit PRESQU'ÎLE comme ‘Partie saillante d'une côte, rattachée à la terre par un isthme, une langue de terre’ et donne l'exemple suivant : *La presqu'île de Quiberon.* La définition, cet exemple et l'expression *la presqu'île de Repantigny* suggèrent que la lexie PRESQU'ÎLE contrôle une position actancielle que peut occuper une côte.

C) Type IIb.

[8.12]

A1) ma famille

A2) **Chantal Deschamps**

B) La FP de FAMILLE1.b du DiCo0 est :

~ DE L' (ensemble-2->) individu X

Chantal Deschamps, qui est l'antécédent du déterminant du défini et dont INDIVIDU est un hyperonyme, occupe la position de Y.

C) Type IIc.

[8.13]

A1) la mairie de Repentigny

B) La FP de MAIRIE dans le DiCo3 est :

~ DE LA localité X1 DE L'ensemble de individu X2

Repentigny, qui est un ASynS du défini et dont LOCALITÉ un hyperonyme, occupe la position de X1.

C) Type IIb

[8.14]

A1) le fleuve

A2) **le fleuve St-Laurent** [8.7]

B) Le défini est anaphorique, il contient le même nominal que son antécédent.

C) Type Ia.

[8.15]

A1) mes préoccupations

A2) **Chantal Deschamps**

B) La FP de PRÉOCCUPATION du DiCo3 est :

[Y EST] ~ DE L'individu X

Chantal Deschamps, qui est l'antécédent du déterminant du défini et dont INDIVIDU est un hyperonyme, occupe la position de X.

C) Type IIc.

[8.16]

A1) la préservation du littoral

B) Le NPR définit PRÉSERVATION comme 'Action, moyen de préserver, de se garantir' et donne les exemples suivants : *Préservation des droits, des intérêts, du patrimoine d'une personne. Préservation des sites, de l'environnement.* L'entité protégée/préservée n'est qu'implicite dans cette définition (préserver quelque chose) mais elle est explicitée par les exemples (*les droits, des intérêts, du patrimoine d'une personne, etc.*). Nous en concluons que la FP de PRÉSERVATION contient une position actancielle pour l'entité protégée/préservée.

On retrouve également cette position actancielle dans la FP de PROTECTION qui, selon le NPR, est un synonyme de PRÉSERVATION. Or, dans le DiCo1, PROTECTIONI.1a a la FP suivante :

~ PAR entité X DE entité Y CONTRE Z

le littoral, qui est un ASynS du défini, occupe cette position actancielle.

C) Type IIb.

[8.17]

A1) le littoral

A2) *les rives* [8.6]

B) L'article RIVE du DiCo3 contient la FL suivante :

{Syn} côte; littoral, RIVAGE#1.

Autrement dit, LITTORAL et RIVE sont synonymes.

C) Type Ib.

[8.18]

A1) la population

A2) entre les *rives* (6) du fleuve (7) St-Laurent et de la rivière (8) L'Assomption

B) La FP de POPULATION est :

[ensemble-2->individu X] ~ DE territoire Y.

L'ancre, dont TERRITOIRE est un hyperonyme, occupe la position de X.

C) Type IIb.

Neuvième texte

Passy-sur-Seine se passe de campagne électorale pour les municipales (1)

lundi 19 février 2001, 13h25

PASSY-SUR-SEINE, 19 fév (AFP) - Les 36 habitants (2) de Passy-sur-Seine, plus petite commune de Seine-et-Marne, se passent de campagne électorale, et savent que de toute façon, le soir (3) du scrutin (4), pratiquement toutes les familles (5) du village (6) seront représentées au conseil municipal (7).

Le maire (8) sortant Philippe Germann, sans étiquette mais qui se dit divers droite, brigue un quatrième mandat. « Ce sera le dernier » (9) promet cet agriculteur (10) âgé de 66 ans, fier d'être né dans la chambre (11) où il se couche chaque soir.

*Il mène l'unique liste (12) candidate pour conduire les destinées (13) de ce petit vil-
lage (14) qui ne compte aucun commerce.*

[9.1]

A1) les municipales

A2) **Passy-sur-Seine**

B) Selon le NPR, l'expression *les municipales* se définit comme 'élection des conseillers municipaux par les citoyens'. Il est clair que cette élection est liée à une localité (*les municipales de la ville ou du village X*). Nous en concluons que la définition de MUNICIPALES contient une position actancielle qui dénote une municipalité/localité.

Passy-sur-Seine, la seule lexie du texte dont LOCALITÉ est un hyperonyme, occupe ici cette position.

C) Type IIc.

[9.2]

A1) *les 36 habitants de Passy-sur-Seine, plus petite commune de Seine-et-Marne*

B) Selon le NPR, la définition de HABITANT est ‘Personne qui réside habituellement en un lieu déterminé’. La situation évoquée par cette définition implique deux participants : quelqu’un est habitant d’un lieu. Par ailleurs, selon ce même dictionnaire, HABITANT et POPULATION sont synonymes. Nous avons mentionné dans [8.18] que la FP de cette dernière lexie est :

[ensemble-2->individu X] ~ DE territoire Y

Comme POPULATION, HABITANT a également une position actancielle pour le territoire habité : on le voit dans les expressions comme *les habitants d'un pays; les habitants de Sens s'appellent les Sénonais; habitants des villes, des banlieues, d'une île, de la campagne*, etc. proposées par le NPR.

Passy-sur-Seine, qui est un ASyntS du défini, occupe ici cette position actancielle.

C) Type IIb.

[9.3]

A1) *le soir du scrutin*

B) La FP de SOIRI du DiCo3 est :

~ DE LA journée X.

le scrutin, qui est un ASynS du défini, occupe ici la position de X. Typiquement, l’actant X désigne une journée mais il peut aussi – comme c’est le cas maintenant – désigner un autre repère temporaire. Implicitement, on interprète ce défini comme dénotant le soir du jour du scrutin.

C) Type IIb.

[9.4]

A1) *le scrutin*

A2) *les municipales*

B) Nous avons mentionné dans [9.1] que, selon le NPR, l'expression *les municipales* se définit comme 'élection des conseillers municipaux par les citoyens'. SCRUTIN, quant à lui, est défini dans le NPR par 'Opération électorale, comprenant le dépôt des bulletins, le dépouillement, la proclamation des résultats; modalité particulière des élections'. Ces deux définitions permettent d'établir un lien de quasi-synonymie entre SCRUTIN et *les municipales*, les deux étant une forme d'élection.

C) Type Ib.

[9.5]

A1) *pratiquement toutes les familles du village*

B) Le village peut être identifié comme Passy-sur-Seine, étant donné que ce dernier est le nom propre d'un village. Le référent du défini inclut toutes les entités (c'est-à-dire toutes les familles) qui correspondent à la description que contient cette expression.

C) Type Vb.

[9.6]

A1) *le village*

A2) *plus petite commune de Seine-et-Marne*

B) L'article COMMUNE du NPR contient la définition suivante : 'La plus petite subdivision administrative du territoire, administrée par un maire, des adjoints et un conseil municipal' et renvoie à des lexies comme MUNICIPALITÉ, VILLAGE, VILLE. Autrement dit, COMMUNE est un quasi-synonyme de VILLAGE.

C) Type Ib.

[9.7]

A1) *le conseil municipal*

A2) *plus petite commune de Seine-et-Marne*

B) Le NPR définit CONSEIL MUNICIPAL comme 'composé de membres élus, chargés de régler les affaires de la commune'. La définition indique explicitement

un participant de la situation évoquée qui dénote une commune. Ce participant est facilement exprimable : *le conseil municipal du village X, de la ville Y, etc.* Cela suggère que la FP de CONSEIL MUNICIPAL contient une position actancielle pour la commune/municipalité dont c'est le conseil municipal.

Par ailleurs, le DiCo1 définit la FP de CONSEIL **III.1** de la manière suivante :

~ DE ensemble-2->individu X DE communauté Y DE W
 CONSEIL MUNICIPAL n'a plus la position actancielle équivalente à la position Y de CONSEIL (il est justement remplacé par l'adjectif MUNICIPAL), par contre, il a l'équivalent de la position W de CONSEIL et cette position peut être occupée par une commune/municipalité.

Nous concluons que COMMUNE est un nom typique de l'actant qui occupe la position actancielle de CONSEIL MUNICIPAL en question.

C) Type IIa.

[9.8]

A1) *le maire sortant Philippe Germann*

A2) *le village* [9.6]

B) Rappelons que dans [8.5] nous avons vu que la définition de MAIRE, comme celle de tout autre nom de responsable, contient au moins deux positions actanciell-les. *Philippe Germann*, qui est un ASynS du défini *le maire*, occupe ici la position de l'individu X. *village*, la seconde ancre, est un nom typique de l'actant qui occupe la position Y de la FP de MAIRE.

C) Type IIa (pour Y) et type IIb (pour X).

[9.9]

A1) *le dernier*

A2) *un quatrième mandat*

B) Le défini est une expression elliptique équivalent à l'expression *le dernier mandat*. Grâce à diverses parties du cotexte (*campagne électorale pour les municipa-*

les... le maire sortant... brigue un quatrième mandat), le lecteur peut l'interpréter comme *le dernier mandat pour être maire de Passy-sur-Seine*.

C) Type VIII.

[9.10]

A1) *cet agriculteur âgé de 66 ans*

A2) *le maire* [9.8]

B) Étant donné que le déterminant du nominal est un adjectif démonstratif ici, le lecteur doit chercher dans le texte un antécédent pour le défini. Le seul antécédent est *le maire* bien qu'il n'y ait aucun lien lexical entre AGRICULTEUR et MAIRE – à part le fait que INDIVIDU est l'hyperonyme des deux.

Le lecteur peut interpréter le défini comme faisant référence au maire sortant.

C) Type VIII.

[9.11]

A1) *la chambre où il se couche chaque soir*

B) La proposition subordonnée qui est rattachée au défini permet d'identifier son référent de manière unique dans ce texte.

C) Type Vb.

[9.12]

A1) *l'unique liste candidate*

B) Le NPR défini LISTE comme 'Suite de mots, de noms, de signes, généralement inscrits les uns au-dessous des autres' et donne toute une série d'exemples : *faire la liste des personnes présentes; liste des membres d'une société; liste méthodique et détaillée d'objets; liste des mets; liste de proscription; liste d'attente*. La définition et ces exemples indiquent que la FP de LISTE contient une position actancielle pour ce qui constitue cette liste.

On retrouve également cette position actancielle dans la FP de REGISTRE(1), qui est un quasi-synonyme de LISTE, suivant le DiCo3 :

~ DE Y [FAIT PAR X]

candidate, qui est un ASynS du défini *l'unique liste*, occupe cette position.

C) Type IIb.

[9.13]

A1) *les destinées de ce petit village*

B) Le NPR définit DESTINÉE comme ‘Destin particulier d'un être’ et l'illustre par les exemples suivants : *tenir entre ses mains la destinée de qqn; voici des milliers de siècles que notre pauvre humanité accomplit sa destinée sur la terre ; mon père se faisait de l'âme humaine et de sa destinée une idée sublime ; c'est sa destinée d'être parfaitement aimé ; etc.* La définition et les exemples indiquent que la FP de DESTINÉE contient une position actancielle pour la personne dont c'est la destinée. On retrouve cette position actancielle dans la FP de ÉTOILEIII (comme dans *Toutefois, elle pouvait remercier sa bonne étoile*) qui, suivant le DiCo3, est un quasi-synonyme de DESTINÉE :

~ DE personne X

ce petit village, qui est un ASynS du défini *les destinées* et qui peut aussi désigner les habitants de Passy-sur-Seine, occupe la position de X.

C) Type IIb.

[9.14]

A1) *ce petit village qui ne compte aucun commerce*

A2) *le **village*** [9.6]

B) Le défini est anaphorique, l'antécédent contient le même nominal que le défini.

C) Type Ia.

Dixième texte

Le journal (1) des municipales (2)

19 février 2001

Lyon.- Raymond Barre, député-maire depuis 1995, préside ce lundi (3) sa dernière séance (4) du conseil municipal (5). La passation de pouvoirs (6) avec son successeur (7) est prévue le 23 mars (8). Dans une lettre à ses administrés (9), le maire (10) sortant a appelé à voter « dès le premier tour » (11) pour le candidat (12) UDF-RPR-DL Michel Mercier : « Ainsi, Lyon ne basculera pas à gauche ».

[10.1]

A1) le journal des **municipales**

B) La FP de JOURNAL^{III} dans le DiCo3 est :

~ PAR LA chaîne X SUR Y

les municipales, qui est un ASynS du défini *le journal*, occupe la position de Y.

C) Type IIb.

[10.2]

A1) les municipales

B) Il s'agit des municipales qui se préparent au moment de l'énonciation – le défini est déictique.

C) Type VI.

[10.3]

A1) ce lundi

B) Le défini doit être interprété par rapport au moment de l'énonciation, il est déictique.

C) Type VI.

[10.4]

A1) *sa dernière séance du conseil municipal*

B) La FP de SÉANCE1 dans le DiCo3 est :

~ DES individus X À PROPOS DE Y

le conseil municipal, qui est un ASynS du défini *sa dernière séance*, occupe ici la position de X.

C) Type IIb.

[10.5]

A1) *le conseil municipal*

A2) *Lyon*

B) Nous avons déjà mentionné dans [9.7] que la définition de CONSEIL MUNICIPAL contient une position actancielle qu'occupe la commune/municipalité dont c'est le conseil municipal. *Lyon*, dont un des termes génériques est également MUNICIPALITÉ, occupe cette position dans notre exemple.

C) Type IIc.

[10.6]

A1) *la passation de pouvoirs avec son successeur*

A2) *Raymond Barre* et *député-maire*

B) La NPR définit l'expression *passation des pouvoirs* comme la 'transmission des pouvoirs à un autre, à d'autres' alors que TRANSMISSION y est définie comme 'action de transmettre' et illustrée par l'exemple *transmission d'un bien, d'un droit à une autre personne*. L'article du NPR de TRANSMISSION contient également l'expression *la transmission des pouvoirs* et la renvoie vers PASSATION. La situation que ces lexies décrivent implique trois participants : il s'agit de la passation d'un certain type de pouvoir d'une personne à une autre. Ces trois participants sont

exprimables auprès de la lexie : *la passation de pouvoirs de chef d'une personne à une autre*.

Le DiCo3 définit la FP de DÉLÉGATIONI, qu'il considère comme un quasi-synonyme de PASSATION, de façon semblable :

~ PAR LA personne X DU fait Y À LA personne Z

Tout cela suggère que la définition de la lexie PASSATION contient trois positions actanciennes : la personne X passe/transmet les pouvoirs Y à la personne Z.

Dans notre exemple, *son successeur* occupe la position de Z, *Raymond Barre* la position X et *député-maire* la position Y.

C) Type IIb (pour Z) et IIc (pour X et Y).

[10.7]

A1) *son successeur*

A2) ***Raymond Barre***

B) Antidote définit SUCCESSEUR comme 'Personne qui en remplace définitivement une autre dans une fonction' On voit que la situation décrite par SUCCESSEUR implique trois participants : une personne qui succède à une autre personne dans une fonction. Ces trois participants sont facilement exprimables auprès de la lexie : *Paul Martin est le successeur de Jean Chrétien*. Cela indique que la définition de SUCCESSEUR a trois positions actanciennes : une personne X succède à une autre personne Y dans une fonction Z.

Raymond Barre occupe la position de Y. L'ancre est syntaxiquement indépendante du défini mais elle est facilement identifiable car elle est l'antécédent du déterminant (*son*) du défini.

C) Type IIc.

[10.8]

A1) *le 23 mars*

B) Ce défini s'interprète en fonction de la date d'énonciation, il est déictique.

C) Type VI.

[10.9]

A1) *ses administrés*

A2) *le maire* [10.10]

B) Le NPR définit ADMINISTRÉ comme ‘personne soumise à une autorité administrative’ et donne comme exemple *le maire et ses administrés*. La définition indique que la situation évoquée par ADMINISTRÉ contient comme participant une autorité administrative et ce participant est exprimable, comme l’illustre l’exemple.

Par ailleurs, la FP de MAIRIE dans le DiCo3 est :

~ DE L’agglomération Y1 DE L’ensemble-2->individu
Y2

Parmi ses FL, on trouve la suivante :

{S2} arrondissement, localité, municipalité, vil-
lage, ville | Y1; administré | Y2

Tout cela indique que MAIRE est un nom typique d’un actant de ADMINISTRÉ.

C) Type IIa.

[10.10]

A1) *le maire sortant*

A2) *Lyon* et *Raymond Barre, député-maire*

B) Nous avons mentionné dans [8.5] les positions actantielles de la FP de MAIRE. *Lyon*, dont MUNICIPALITÉ est un hyperonyme, occupe la place de Y. *député-maire*, la deuxième ancre, est une expression coréférentielle avec le défini qui contient le même nominal.

C) Type IIc (pour *Lyon*) et type Ia (pour *député-maire*).

[10.11]

A1) *le premier tour*

A2) *les municipales* [10.2]

B) Le NPR définit TOUR (ou *tour de scrutin*) comme ‘chaque vote, dans un scrutin qui en comporte plusieurs’ et donne comme exemple : *le premier, le second tour (d’élections)*. SCRUTIN lui-même y est défini comme ‘opération électorale...’ Tout cela suggère que TOUR a une position actancielle : les élections. Or, nous avons mentionné dans [9.1] que MUNICIPALES dénote également une forme d’élections : son occurrence occupe ici la position actancielle de l’expression *le premier tour*.

C) Type IIc.

[10.12]

A1) *le candidat UDF-RPR-DL Michel Mercier*

B) Selon le NPR, CANDIDAT se définit comme ‘Personne qui postule une place, un poste, un titre’. Cette définition suggère deux participants de la situation décrite par CANDIDAT : une personne X est candidat pour une place/un poste/un titre Y. Les expressions comme *être candidat* ou *se porter candidat* confirment l’existence de la position X.

Michel Mercier, qui est un ASynS du défini *le candidat*, occupe cette position X.

C) Type IIb.

5.4.4 Quatrième lieu : le jardin public

Nous analysons ici deux textes contenant 35 expressions définies.

Onzième texte

François Gibault, 65 ans, avocat de causes troubles et d'une fille de Yann Piat, biographe de Céline, publie son autobiographie (1).

Le titre (2) du livre (3), Interdit aux Chinois (4) et aux chiens (5), il l'a vu sur une pancarte à Shanghai, il y a plus de vingt ans, à l'entrée (6) d'un jardin public réservé aux étrangers (7).

[11.1]

A1) son autobiographie

A2) **François Gibault**

B) Le NPR définit AUTOBIOGRAPHIE comme 'Biographie de l'auteur faite par lui-même' ce qui indique que l'auteur en question (un individu) est un participant de la situation dénotée par AUTOBIOGRAPHIE. Ce participant est facilement exprimable : *l'autobiographie de Gandhi, de l'auteur, du Président de la république, etc.* La définition et les exemples indiquent que AUTOBIOGRAPHIE a une position actancielle occupée par l'individu dont c'est l'autobiographie.

L'ancre, qui est l'antécédent du déterminant du défini et dont INDIVIDU est un hyperonyme, occupe ici cette position actancielle.

C) Type IIc.

[11.2]

A1) le titre du livre

B) La FP de TITRE(1) dans le DiCo3 est :

[X EST] ~ DU texte Y

le livre, qui est un ASynS du défini, occupe la position de X.

Par ailleurs, le NPR définit TITRE comme ‘désignation du sujet traité (dans un livre); nom donné (à une œuvre littéraire) par son auteur, et qui évoque plus ou moins clairement son contenu’ et donne l’exemple suivant : *les titres des livres sont souvent d’effrontés imposteurs*. Le fait que LIVRE se trouve même dans la définition proposée par le NPR suggère qu’il est le nom typique de l’actant Y.

C) Type IIa.

[11.3]

A1) *le livre*

A2) *son **autobiographie***

B) Nous avons donné dans [11.1] la définition du NPR pour AUTOBIOGRAPHIE. Or, l’article LIVRE du NPR considère comme synonyme la lexie BIOGRAPHIE et l’expression ‘*Livre racontant des événements, une vie, la vie d’hommes illustres*’. BIOGRAPHIE étant donc une sorte de LIVRE et AUTOBIOGRAPHIE étant une sorte de BIOGRAPHIE, on peut également conclure que AUTOBIOGRAPHIE est une sorte de LIVRE. Le défini (*le livre*) est anaphorique, son antécédent (*son autobiographie*) étant un de ces hyponymes.

C) Catégorie Ic.

[11.4]

A1) *les Chinois*

B) En l’absence d’autres informations, *les Chinois* s’interprète comme étant un générique.

C) Type VIIa.

[11.5]

A1) *les chiens*

B) En l'absence d'autres informations, *les chiens* s'interprète comme étant un générique.

C) Type VIIa.

[11.6]

A1) *l'entrée d'un **jardin public** réservé aux étrangers*

B) La FP de ENTRÉE est :

~ DU lieu Y UTILISÉE PAR individu X

jardin public, qui est un ASynS du défini et dont LIEU est un hyperonyme, occupe la position de Y.

C) Type IIb.

[11.7]

A1) *les étrangers*

A2) *interdit aux **Chinois*** [11.4]

B) Le NPR définit ÉTRANGER comme 'Personne dont la nationalité n'est pas celle d'un pays donné (par rapport aux nationaux de ce même pays)'. Cela suggère que la situation décrite par ÉTRANGER implique deux participants : le premier est une personne qui est étranger dans un pays. C'est le deuxième participant qui est intéressant pour nous : il est facilement exprimable auprès de ÉTRANGER, comme dans l'exemple *Il y a beaucoup plus d'étrangers en France / en Allemagne / en Espagne, etc. aujourd'hui qu'il y a cinquante ans*. La définition du NPR et l'exemple suggèrent que la FP de ÉTRANGER contient une position actancielle disponible pour un pays ou les nationaux de ce pays. L'ancre (*les Chinois*), qui dénote les nationaux de la Chine, occupe cette position.

C) Type IIc.

Douzième texte

Central Park fait appel à la charité (1)

AFP New York - 7 juin 1993

Frappé par de fortes restrictions budgétaires, le poumon (2) de New York, Central Park, le grand jardin public (3) au cœur (4) de Manhattan, est contraint de multiplier les appels (5) à la charité (6) et au bénévolat (7) pour assurer un entretien minimum alors que se pose le problème (8) de la reconversion (9) d'un immense lac artificiel en son centre (10). Entre 1990 et 1993, le budget (11) alloué par la ville (12) de New York aux espaces verts (13) est passé de 195 à 156 millions de dollars. Le nombre (14) des salariés (15) chargés de l'entretien (16) des jardins publics (17) a baissé de 1400, pour tomber à 3260. Et la ville (18) veut réduire ces effectifs (19) à 2600 en 1997. Si l'herbe (20) est encore régulièrement tondue, les massifs (21) de fleurs se font rares, des arbres déracinés par les tempêtes (22) de l'hiver gisent toujours sur le sol (23), les poubelles (24) sont moins souvent vidées et 30 p. cent des fontaines (25) publiques ne marchent pas, indique la porte-parole (26) d'une association de défense des parcs (27) new-yorkais, Linda Davidoff.

[12.1]

A1) la charité

B) Il s'agit de la charité en général.

C) Type VIIb.

[12.2]

A1) le poumon de New York, Central Park

B) Le NPR définit POUMON (au sens figuré) comme 'Ce qui fournit de l'oxygène' et donne les exemples suivants : *Ce parc est le poumon de la ville. Région qui est le poumon économique d'un pays.* Cette définition et les exemples suggèrent que

POUMON a deux positions actanciennes : quelque chose fournit l'oxygène à quelque chose.

New York, qui est un ASynS du défini, occupe la deuxième position actancielle. Qui plus est, un nom propre (*Central Park*) est apposé au défini.

C) Type IIb (pour *New York*) et type Va (pour *Central Park*).

[12.3]

A1) *le grand jardin public au cœur de Manhattan*

A2) *le poumon de New York* [12.2]

B) JARDIN PUBLIC est un nom typique du premier actant de POUMON, mentionné dans [12.2].

C) Type IIa.

[12.4]

A1) *le cœur de Manhattan*

B) Le défini et l'ancre sont liés par une FL : **Centr**(*Manhattan*) = *le cœur*. Cette FL nécessite ici l'article défini.

C) Type III.

[12.5]

A1) *les appels à la charité et au bénévolat*

B) La FP de APPELI.1 dans le DiCo3 est :

~ DE X À Y À Z (Y)

la charité et le bénévolat occupent ici la position de Z.

C) Type IIb.

[12.6]

A1) *la charité*

B) En l'absence d'autre information, ce défini doit être interprété de façon générale.

C) Type VIIb.

[12.7]

A1) *le bénévolat*

B) Comme le défini précédent, ce défini doit être interprété de façon générale.

C) Type VIIb.

[12.8]

A1) *le problème de la reconversion*

B) Le NPR définit PROBLÈME comme ‘Difficulté qu’il faut résoudre pour obtenir un certain résultat; situation instable ou dangereuse exigeant une décision’. Cette définition suggère deux participants : quelqu’un (X) doit résoudre quelque chose (Y). Le premier participant est exprimable auprès de la lexie à l’aide des verbes supports (*avoir un problème, faire face à un problème*) et de réalisation (*régler un problème, résoudre un problème*). Les expressions comme *problème de l’alcoolisme, du logement, du chômage* ou encore celles comme *problèmes familiaux, professionnels, sexuels* etc. montrent que le deuxième participant est également exprimable. Tout cela indique que PROBLÈME contrôle deux positions actancielles : X a le problème Y.

Par ailleurs, la FP de QUESTION2 (qui est quasi-synonyme de PROBLÈME) est semblable selon le DiCo3 :

~ DE Y POUR personne X

la reconversion, qui est un ASynS du défini, occupe ici la position de Y. (De façon moins directe, on pourrait même interpréter *Central Park* comme occupant la position X.)

C) Type IIb.

[12.9]

A1) *la reconversion d’un immense lac artificiel en son centre*

B) Le NPR définit RECONVERSION par ‘Conversion, transformation qui rétablit l’état primitif d’une organisation transformée’ et donne l’exemple suivant : *Re-*

conversion d'une fabrique de tanks en usine d'automobiles. La définition (notamment la transformation de quelque chose en quelque chose) et l'exemple suggèrent l'existence de deux positions actanciellles : la reconversion de X en Y.

un immense lac, qui est un ASynS du défini, occupe ici la position de X.

C) Type IIb.

[12.10]

A1) *son centre*

A2) ***Central Park***

B) La FP de CENTREL.1 suivant le DiCo3 est :

~ DE X

Central Park occupe ici la position de X

C) Type III.

[12.11]

A1) *le budget alloué par la ville de New York aux espaces verts*

B) Le DiCo ne contient pas encore l'article BUDGET mais on y trouve ENVELOPPEIII, son quasi-synonyme. En effet, les FL synonymiques de cette dernière sont les suivantes :

{=} enveloppe budgétaire

{QSyn} budget

Selon le DiCo, la FP de ENVELOPPEIII (dont l'étiquette sémantique est 'somme d'argent') est :

~ DE valeur X [DONNÉE PAR LA personne Y] À LA personne Z POUR FAIRE W

Parmi les quatre positions actanciellles de ENVELOPPE, deux sont importants pour BUDGET dans le cadre de notre analyse : quelqu'un (Y) qui alloue un budget à quelqu'un (Z) d'autre.

la ville de New York occupe la position actancielle de Y et *les espaces verts* occupe celle de Z. Bien que les ancres ne soient pas des ASynS du défini, elles sont identifiables car le défini et les ancres sont des ASynS de l'occurrence de ALLOUER.

C) Type IIc.

[12.12]

A1) *la ville de New York*

B) Le défini est spécifié par un nom propre, les deux sont coréférentiels.

C) Type Va.

[12.13]

A1) (*le budget alloué par la ville de New York*) *aux espaces verts*

B) Bien que cela ne soit pas spécifié de façon explicite, on interprète ce défini comme faisant référence aux espaces verts de New York (et pas aux espaces verts en général). L'information nécessaire à l'interprétation est dans le texte mais elle demande un traitement plus complexe.

C) Type VIII.

[12.14]

A1) *le nombre des salariés*

B) La FP de NOMBREI.2 selon le DiCo3 est :

~ DE Y [AYANT LA valeur X]

les salariés, qui est un ASynS du défini, occupe la position de Y.

C) Type IIb.

[12.15]

A1) *les salariés chargés de l'entretien des jardins publics*

B) Le défini est spécifié à l'intérieur de son groupe nominal : il fait référence à tous les individus qui correspondent à la description.

C) Type Vb.

[12.16]

A1) *l'entretien des jardins publics*

B) Le NPR définit ENTRETIEN comme 'Soins, réparations, dépenses qu'exige le maintien (de qqch.) en bon état.' Cette définition suggère que la situation décrite par ENTRETIEN a un participant qui dénote ce qui est entretenu. Ce participant est facilement exprimable : *l'entretien d'une maison, l'entretien d'une machine*, etc., ce qui indique qu'il correspond à une position actancielle.

Le DiCo ne contient pas la lexie ENTRETIEN mais on y trouve MÉNAGEI, son quasi-synonyme, dont la FP contient également une telle position actancielle Y :

~ DANS L'endroit Y [FAIT PAR L'individu X]

les jardin public, qui est un ASynS du défini, occupe cette position.

C) Type IIb.

[12.17]

A1) *les jardins publics*

A2) *les espaces verts* [12.13]

B) L'article JARDIN PUBLIC du DiCo3 contient la FL suivante :

{Gener} _espace vert_

Autrement dit, l'ancre est un terme générique du défini.

C) Type Ic.

[12.18]

A1) *la ville*

A2) *la ville de New York* [12.12]

B) Le défini est anaphorique, son antécédent contient le même nominal.

C) Type Ia.

[12.19]

A1) *ces effectifs*

A2) le **nombre** des salariés [12.14]

B) Selon le NPR, le sens de EFFECTIF est ‘nombre de personnes (constituant un groupe défini)’. L’ancre (*le nombre des salariés*) est un synonyme du défini (*ces effectifs*), étant donné que l’expression *les salariés* désigne un groupe de personnes.

C) Type Ib.

[12.20]

A1) *l’herbe*

B) Le lecteur interprète *l’herbe* comme faisant référence à l’herbe de Central Park.

C) Type VIII.

[12.21]

A1) *les massifs de fleurs*

B) Le lecteur interprète *les massifs de fleurs* comme faisant référence aux massifs de fleurs de Central Park.

C) Type VIII.

[12.22]

A1) *les tempêtes de l’hiver*

B) Le lecteur interprète *les tempêtes de l’hiver* en fonction du seul contenu de l’expression.

C) Type Vb.

[12.23]

A1) *l’hiver*

B) Le défini doit être compris par rapport au moment de l’énonciation : comme faisant référence au dernier hiver qui précède l’énonciation.

C) Type VI.

[12.24]

A1) le sol

A2) **Central Park**

B) La FP de SOL1 dans le DiCo3 est :

~ DE L'endroit X

Central Park, dont ENDROIT est un hyperonyme, occupe la position de X.

C) Type IIc.

[12.25]

A1) les poubelles

B) Le lecteur interprète *les poubelles* comme faisant référence aux poubelles de Central Park.

C) Type VIII.

[12.26]

A1) les fontaines publiques

B) Le lecteur interprète *les fontaines publiques* comme faisant référence aux fontaines publiques de Central Park.

C) Type VIII.

[12.27]

A1) la porte-parole d'une association

B) Le NPR définit PORTE-PAROLE comme 'Personne qui prend la parole au nom de qqn d'autre, d'une assemblée, d'un groupe' et donne les exemples suivants : *Les porte-parole du personnel auprès de la direction* et *Porte-parole du gouvernement*.

La définition et les exemples indiquent que PORTE-PAROLE a une position actancielle pour l'entité dont quelqu'un est le/la porte-parole.

une association, qui est un ASynS du défini et dont ORGANISATION est un quasi-synonyme, occupe ici cette position actancielle.

C) Type IIb.

[12.28]

A1) *les parcs new-yorkais*

B) Le défini est spécifié à l'intérieur de son groupe nominal de façon non ambiguë.

C) Type Vb.

5.5 Évaluation des analyses

Notre évaluation des analyses consiste à récapituler et à examiner les résultats obtenus et à présenter quelques-unes des limites de ces analyses.

5.5.1 Résultats des analyses

Nous présentons d'abord des statistiques sur les résultats et ensuite nous les commentons type par type. À la fin de la sous-section, nous proposons une évaluation globale des analyses.

5.5.1.1 Statistiques

Dans douze textes où il est question de quatre lieux publics, nous avons analysé et classé 181 expressions définies selon le type d'identification de leur référent. Parmi les définis qui ont plusieurs ancrés, dix étaient classés dans deux types (par exemple, *le premier ministre... son chef de cabinet, Jean Pelletier* [4.6]), ce qui donne au total 191 identifications. Les résultats des analyses sont contenus dans le Tableau 1 :

Tableau 1

Type	Sous-type	Nombre de types trouvés	Nombre de sous-types trouvés
I		26	
	Ia		10
	Ib		9
	Ic		7
II		98	
	IIa		8
	IIb		62
	IIc		28
III		13	
IV		6	
V		11	
	Va		5
	Vb		6
VI		11	
VII		15	
	VIIa		7
	VIIb		8
VIII		11	

La première colonne du tableau contient les numéros de types, la deuxième les numéros de sous-types, la troisième le nombre de types trouvés, la quatrième le nombre de sous-types trouvés.

5.5.1.2 Commentaires type par type

Nous présentons ici des commentaires sur chaque type d'identification du défini.

Type I : identification par relation anaphorique.

26 identifications sur 191 relèvent du type I : 10 du sous-type Ia, 9 du sous-type Ib et 7 du sous-type Ic. Ce type fait clairement appel aux CL pour ses sous-types Ib (synonymes) et Ic (hyperonyme-hyponymes)³⁰. Il est intéressant de contraster nos résultats avec Bergler (1997) qui s'intéresse à la résolution de l'anaphore en utilisant des stratégies simples et sans l'utilisation d'un lexique. Dans cet article, ce que nous appelons les anaphores identiques (*equal to first reference* et *close to first reference*) représentent plus que la moitié des relations coréférentielles, alors que les synonymes ne représentent que 2% et les hyperonymes 4%. Selon les résultats de notre analyse, la proportion des synonymes et des hyperonymes est nettement plus importante : cette différence pourrait s'expliquer aussi bien par la différence des natures de textes (fait divers dans notre corpus, article de *Wall Street Journal* dans le corpus de Bergler) que par la différence entre les modèles linguistiques utilisés :

³⁰ Le sous-type Ia pourrait également faire appel aux CL, notamment lorsque le défini et son antécédent ont le même nominal principal et des modificateurs différents (*le Président américain* vs *le Président des Etats-Unis*).

l'utilisation d'un lexique riche, tel qu'il est postulé par la LEC, permet d'établir un nombre considérable de liens synonymiques et hyperonymiques.

Type II : identification par lien actanciel

98 identifications sur 191 relèvent du type II : 8 du sous-type IIa, 62 du sous-type IIb et 28 du sous-type IIc. Autrement dit, ce type totalise plus que la moitié des identifications et le sous-type IIb lui seul presque le tiers ! Cela indique clairement que les liens actanciels jouent un rôle particulièrement important dans l'interprétation du défini. Il est évident que c'est dans le cas du sous-type IIa que les liens actanciels sont les plus forts : l'ancre est un nom typique d'un actant du défini – toutefois, ce sous-type est relativement rare. La particularité du sous-type IIb est que l'ancre est à la fois un ASynS du défini (ce qui rend son identification pratiquement triviale) et un ASém. C'est cependant pour les définis du sous-type IIc qu'il est le plus indispensable d'établir les liens lexicaux. Ces définis ont une ancre mais elle est en dehors du groupe nominal du défini (*Lyon... le conseil municipal* [10.5]) ; il faut faire appel aux CL pour trouver ces liens.

Type III : identification par une FL autre que Syn, Gener et S_i

Treize identifications relèvent du type III. Bien que ces FL puissent être perçues comme représentant un lien lexical moins fort que **Syn**, **Gener** et **S_i**, elles peuvent jouer un rôle important dans l'identification du défini, notamment lorsque le défini

et l'ancre sont syntaxiquement indépendants l'un de l'autre (*gare de Montpellier... les voies* [1.7]).

Type IV : identification à l'intérieur de certaines expressions idiomatiques

Seulement six identifications relèvent du type IV. Il faut cependant souligner l'importance de CL ici face aux CM. Lorsque c'est l'expression idiomatique qui contrôle le choix de l'article (*brouiller les cartes* [7.16]), seules les CL peuvent fournir une explication de la présence du défini : les CM ne contiennent pas d'information à ce sujet.

Type V : identification par spécification

Onze identifications relèvent du type V : cinq du sous-type Va et six du sous-type Vb. Comme l'identification se fait à partir du contenu du groupe nominal du défini, l'importance des CL réside ici dans une bonne analyse de ce contenu.

Type VI : identification déictique

Onze identifications relèvent de ce type. Comme nous l'avons remarqué plus haut, ce type de défini fait appel avant tout aux CM, aux paramètres de l'énonciation.

Type VII : identification générique

Quinze identifications relèvent du type VII : sept du sous-type VIIa et huit du sous-type VIIb. Dans l'état actuel de nos connaissances, il est loin d'être clair ce qui permet d'identifier le défini comme appartenant à ce type ; une fois cette question mieux éclairée, nous pourrions évaluer la place des CL dans leur interprétation.

Type VIII : identification complexe

Onze définis font partie du type VIII. Le fait de parler d'identification complexe pourrait suggérer à certains lecteurs qu'il s'agit d'une catégorie fourre-tout ; toutefois, il n'en est rien. Tous les définis de ce type ont en commun le fait que le cotexte contient les informations nécessaires à leur interprétation mais cette interprétation demande un processus relativement complexe. Ce processus est complexe, d'une part, parce que le défini ne peut pas être lié à une lexie (ou une expression) particulière – comme c'est le cas des cinq premiers types – mais doit être lié plutôt à tout un ensemble de lexies et d'expressions ; d'autre part, parce que ces lexies et ces expressions sont, le plus souvent, syntaxiquement indépendantes du défini.

Après ces considérations spécifiques à chaque type, nous pouvons proposer quelques remarques générales concernant les analyses.

5.5.1.3 Évaluation générale des analyses

En premier lieu, nous tenons à souligner que dans le cas des cinq premiers types d'identification, qui totalisent 154 sur 191, on peut établir un lien linguistique (surtout sémantique, mais aussi parfois syntaxique) clair et net avec une ou quelques autres lexies ou expressions dans le texte qui joue un rôle central dans l'interprétation du défini. On peut même ajouter à cela les onze occurrences du type VIII où de tels liens existent aussi mais demandent un traitement relativement complexe. Autrement dit, dans l'écrasante majorité des cas, les liens sont présents.

On pourrait nous objecter que pour dire que, par exemple, l'expression *la gare d'Ottawa* dénote la gare d'Ottawa ou pour dire que *la ville de New York* dénote la ville de New York, on n'a pas besoin de savoir quel est le lien lexical entre GARE et OTTAWA ou entre VILLE et NEW YORK. Nous pouvons répondre à une telle objection que l'intérêt des liens lexicaux dans de tels cas devient crucial dès que l'on cherche à accomplir des tâches un peu plus complexes, par exemple, la résolution de l'anaphore. En effet, une analyse purement syntaxique ne verra pas de différence entre *la ville de New York* et *la gare d'Ottawa*, dans les deux cas, la structure syntaxique est pareille : un nom propre est rattaché à l'aide de la préposition DE à un groupe nominal, composé de l'article défini et d'un nom commun. Notamment, elle ne nous permet pas de dire que *la ville de New York* et *New York* sont des expressions coréférentielles alors que *la gare d'Ottawa* et *Ottawa* ne le sont pas – pour ce faire, nous avons besoin de connaissances lexicales.

Il est intéressant de comparer nos résultats avec ceux obtenus par Poesio et Vieira (1998), qui ont étudié plus de 1400 descriptions définies. Leur corpus était constitué de 33 articles choisis au hasard dans le *Wall Street Journal*. Selon leurs résultats, environ 30% des DD étaient anaphoriques, 18% associatives, et près de 50% étaient apparues pour la première fois demandant des connaissances spécifiques ou générales. Les auteurs considéraient comme anaphoriques seulement les DD qui avaient la même tête (le même nominal principal) que leur antécédent alors que les DD associatives incluaient aussi bien des synonymes que des associations entre le tout et ses parties. Les expressions idiomatiques ne représentaient qu'un pourcentage minime. Ces résultats sont très différents des nôtres : à part l'importance considérable des liens anaphoriques, les définis sont beaucoup moins en relation avec d'autres parties du texte que dans notre cas. Comme nous l'avons déjà mentionné dans le cas du défini anaphorique et l'étude de Bergler(1997), nous croyons que ce contraste est dû, au-delà de la différence qui existe entre les types de textes, à l'utilisation d'un lexique riche, tel qu'il est modélisé par la LEC.

Nous n'avons pas eu l'occasion d'écrire un programme qui accomplirait la tâche d'interprétation du défini mais nous estimons que l'approche de la LEC, utilisée dans ce travail, pourrait permettre la réalisation de telles applications. Les informations concernant l'ancrage des définis sont – la plupart du temps – faciles à trouver, même pour une machine. Les analyseurs actuels sont assez efficaces pour ce qui est d'établir la structure syntaxique des phrases et de reconnaître les ASynS. Les liens

lexicaux devraient également être relativement faciles à établir, une fois en possession d'un dictionnaire complet. Bien entendu, les stratégies de recherche données lors de la présentation de chaque type de défini ne sont qu'à l'état embryonnaire mais nous sommes convaincu qu'ils représentent des pistes valables pour l'interprétation.

Qui plus est, les informations linguistiques que l'on peut recueillir grâce à l'approche de la LEC devraient permettre d'accomplir – avec une bonne efficacité – des tâches très variées, telles la rédaction des résumés, la recherche d'information, les questions/réponses, etc. Concernant cette dernière tâche par exemple, il devrait être relativement facile de donner des réponses (36b), (37b) et (38b) aux questions (36a), (37a) et (38a), portant sur le premier texte analysé :

(36a) *De quelle gare est-il question ? (Où se trouve la gare ?)*

(36b) *De la gare de Montpellier. (Elle se trouve à Montpellier.)*

(37a) *De quelles voies est-il questions ? (Où sont ces voies ?)*

(37b) *Des voies de la gare de Montpellier. (Elles sont à la gare de Montpellier.)*

(38a) *Qui sont les protestataires ?*

(38b) *Les protestataires sont des infirmiers anesthésistes grévistes.*

Nos analyses nous suggèrent également une remarque concernant ce que nous pourrions appeler l'opposition approche globale vs approche locale. En IA, c'est une approche courante, et même triviale, que de chercher à accomplir une tâche, considérée comme relative à un domaine, à l'aide des connaissances sur ce domaine.

C'est notamment le cas des systèmes experts. Cette approche pourrait paraître prometteuse en TAL aussi : il est clair par exemple que pour étudier des textes qui parlent de la gare, ce sont des connaissances de la lexie GARE et des CM concernant la gare qui sont les plus utiles. Cependant, nos analyses suggèrent plutôt que les connaissances d'un domaine ne couvrent qu'une très petite partie des connaissances nécessaires à interpréter un texte où il est question de ce domaine. En effet, une grande partie des définis n'est pas directement liée au lieu étudié. Cela suggère que l'interprétation du défini doit être traitée de façon globale plutôt que de façon locale. Nous entendons par *façon locale* une approche qui consisterait à dire que chaque texte relève d'un certain domaine (par exemple, du champ sémantique de GARE) et que connaître ce domaine permettrait de comprendre une bonne partie du texte. L'approche globale – qui nous semble donc mieux adaptée – dirait au contraire que la compréhension d'un texte ne peut pas être faite en se limitant à un domaine, à un champ mais nécessite plutôt un dictionnaire complet de la langue générale.

5.5.2 Limites de l'analyse

Notre analyse porte sur un phénomène linguistique précis dans un corpus donné. Nous avons évalué la place des CL dans des textes (le plus souvent, des faits divers) sur Internet. Il reste à voir si, dans d'autres types de texte, elles occupent une place semblable. Il se pourrait que, dans des dialogues entre les personnes qui se connaissent bien et qui partagent un plus grand nombre de CM (amis ou collègues par exemple), le locuteur fasse davantage appel aux CM de son auditeur. Il se pourrait

également que dans certains types de textes, l'approche locale puisse s'avérer plus efficace.

La présence de l'article défini s'explique parfois par le fait qu'il s'agit d'une expression déictique, anaphorique ou générique. Bien entendu, cela nécessite un modèle qui décrit comment reconnaître de telles expressions. Il faut admettre que, pour le moment au moins, nous n'en disposons pas.

Établir un lien lexical ou, de façon plus générale, un lien sémantique entre deux unités lexicales peut aussi nous amener sur une mauvaise piste. Considérons les énoncés (39a) et (39b) :

(39a) *Sarah est allée en Angleterre. Elle est dans la capitale maintenant.*

(39b) *Sarah est revenue de l'Angleterre. Elle est dans la capitale maintenant.*

Dans le premier exemple, *la capitale* fait référence à la capitale de l'Angleterre alors que dans le second, à une capitale autre que celle d'Angleterre. Le simple fait d'établir une relation sémantique entre CAPITALE et ANGLETERRE ne permet pas de saisir cette différence ; il faut prendre en considération le sens de chacun des énoncés, notamment le sens des lexies ALLER et REVENIR.

Nous n'avons pas de modèle qui permettrait de lier les définis à des éléments d'une base de connaissances. Cela s'explique non seulement par l'absence d'une telle base

mais aussi par le fait que, dans bien des cas, les humains eux-mêmes en seraient incapables. On sait par exemple, que les protestataires dont il est question dans le premier texte sont les infirmiers anesthésistes grévistes. Cependant, avec les informations dont on dispose, on n'a aucun moyen de les lier à une base de connaissances qui contiendrait le nom de tous les infirmiers anesthésistes.

Certaines des limites mentionnées sont plus faciles à dépasser que d'autres. Toutefois, malgré toutes ces limites, les analyses constituent une preuve solide pour affirmer que les CL sont, potentiellement, quasi omniprésentes dans l'interprétation du défini.

6 Conclusion

Dans cette conclusion, nous proposons d'abord un aperçu du travail accompli et mentionnons ensuite quatre questions qui restent à étudier : elles pourraient être une suite logique de notre travail.

6.1 Aperçu du travail accompli

Le but de notre travail était d'étudier la question suivante : étant donné les informations contenues dans les textes, dans quelle mesure les machines dotées de connaissances linguistiques seraient-elles capables d'interpréter ces textes ? Plus particulièrement, dans quelle mesure seraient-elles capables d'identifier le référent des expressions définies ? Bien entendu, il faut comprendre *référent* au sens large. Identifier le référent de l'expression *le X* signifie pouvoir répondre à la question : de quel X parle-t-on ? Autrement dit, il ne s'agit pas tant de mettre ce X en relation avec une base de données ou une base de connaissances qui décrirait le monde mais de l'identifier de façon unique dans le contexte de l'énonciation.

À vrai dire, dans une optique TAL, il importe relativement peu de savoir comment les humains identifient le référent de tel ou tel défini : à l'aide de leurs CM ou à l'aide de leurs CL ? Ce qui compte, c'est de savoir dans quelle mesure une certaine

approche (dans le cas de ce travail, une approche lexicale) peut être suffisante pour identifier ces référents.

Nous avons commencé notre travail par établir une distinction entre les CL et les CM et par présenter des projets dont le but est de les recenser. Après cela, nous avons présenté l'approche de la LEC et expliqué pourquoi nous l'avons choisie pour évaluer l'importance des CL dans le processus d'interprétation. Nous avons examiné de près les deux notions centrales de notre travail : *énoncé* et *interprétation*. Nous nous sommes également penché sur ce que sont les expressions définies et nous avons fait des analyses sur les définis d'un corpus. Ces analyses indiquent que de nombreuses CL – notamment des liens actanciels mais aussi d'autres FL – sont disponibles pour l'interprétation de presque la totalité des expressions définies.

Ce travail n'est pas le premier à utiliser une approche lexicale pour proposer une alternative à l'utilisation des CM mais il est le premier à le faire dans le cadre de la LEC. Nous utilisons la LEC parce que nous y trouvons une puissance et une rigueur inégalée aussi bien en lexicologie qu'en lexicographie. Les CL, telles que modélisées par la LEC permettent notamment d'établir des liens précis entre les différentes lexies ou expressions des textes où elles se trouvent. L'utilisation des CL a plusieurs avantages par rapport à l'utilisation des CM. Un de ces avantages est qu'elles peuvent être définies selon des critères précis et représentées de manière uniforme.

Concernant la place des CL face aux CM, notre conclusion principale est que les textes et les CL contiennent une très grande quantité de connaissances nécessaires à l'interprétation des énoncés. Les analyses que nous avons faites des expressions définies, notamment celles du type VIII, suggèrent que le véritable défi de l'interprétation ne réside pas tant dans la possibilité d'utiliser des CM mais dans la possibilité de pouvoir mieux rendre compte des relations linguistiques (aussi bien lexicales que syntaxiques) qui existent entre les différentes parties du texte. Bien entendu, il ne s'agit pas de nier l'importance des CM mais de soutenir qu'une grande partie des connaissances nécessaires à l'interprétation est encodée lexicalement.

6.2 Quatre questions qui restent sans réponse

Une suite logique de notre travail pourrait être d'examiner les quatre questions suivantes : comment l'identification des lexies se fait-elle ? Comment lier le référent d'une expression définie à un élément d'une base de connaissances ? Comment modéliser l'interaction des CM et des CL dans le processus de la compréhension ? Quel rôle attribuer aux approches statistiques ?

Nous nous sommes intéressé à l'interprétation au sens plus restreint : établir les liens entre les référents des occurrences. Dans ce processus, nous avons pris pour acquis l'identification des lexies ; la désambiguïsation lexicale. Or, nous reconnaissons

qu'un modèle qui permettrait cette identification reste encore à trouver. Sur ce point, nous ne pouvons que partager la position de Fodor (2001) : on ne sait pas vraiment comment le cerveau fonctionne. Dans cet ouvrage, Fodor se penche sur la CTM (*Computational Theory of Mind*), une théorie qui postule que les opérations mentales sont des calculs, structurés et effectués de façon modulaire. La conclusion de l'auteur est que, bien que la CTM soit la meilleure théorie que nous ayons du fonctionnement du cerveau humain, elle ne peut couvrir qu'une petite partie de la vérité.

L'interprétation au sens plus large devrait également être étudiée : lier les référents à une base de connaissances. Bien entendu, cela nécessite entre autres la création préalable d'une énorme base de connaissances dont le contenu serait comparable aux connaissances d'un être humain. Il est clair que pour le moment nous n'avons pas de raison d'espérer qu'une telle base puisse être créée dans les années (voire dans les décennies) à venir.

Certes, l'interprétation, au sens plus restreint comme au sens plus large, est bien plus limitée que la compréhension. Toutefois, elle est une première étape incontournable vers d'autres tâches linguistiques, notamment la compréhension. On peut s'attendre à ce que dans le processus de la compréhension (incluant les raisonnements sur des énoncés, l'évaluation de leur vraisemblance, la prise de décisions, etc.) les CM jouent un rôle plus incontournable que dans celle de l'interprétation.

Il n'était pas possible d'examiner de près dans ce travail les approches statistiques, actuellement très utilisées en linguistique informatique, cf. Klavans & Resnik (1996). La raison principale en est la différence fondamentale qui existe entre ces projets et ceux qui ont été étudiés dans ce travail. Nous nous intéressons aux connaissances nécessaires à l'interprétation des énoncés (et aux projets qui les créent et qui les utilisent) alors que la linguistique informatique statistique s'y prend de façon diamétralement opposée. Cette différence peut être illustrée par une citation de Mitkov *et al.* (2001 : 473) qui s'intéressent particulièrement au problème de résolution d'anaphore :

« Much of the earlier work in anaphora resolution heavily exploited domain and linguistic knowledge (...) which was difficult both to represent and to process and which required considerable human input. However, the pressing need for the development of robust and inexpensive solutions to meet the demands of practical NLP systems encouraged many researchers to move away from extensive domain and linguistic knowledge and to embark instead upon knowledge poor anaphora resolution strategies. A number of proposals in the 1990s deliberately limited the extent to which they relied on domain and or linguistic knowledge and reported promising results in knowledge-poor operational environment. »

Malgré cette différence fondamentale dans l'utilisation ou la non-utilisation d'une grande quantité de connaissances, les approches statistiques pourraient représenter une piste intéressante aussi bien pour l'identification des lexies que pour les raisonnements sur les énoncés.

Index

- absence d'opposition**, 148
- actant sémantique**, 91
- anaphore hyperonymique**, 172
- anaphore identique**, 171
- anaphore synonymique**, 171
- ancre**, 167
- base**, 97
- collocatif**, 97
- collocation**, 97
- compréhension**, 108
- contexte**, 123
- cotexte**, 123
- définition lexicographique**, 69
- dérivation sémantique**, 94
- description définie**, 159
- encodée dans une langue (être)**, 73
- énoncé**, 107
- étiquette sémantique**, 103
- exprimable auprès de**, 90
- fonctions lexicales**, 71
- forme propositionnelle**, 67
- interprétation**, 117
- langue**, 106
- lexème**, 61
- lexicologie**, 62
- lexie**, 62
- lexique**, 62
- lexique génératif**, 44
- locution**, 61
- méronymie**, 16
- mot-clé**, 101
- objet riche**, 24
- opposition effective**, 147
- opposition neutralisée**, 148
- parole**, 106
- participant obligatoire**, 87
- pertinence**, 115
- posé**, 112

positions actancielles, 88

présupposé, 112

situation, 123

situation linguistique, 86

sous-entendu, 112

structure actancielle, 103

tableau de régime, 89

token, 107

troponyme, 56

type, 107

vocable, 62

Bibliographie

- ARISTOTE (1939) *Organon V – Les topiques*, Vrin, Paris.
- AUSTIN, John L. (1970) *How to do things with words*, Oxford University Press, N.Y.
- BARTLETT, Frederic (1932) *Remembering*, CUP, London.
- BEAUCHESNE, Jacques (2001) *Dictionnaire des cooccurrences*, Guérin, Montréal/Toronto.
- BERGLER, Sabine (1997) Towards Reliable Partial Anaphora Resolution, in: *Operational Factors in Practical, Robust Anaphora Resolution for Unrestricted Texts* Proceedings of a workshop sponsored by the Association for Computational Linguistics, Madrid, Spain.
- BRACHMAN, Ronald J. (1979) On the epistemological status of semantic networks, dans BRACHMAN & LEVESQUE (1985) *Readings in Knowledge Representation*, Morgan Kaufmann, Los Altos, CA.
- BROOKS, Rodney A. (1991) Intelligence without reason, Proceedings of IJCAI-91, pp. 569-595, Morgan Kaufman, San Mateo, CA.
- CARLSON, Lynn & NIRENBURG, Sergei (1992) *World Modeling for NLP*, Proceedings of the 3rd Conference on Applied Natural Language Processing, Trento, Italy.

- CHAFE, Wallace L. (1976) Givenness, Contrastiveness, Definiteness, Subjects, Topics, and Point of View in *Subject and Topic* edited by Charles N. Li, Academic Press, New York.
- COLE, Peter and MORGAN, Jerry (eds) (1975) *Syntax and semantics 3: Speech acts*. Academic Press, New York.
- COLLINS, Allan & QUILLIAN, M. Ross (1969) Retrieval time for semantic memory. *Journal of Verbal Learning & Verbal Behavior*, 8: 240-247.
- CORNULIER, Benoît de (1985) *Effets de sens*, Éditions de minuit, Paris.
- DALE, Robert, MELLISH, Chris & ZOCK, Michael, eds. (1990) *Current Research in Natural Language Generation*, Academic Press, London.
- DRUIDE INFORMATIQUE (2001) *Antidote MP*.
- DUCROT, Oswald (1991) *Dire et ne pas dire : principes de sémantique linguistique*, Hermann, Paris.
- DUCROT, Oswald & SCHAEFFER, Jean-Marie (1995) *Nouveau dictionnaire encyclopédique des sciences du langage*, Éditions du Seuil.
- ECO, Umberto (1988) *Sémiotique et philosophie du langage*, PUF, Paris.
- FARQUHAR, Adam, FIKES, Richard, RICE, James (1996): The Ontolingua Server: A Tool for Collaborative Ontology Construction.
<http://ksi.cpsc.ucalgary.ca/KAW/KAW96/farquhar/farquhar.html>.
- FELLBAUM, Christiane, ed. (1998) *Wordnet – An Electronic Lexical Database*, The MIT Press, Cambridge, Mass.
- FISSETTE, Jean (1990) *Introduction à la sémiotique de C. S. Peirce*, Collection Études et documents, Montréal

- FODOR, Jerry (2001) *The Mind Doesn't Work That Way – The scope and Limits of Computational Psychology*, The MIT Press, Cambridge, Massachusetts.
- FRADIN, Bernard (1996) *L'identification des unités lexicales*, dans Piotrowski et Stockinger(1996).
- FREGE, Gottlob (1892), Über Sinn und Bedeutung, dans *Zeitschrift für Philosophie und philosophische Kritik* (traduction française : C. IMBERT, Sens et dénotation dans *Écrits logiques et philosophiques*, Seuil, Paris).
- FUCHS, Catherine (1982) *La paraphrase*, Presses Universitaires de France, Paris.
- GOODMAN, Kenneth & NIRENBURG, Sergei, Eds. (1991) *THE KBMT Project: A case study in Knowledge-Based Machine Translation*, Morgan Kaufmann Publishers, San Mateo, California.
- GREVISSE, Maurice (1986) *Le bon usage*, Duculot, Paris.
- GRICE, Herbert Paul (1975) Logic and conversation, dans Cole and Morgan (1975: 41-58).
- HAIMAN, John (1980) Dictionnaires and encyclopedias, *Lingua* 50.
- HENRY, Françoise (1996) *Les paramètres de l'analyse dans la pratique lexicographique*, dans Piotrowski et Stockinger (1996).
- HOFSTADTER, Douglas R. (1995) *Fluid Concepts and Creative Analogies*, Basic Books, New York.
- KAMP, Hans (1981). A Theory of Truth and Semantic Representation in *Formal Methods in the Study of Language*, J. Groenedijk, J. Janssen, M. Stokhof, eds. Amsterdam: Mathematical Center Tracts.

- KATZ, Jerrold, FODOR, Jerry (1963) The structure of a semantic theory, *Language*, Vol. 39, Num. 2.
- KAYSER, Daniel (1997) *La représentation des connaissances*, Hermes, Paris.
- KLAVANS, Judith & RESNIK, Philip (eds) (1996) *The Balancing Act – Combining Symbolic and Statistical Approaches to Language*, The MIT Press, Cambridge, Massachusetts.
- LAGORGETTE, Dominique, LARRIVÉE, Pierre, dir. (2002) *Représentation du sens linguistique*, Lincom Europa, Munich.
- LAKOFF, George (1987) *Women, Fire and Dangerous Things*, The University of Chicago Press, Chicago and London.
- LAROUSSE (2001) *Le Petit Larousse*, version électronique.
- LEECH, Geoffrey (1974) *Semantics*, Penguin, Harmondsworth.
- LENAT, Douglas B. & GUHA Ramanathan V. (1990) *Building large knowledge-based systems*, Addison-Wesley Publishing Company, Inc., Reading, Massachusetts.
- LE PESANT, Denis (1996) Principes d'organisation des données lexicales dans un dictionnaire électronique, dans Piotrowski et Stockinger(1996).
- LINSKY, Leonard (1974) *Le problème de la référence*, Seuil, Paris.
- LUGER, F. George (2002) *Artificial Intelligence – Structures and Strategies for Complex Problem Solving*, Addison-Wesley, Reading, Massachusetts.
- MEL'CUK Igor *et al.* (1984, 1988, 1992, 1999) *Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques*, vol. I à IV, Les Presses de l'Université de Montréal, Montréal.

- MEL'ČUK, Igor (1993) *Cours de morphologie générale*, Presses de l'Université de Montréal, Montréal.
- MEL'ČUK, Igor (1997) *Vers une linguistique Sens-Texte*, Collège de France, Paris.
- MEL'ČUK, Igor (2001) *Communicative Organisation in Natural Language*, John Benjamins Publishing Company, Amsterdam/Philadelphia.
- MEL'ČUK, Igor (2003) *Actants*, MTT, Paris.
- MEL'ČUK, Igor (à paraître) Actants in Semantics and Syntax, *Language*.
- MEL'ČUK, Igor, CLAS, André et POLGUÈRE, Alain (1995) *Introduction à la lexicologie explicative et combinatoire*, Duculot, Louvain-la-Neuve.
- MEL'ČUK, Igor et POLGUÈRE, Alain, (à paraître) *Lexique actif du français*, Louvain-la-Neuve, Duculot.
- MINSKY, Marvin (1975) A framework for representing knowledge, dans BRACHMAN & LEVESQUE (1985) *Readings in Knowledge Representation*, Morgan Kaufmann, Los Altos, CA.
- MITCHELL, Melanie (1993) *Analogy-Making as Perception*, MIT Press, Cambridge, MA.
- MITKOV, Ruslan; BOGURAEV, Branimir et LAPPIN, Shalom (2001) Introduction to the Special Issue on Computational Anaphora Resolution, dans *Computational Linguistics*, V. 27, N. 4, December 2001.
- MURPHY, Lynne (2000) Knowledge of words versus knowledge about words: The conceptual basis of lexical relations, dans Peeters (2000).
- NIRENBURG & al. (1992) *Machine Translation: A Knowledge-Based Approach*, Morgan Kaufmann Publishers, San Mateo, California.

- ONYSHKEVYCH, Boyan & NIRENBURG, Serguei (1995) A Lexicon for Knowledge-Based MT, *Machine Translation*, 10:1-2.
- OSTERTAG, Gary, Ed. (1998) *Definite Descriptions – A Reader*, The MIT Press, Cambridge, Mass.
- PARIENTE, Jean-Claude (1985) *L'Analyse du langage à Port-Royal*, Éditions de Minuit, Paris.
- PEETERS, Bert (2000) Some recent milestones in the lexicon-encyclopedia debate dans PEETERS (ed) *The Lexicon-Encyclopedia Interface*, Elsevier, Amsterdam.
- PEETERS, Bert (2002) Les constructions du type *commencer un livre* : état de la question et nouvelles perspectives, dans LAGORGETTE et LARRIVÉE (2002).
- PEIRCE, Charles S. (1978) *Ecrits sur le signe*, rassemblés, traduits et commentés par Gérard Deledalle, Editions du Seuil, Paris.
- PIOTROWSKI, David et STOCKINGER, Peter, ed. (1996) Savoirs lexicaux et savoirs encyclopédiques – t.1. Légitimer la description lexicale, *Sémiotiques*, n.11.
- POESIO, Massimo et VIEIRA, Renata (1998) A Corpus Based Investigation of Definite Description Use dans *Computational Linguistics*, V. 24, N. 2, June 1998.
- POLGUÈRE, Alain (1998) La Théorie Sens-Texte dans *Dialangue*, vol. 8-9, Université du Québec à Chicoutimi.

- POLGUÈRE, Alain (2000) Towards a theoretically-motivated general public dictionary of semantic derivations and collocations for French, *Proceedings of EURALEX'2000*, Stuttgart.
- POLGUÈRE, Alain (2003) Étiquetage sémantique des lexies dans la base de données DiCo, TAL, Vol. 44 No 2.
- POLGUÈRE, Alain (à paraître) Lexical function standardness. In L. Wanner (ed.): *Festschrift in Honour of Igor Mel'čuk*, John Benjamins, Amsterdam/Philadelphia.
- PUSTEJOVSKY, James (1995) *The Generative Lexicon*, The MIT Press, Cambridge, Massachusetts.
- QUINE, Willard Van Ornam (1953) Reference and Modality dans *From a Logical Point of View*, Harvard University Press, Cambridge.
- QUINE, Willard Van Ornam (1960) *Word and Object*, MIT Press, Cambridge, Mass.
- REITER, Ehud (1990) *Generating Descriptions that Exploit a User's Domain Knowledge* in DALE et al.(1990).
- REITER, Ehud et DALE, Robert (2000) *Building Natural Language Generation Systems*, Cambridge University Press, Cambridge.
- REY-DEBOVE, Josette, REY, Alain, (dir) (1996) *Le Nouveau Petit Robert*, version électronique.
- RIEGEL, Martin, PELLAT, Jean-Christophe, RIOUL, René (1994) *Grammaire méthodique du français*, Presses Universitaires de France, Paris.
- RUSSELL, Bertrand (1905), On Denoting dans *Mind 14* (traduction française : De la dénotation dans *Écrits de logique philosophique*, PUF, Paris).

- SABAH, Gérard (1988) *L'intelligence Artificielle et le langage*, Hermès, Paris
- SANFORD, Anthony J. & GARROD, Simon C. (1981) *Understanding Written Language*, John Wiley & Sons, Chichester.
- SAUSSURE (de), Ferdinand (1975) *Cours de linguistique générale*, Hatier, Paris.
- SCHANK, Roger & ABELSON, Robert (1977) *Scripts Plans Goals and Understanding*, Hillsdale.
- SCHANK, Roger & REIGER, Chuck J. (1974) Inference and the computer understanding of the natural language, *Artificial Intelligence* 5(4) : 373-412.
- SOWA, John F. (1984) *Conceptual Structures: Information Processing in Mind and Machine*, Addison-Wesley, Reading, MA.
- SPERBER, Dan & WILSON, Roger (1986) *Relevance*, Basil Blackwell, Oxford.
- ST-GERMAIN, Jean (1995) *Incidence de la structure sémantique et communicative sur la structure syntaxique profonde des énoncés causatifs du français contemporain*, Ph. D., Université de Montréal.
- SWARTOUT, Bill, PATIL, Ramesh, KNIGHT, Kevin, RUSS, Tom (1996): Toward Distributed Use of Large-Scale Ontologies, *Proceedings of Tenth Knowledge Acquisition for Knowledge-Based Systems Workshop*,
http://ksi.cpsc.ucalgary.ca/KAW/KAW96/swartout/Banff_96_final_2.html.
- TAYLOR, John R. (2000) The network model and the two-level model in comparison in PEETERS (ed) *The Lexicon-Encyclopedia Interface*, Elsevier, Amsterdam.
- VERSPOOR, Cornelia, Maria (1997) *Contextually-Dependent Lexical Semantics*, Ph.D, University of Edinburgh.

WIERZBICKA, Anna (1972) *Semantics primitives*, Frankfurt a. M., Athenäum.

WIERZBICKA, Anna (1996) *Semantics – Primes and Universals*, OUP, Oxford.

Annexe : articles du DiCo

Nous présentons ici les quatre articles du DiCo qui portent sur les quatre lieux étudiés. Ces articles sont sans formatage (c'est-à-dire, en ASCII) dans le DiCo mais nous avons mis en gras les lexies qui ont un lien avec la lexie vedette (qui sont des valeurs d'une FL) afin d'en faciliter la lecture.

Premier article : GARE

Nom : GARE

cs : bâtiment public: ~ DU lieu Y [POUR LE train X]

cg : nom, fém

tr :

X = I = ---

Y = II = de N, A-poss, Adj | Adj = A0(N)

syn : {QSyn} **halte, station, terminus**

fl :

/*Nom pour X*/

{S1} **train; _chemin de fer_**

/*Utilisateur d'une G.*/

{S1Real²} **voyageur**

```

/*Nom pour Y*/
{S2} localité, village, ville
/*Individu responsable d'une G.*/
{Cap} _chef de ~_
/*Dans une G.*/
{Loc-in} à [la ~]; dans [ART ~]; en [~] | pour le train
X
{Qui sert aux trains de voyageurs} de voyageurs
{Qui sert aux trains de marchandises} de marchandises
{Qui sert aux trains de banlieue} de banlieue
/*[X] arriver dans une G.*/
{IncepReall} arriver, entrer [en ~]
/*[X] être dans une G.*/
{Reall} être [en ~]
/*[X] partir d'une G.*/
{FinReall} quitter [ART ~] //partir
/*G. est un lieu d'arrivée pour X*/
{Pour X IncepReall Fact1--usual} accueillir [N=X]
["Cette gare accueille les trains venant de Bretagne."]
{Local principal d'une G.} hall
{Local d'une G. où les voyageurs attendent les trains}
salle d'attente
{Endroit dans une G. où les voyageurs montent dans les
trains} quai
{Endroit dans une G. où roulent les trains} voie
{Endroit dans une G. où l'on peut 1) acheter des bil-
lets, 2) manger, 3) boire, 4) acheter de la lecture, 5)
laisser des bagages, 6) récupérer un objet perdu}
1) guichet#1; billetterie
2) buffet
3) buvette
4) kiosque à journaux
5) consigne
6) (bureau des) objets

```

trouvés

["Ta valise est peut-

être aux objets trouvés."]

{Employé d'une G. chargé de transporter les bagages}

porteur

{Mention sur un panneau dans une G. qui sert à désigner les trains

- 1) qui arrivent,
- 2) qui partent,
- 3) desservant la banlieue,
- 4) desservant la province ou l'étranger}

1) "Arrivées"

2) "Départs"

3) "Banlieue"

4) "Grandes

lignes"

ph :

chef de gare

gare fluviale

gare routière

gare maritime

ex :

Un homme est tombé du quai juste avant que le train n'entre en gare. Le conducteur a fait fonctionner sans arrêt le sifflet du train bien avant son entrée en gare afin d'avertir les voyageurs. Deux bombes ont été déposées sur des rails à proximité d'une gare, a annoncé la police.

Deuxième article : ÉCOLE

Nom : ÉCOLE

Numéro de lexie : I.1.b

cg : nom, fém

cs : établissement: ~ [DE personne X DE matière Y] POUR
personne Z

nb : [= S-loc--usual(enseignement#1a)]

tr :

X = I = ---

Y = II = ---

Z = III = de+N-pl, pour+N-pl, A-poss

Syn :

{Syn} "spéc" **établissement scolaire**

{QSyn} COURS#4; "fam" **_boîte à bac_**

{Gener} **établissement d'enseignement#1a**

fl :

{A0} d' [~] //scolaire

{Mult} **réseau d' [~s] | É. au pl //ÉCOLE#I.2**

{Cap} **directeur [de ART ~]; principal; proviseur**

{assistant de Cap} **sous-directeur; surveillant général;**

censeur; conseiller principal d'éducation; conseiller

d'orientation; intendant gestionnaire

{Equip} //corps(2)#I professoral, corps(2)#I ensei-
gnant#2

{Bon} **bonne** (prépos); **excellente; modèle**
 {Loc-in--actual} **à, dans** [ART ~]
 {Loc-in--usual} **à** [l'~]
 {en instituant les méthodes d'enseignement#1a et les traditions, CausFunc0} **fonder** [ART ~]
 {LiquFunc0} **fermer, supprimer** [ART ~]
 {Real1} **enseigner#1, travailler** [dans <à> ART ~]
 {Real1(É. élémentaire)} //faire l'ÉCOLE#I.1c, faire la **classe**
 {S1QReal1} **maître** [d'~] //enseignant#1; instituteur; **professeur**, "fam" prof
 {Real2} **s'enseigner** [à <dans> ART ~]
 {Real3} **aller** [à ART ~], **fréquenter** [ART ~], **faire ses études, se former** [à <dans> ART ~]
 {S0Real3} **fréquentation** [de ART ~], **formation** [à <dans> ART ~]
 {S0QReal3} //scolarité
 {S1Real3} **élève, écolier**
 {ancien S1Real3} [un] **ancien**
 {nouveau S1Real3} [un] **nouveau**
 {IncepReal3} **entrer** [à ART-déf ~ <dans ART-indéf ~>]; **prendre le chemin de** [l'~]
 {Perm1IncepReal3} **accueillir, admettre** [N à ART-déf ~ <dans ART-indéf ~>]
 {prématurément FinReal3} **abandonner, laisser, quitter** [ART ~]
 {CausReal3} **envoyer** [N à ART-déf ~ <dans ART-indéf ~>]
 {PerfCausReal3} **inscrire, "fam" mettre** [N à ART-déf ~ <dans ART-indéf ~>]
 {LiquReal3} **retirer** [N de ART ~]
 {Liqu1Real3} **renvoyer, expulser** [N de ART ~], "fam" **mettre** [N à la porte de ART ~]
 {PerfReal3} **passer** [par ART ~]
 {F1=ayant épuisé le programme et rempli(2) les exigences de l'É., PerfReal3} **finir, terminer** [l'~]

{document certifiant F1} **diplôme**
 {F2=peu de temps après F1} **_à la sortie_#1 [de ART ~] |**
 C3=Ø
 {S1F1} **finissant#2, diplômé [de ART ~]**
 {A1(F1 tout récemment)} **frais émoulu [de ART ~]**
 {S1ProxF1} **finissant#1**
 {Real3--actual} **//être en classe**
 {nonReal3--actual} **être absent [de l'~]**
 {Caus3nonReal3--actual} **s'absenter [de l'~]**
 {IncepFact0} **commencer, ouvrir**
 {FinFact0} **finir, fermer**
 {CausFact0} **ouvrir [ART ~]**
 {É. élémentaire} **la petite ~**
 {pouvoir l'ensemble de Z d'É. en les dotant d'un ensei-
 gnement#1a régulier} **//scolariser**
 {local de l'É. destiné à l'enseignement#1a} **//(salle de)**
classe
 {document périodique préparé par les professeurs
 (S1QReal1) et l'administration d'une É. et contenant les
 notes de travail et de conduite de l'élève (S1Real3)}
//bulletin <relevé> (de notes <scolaire>)
 {punition donnée à l'élève (S1Real3) qui s'est mal com-
 porté ou qui n'a pas fait son travail, et qui consiste à
 venir ou à rester à l'É. (Real3--actual) en dehors des
 heures de COURS(1)#1} **//retenu, "arg scol" colle**
 {à l'É. élémentaire, sorte d'auvent construit dans la
 cour de récréation où les élèves (S1Real3) peuvent
 s'abriter en cas de mauvais temps} **préau ([de ART ~])**
 {F3=près de la porte de sortie de l'É.} **à la sortie [de**
ART ~]
 {reprise des activités scolaires dans une É., après les
 grandes vacances (= S0(après les grandes vacances, de
 nouveau IncepFunc0))} **//rentrée (des classes <scolaire>)**

ex :

Les 900 000 enfants indigènes qui se trouvent aujourd'hui sans écoles auraient été scolarisés [A. Camus]. Il n'avait été en classe que quelques semaines en tout; il gardait cependant un souvenir persistant mais douloureux et qui ne lui laissait pas de repos [G. Roy]. Toutes les écoles auront quatre jours de congé à l'occasion des fêtes de Pâques. Les enfants apprennent plus par la télévision qu'à l'école.

Troisième article : MAIRIE

Nom : MAIRIE

cg : nom, fém

cs : bâtiment public/administration: ~ DE LA localité X1
DE L'ensemble de individu X2

tr : X = I = de N, A-poss

Syn : {QSyn} **_hôtel de ville_**

fl :

/*Nom pour X*/

{S2} **arrondissement, localité, municipalité, village, ville** | X1; **administré** | X2

/*Chef de la M.*/

{Cap} **maire**

/*Personnel de la M.*/

{Equip} **personnel [de ART ~]**

{SingEquip} **secrétaire [de ~] //maire-adjoint = adjoint au maire, _conseiller municipal_**

{QEquip} **conseil municipal, élus, cabinet du Maire**

/*Localisé dans la M.*/

{Loc-in} **à [la ~]**

{Place publique se trouvant devant une M.} **Place de la ~**

{Mariage célébré dans une M.} **mariage civil**

{Service d'une M. fonctionnant de façon continue} **permanence [de ART ~]**

{Services d'une M. s'occupant de l'emploi, du logement et de la santé} **services sociaux [de ART ~]**

{Services d'une M. s'occupant des formalités administratives} **services administratifs [de ART ~]**

ex :

Elle présente, dans le hall de la mairie, les photos réalisées en Roumanie par Yann Pouliquen. La cathédrale Saint-Cyr se trouve face à la mairie, à 200 mètres du palais ducal. Cette somme sera dispensée dans les écoles primaires de la ville, précise-t-on à la mairie.

L'Association des jardins familiaux a déposé, samedi matin en mairie, 90 lettres-pétitions destinées à accélérer le processus. Dans les dossiers déposés en mairie, les riverains du projet trouveront l'étude d'impact, sans oublier l'étude économique avec l'influence des échangeurs. La mairie a décidé de lâcher d'autres furets pour s'attaquer aux lapins qui campent aux abords de la ligne de chemin de fer. La mairie a dépensé environ 1 million de francs en frais d'études. Il n'a toujours pas décidé de l'éventualité de sa candidature à la mairie de Vienne.

Quatrième article : JARDIN PUBLIC

Nom : JARDIN PUBLIC

cg : nom, masc

cs : endroit aménagé: ~ [MAINTENU PAR L'individu X POUR L'individu Y]

tr : Y = II = ---

Syn : {QSyn} **parc; _jardin de ville_, _jardin d'agrément_; square**

fl :

{Gener} **_espace vert_**

/*Nom pour X*/

{S1} **jardinier; gardien**

/*Nom pour Y*/

{S2} **visiteur**

/*Localisé dans le J. P.*/

{Loc-in} **dans [ART ~]**

/*Créer un J. P.*/

{CausFunc0} **aménager [ART ~]**

/*[Y] utiliser un J. P.*/

{Real2} **se promener [dans ART ~]; fréquenter [ART ~]**

{Chemin bordé d'arbres, de massifs ou de verdure dans un J. P.} **allée**

{Groupe d'arbres dans un J. P.} **bosquet**

{Femme qui loue les chaises dans un J. P.} **chaisière**

{Partie d'un J. P. où l'on a aménagé des compartiments de fleurs ou de gazon} **parterre**

{Construction circulaire dans un J. P. ouverte de tous les côtés destinée à recevoir les musiciens d'un concert public en plein air} **kiosque (à musique)**

{Petit étang dans un J.P.} **_pièce d'eau_**

ex :

Driss Moussaoui, ce pauvre hère marocain de quarante-quatre ans, est venu dormir sur un banc de jardin public, après une journée de cette galère métropolitaine qui est son lot depuis près de dix ans. Michel Bécot, le jardinier en chef de la ville de Fougères, présente le trophée au milieu des massifs du jardin public. But de l'opération : aider la collectivité locale à aménager un jardin public. Kiosques à musique, « Un souvenir magique dans un jardin public »: c'est le kiosque à musique chanté par Isabelle Lecomte, chanteur-compositeur.