

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI[®]

**Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600**



Université d'Ottawa • University of Ottawa

Les dictionnaires dans les systèmes de traduction automatique

par Isabelle Guilbault

sous la direction de Madame Roda P. Roberts

Thèse présentée à

l'École des études supérieures et de la recherche

de l'Université d'Ottawa

pour l'obtention de la Maîtrise en Traduction

Université d'Ottawa

École de traduction et d'interprétation

© Isabelle Guilbault, Ottawa, Canada, 1998



**National Library
of Canada**

**Acquisitions and
Bibliographic Services**

**395 Wellington Street
Ottawa ON K1A 0N4
Canada**

**Bibliothèque nationale
du Canada**

**Acquisitions et
services bibliographiques**

**395, rue Wellington
Ottawa ON K1A 0N4
Canada**

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-45222-0

ABSTRACT

Machine translation (MT) is a relatively new field combining linguistics and computer science. From a linguistic perspective, one of the most important components of MT systems is undoubtedly dictionaries. Until now, few studies have been carried out on the role played by dictionaries in such systems. The objective of this thesis is therefore to advance research in machine lexicography, or more precisely, in bilingual machine lexicography.

This thesis examines MT dictionaries in a certain number of systems in order to gain a better understanding of the role played by MT bilingual dictionaries in relation to other parts of the system. This is important since, unlike human translators who can choose to consult bilingual dictionaries whenever they wish, according to their own particular needs, MT systems dictate when consultation takes place and the type of information sought. Furthermore, the type of MT system (direct, transfer or interlingual) has a direct impact on the content and use of bilingual dictionaries. Bilingual dictionaries in direct systems, which do not contain unilingual dictionaries, usually contain more information than those in transfer or interlingual systems. However, the linguistic analysis of bilingual dictionaries in direct systems is less sophisticated than that of unilingual dictionaries in transfer or interlingual systems.

This thesis also compares traditional bilingual dictionaries with MT bilingual dictionaries in order to determine to what extent it is possible to adapt a traditional bilingual dictionary entry for use in MT systems.

REMERCIEMENTS

Je tiens avant tout à remercier Madame Roda P. Roberts pour son encadrement, sa détermination et sa rigueur. Ses encouragements m'ont permis de tenir le coup et de terminer cette thèse dans les délais prévus. Sans elle, il m'aurait tout simplement été impossible de finir en si peu de temps.

En second lieu, je voudrais remercier l'École des études supérieures et de la recherche pour la bourse d'été qu'elle m'a accordée et qui a rendu la rédaction de cette thèse possible.

Il m'aurait été particulièrement difficile de réunir les informations nécessaires à l'élaboration de certains chapitres sans la collaboration de plusieurs consultants : Geert Adriaens, Winfield Scott Bennett, David Farwell, Brian Harris, Pierre Isabelle, Heinz-Dirk Luckhardt, Steve McLaughlin, Ingrid Meyer, Harold Somers, Allison Sutherland, Erwin Valentini et Frank Van Eynde.

J'aimerais aussi remercier quelques proches pour le soutien moral qu'ils m'ont apporté au cours de l'été, notamment mes parents, Michel et Claire, ma belle-mère Michelle, Pierre Seewaldt, Gisèle Béland et Valérie Petit.

Toute ma gratitude à Chantale Grenon-Nyenhuis pour son aide et ses encouragements, ainsi que pour la relecture de ma thèse, et à Kelly Ann Sullivan pour avoir traduit mon résumé.

Enfin, merci à mes petits frères Maxence et Renaud, ainsi qu'à Anya, qui m'ont permis de me changer les idées et de me détendre par leurs jeux et leur affection.

RÉSUMÉ

La traduction automatique est un domaine relativement récent qui allie linguistique et informatique. Du point de vue linguistique, les dictionnaires sont sans aucun doute une des composantes les plus importantes des systèmes de traduction automatique. Jusqu'à présent, peu d'études ont été menées sur le rôle des dictionnaires dans les systèmes de traduction automatique. Le but de cette thèse vise justement à faire avancer les recherches en lexicographie automatique, plus particulièrement en lexicographie bilingue.

Dans un premier temps, cette thèse cherche à examiner les dictionnaires automatiques dans un certain nombre de systèmes afin de mieux cerner le rôle des dictionnaires bilingues automatiques par rapport aux autres parties du système, car au contraire du traducteur humain qui consulte le dictionnaire bilingue à son gré pour répondre à ses besoins variés, le système de traduction automatique dicte le moment de consultation du dictionnaire et le type de renseignements recherchés. En outre, le type de système de traduction automatique (direct, par transfert ou interlangue) influence directement le contenu et l'utilisation du dictionnaire bilingue. Les dictionnaires bilingues des systèmes directs (qui ne possèdent pas de dictionnaires unilingues) contiennent généralement plus d'information que les dictionnaires bilingues des systèmes de transfert et interlangues. Toutefois, l'analyse linguistique des dictionnaires bilingues des systèmes directs est moins élaborée que celles des dictionnaires unilingues des systèmes de transfert et interlangues.

Enfin, dans un second temps, cette thèse effectue une comparaison entre les dictionnaires bilingues traditionnels et les dictionnaires bilingues automatiques afin de voir dans quelle mesure il est possible d'adapter une entrée de dictionnaire bilingue traditionnel pour un dictionnaire bilingue automatique.

TABLE DES MATIÈRES

INTRODUCTION	1
0.1 Présentation du sujet	1
0.2 Problématique	2
0.3 Survol de la documentation	2
0.3.1 <i>Besoins en dictionnaires automatiques</i>	3
0.3.2 <i>Dictionnaires imprimés et dictionnaires automatiques faits manuellement</i>	4
0.3.3 <i>Dictionnaires électroniques et dictionnaires automatiques</i>	5
0.3.4 <i>Conclusion</i>	7
0.4 Objectifs	7
0.5 Méthodologie	8
0.6 Plan de la thèse	8
0.7 Concepts de base	10
0.7.1 <i>Morphologie</i>	10
0.7.1.1 <u>Flexion</u>	10
0.7.1.2 <u>Dérivation</u>	10
0.7.1.3 <u>Composition</u>	11
0.7.2 <i>Lexique et sémantique</i>	11
0.7.2.1 <u>Lexique</u>	11
0.7.2.2 <u>Dictionnaire</u>	11
0.7.2.3 <u>Lemme</u>	11
0.7.2.4 <u>Entrées</u>	12
0.7.2.5 <u>Polysémie</u>	12
0.7.2.6 <u>Homographie</u>	12
0.7.2.7 <u>Propriétés sémantiques</u>	13
0.7.2.8 <u>Sous-catégorisation</u>	13
0.7.2.9 <u>Restrictions de sélection</u>	14
0.7.2.10 <u>Décalage</u>	14
0.7.3 <i>Grammaire</i>	15
0.7.3.1 <u>Grammaires formelles</u>	15
0.7.3.2 <u>Analyse syntagmatique</u>	16
0.7.3.3 <u>Analyse syntaxique</u>	16
0.7.3.4 <u>Grammaire de cas</u>	16
0.7.3.5 <u>Divergence</u>	16
0.7.4 <i>Langage et TA</i>	18
0.7.4.1 <u>Langage contrôlé</u>	18
0.7.4.2 <u>Sous-langage</u>	18
0.7.5 <i>Traduction et automatisation</i>	19
0.7.5.1 <u>Traduction humaine assistée par ordinateur (THAO)</u>	19
0.7.5.2 <u>Traduction automatique assistée par l'homme (TAAH)</u>	19

	0.7.5.3 <u>Traduction entièrement automatique (TEA)</u>	20
0.8	Conclusion	20
Chapitre 1 :	INFORMATION GÉNÉRALE SUR LES SYSTÈMES DE TA	21
1.0	Introduction	21
1.1	Différents types de systèmes	21
1.1.1	<i>Systèmes directs</i>	21
1.1.2	<i>Systèmes de transfert</i>	24
1.1.3	<i>Systèmes interlangues</i>	26
1.1.4	<i>Conclusion</i>	28
1.2	Trois phases dans les systèmes de transfert	29
1.2.1	<i>Analyse</i>	29
1.2.2	<i>Transfert</i>	30
1.2.3	<i>Génération</i>	31
1.3	Différents modes de traduction	32
1.3.1	<i>Traduction en mode différé</i>	32
1.3.2	<i>Traduction interactive</i>	33
1.4	Conclusion	34
Chapitre 2 :	ÉTUDE DES DICTIONNAIRES UNILINGUES ET BILINGUES DANS QUELQUES SYSTÈMES DE TA	35
2.0	Introduction	35
2.1	Analyse des dictionnaires dans quelques systèmes de TA	35
2.1.1	<i>Systèmes directs</i>	35
2.1.1.1	<u>Globalink</u>	36
2.1.1.2	<u>Systran</u>	39
2.1.1.3	<u>Météo</u>	42
2.1.2	<i>Systèmes de transfert</i>	45
2.1.2.1	<u>METAL</u>	45
2.1.2.2	<u>Logos</u>	49
2.1.2.3	<u>Tovna</u>	51
2.1.2.4	<u>SUSY</u>	54
2.1.2.5	<u>Ariane (GETA)</u>	56
2.1.3	<i>Systèmes interlangues</i>	58
2.1.3.1	<u>Rosetta</u>	59
2.1.3.2	<u>DLT</u>	62
2.2	Rôle des dictionnaires dans les systèmes de TA	65
2.2.1	<i>Rôle des dictionnaires unilingues</i>	66
2.2.2	<i>Rôle des dictionnaires bilingues</i>	68
2.3	Conclusion	70
Chapitre 3 :	LES ENTRÉES DE DICTIONNAIRE BILINGUE AUTOMATIQUE	71
3.0	Introduction	71

3.1	Méthodologie	71
3.1.1	<i>Texte pour Globalink et Logos</i>	72
3.2	Globalink	73
3.2.1	<i>Traduction du texte en LD avec les dictionnaires du système</i>	73
3.2.2	<i>Entrées à examiner</i>	73
3.2.2.1	<u>Canadian</u>	74
3.2.2.2	<u>potential</u>	75
3.2.2.3	<u>interest</u>	76
3.2.2.4	<u>performance</u>	76
3.2.2.5	<u>labor</u>	77
3.2.3	<i>Entrées modifiées ou préparées</i>	78
3.2.3.1	<u>interest rate</u>	78
3.2.3.2	<u>performance</u>	79
3.2.3.3	<u>labour market</u>	79
3.2.4	<i>Traduction du texte en LD avec les modifications apportées au dictionnaire</i>	82
3.2.5	<i>Conclusions sur le dictionnaire de Globalink</i>	82
3.3	Logos	83
3.3.1	<i>Traduction du texte en LD avec les dictionnaires du système</i>	84
3.3.2	<i>Entrées préparées</i>	84
3.3.2.1	<u>excess capacity</u>	85
3.3.2.2	<u>cost pressures</u>	87
3.3.3	<i>Traduction du texte en LD avec le dictionnaire d'utilisateur</i>	88
3.3.4	<i>Conclusions sur le dictionnaire de Logos</i>	89
3.4	Conclusion	89

Chapitre 4 :	COMPARAISON ENTRE LES DICTIONNAIRES BILINGUES TRADITIONNELS ET LES DICTIONNAIRES BILINGUES AUTOMATIQUES	91
4.0	Introduction	91
4.1	Constituants des dictionnaires bilingues traditionnels	91
4.2	Constituants des dictionnaires bilingues automatiques	95
4.3	Expérience : adaptation des constituants des dictionnaires bilingues traditionnels pour les dictionnaires bilingues automatiques	101
4.3.1	<i>grocery (division sémantique 1)</i>	102
4.3.1.1	<u>Globalink</u>	103
4.3.1.2	<u>Logos</u>	104
4.3.2	<i>groceries (division sémantique 2)</i>	105
4.3.2.1	<u>Globalink</u>	106
4.3.2.2	<u>Logos</u>	107
4.3.3	<i>to do the groceries</i>	107
4.3.3.1	<u>Globalink</u>	108
4.3.3.2	<u>Logos</u>	109

4.3.4	<i>grocery bill</i>	110
4.3.4.1	<u>Globalink</u>	110
4.3.4.2	<u>Logos</u>	111
4.3.5	<i>grocery store</i>	112
4.3.5.1	<u>Globalink</u>	112
4.3.5.2	<u>Logos</u>	113
4.3.6	<i>Conclusion sur l'expérience</i>	114
4.4	<i>Conclusion</i>	117
CONCLUSION		119
ANNEXE 1		127
BIBLIOGRAPHIE		129
INDEX		141
INDEX DES AUTEURS ET DES CONSULTANTS CITÉS		146

LISTE DES FIGURES

Figure 1	— Système direct	23
Figure 2	— Modules des systèmes de transfert	24
Figure 3	— Système de transfert	26
Figure 4	— Modules des systèmes interlangues	27
Figure 5	— Système interlangue	28
Figure 6	— Processus de traduction de Systran	42
Figure 7	— Processus de traduction de Metal	48
Figure 8	— Processus de traduction dans SUSY	56
Figure 9	— Configuration du système Ariane	58
Figure 10	— Processus de traduction de Rosetta	61
Figure 11	— Processus de traduction de DLT	65
Figure 12	— Entrée <i>Canadian</i> dans Globalink	75
Figure 13	— Entrée <i>potential</i> dans Globalink	75
Figure 14	— Entrée <i>interest</i> dans Globalink	76
Figure 15	— Entrée <i>performance</i> dans Globalink	77
Figure 16	— Entrée <i>labor</i> dans Globalink	77
Figure 17	— Entrée modifiée <i>interest</i> dans Globalink	78
Figure 18	— Entrée modifiée <i>performance</i> dans Globalink	79
Figure 19	— Sélection du dictionnaire dans Globalink	79
Figure 20	— Information sur les flexions dans Globalink	80

Figure 21 — Sélection des attributs dans Globalink	80
Figure 22 — Entrée préparée <i>labour</i> dans Globalink	81
Figure 23 — Entrée préparée <i>excess capacity</i> dans Logos	85
Figure 24 — Sélection de la catégorie pour <i>excess capacity</i> dans Logos	86
Figure 25 — Information sur <i>capacité</i> dans Logos	87
Figure 26 — Entrée préparée <i>cost pressure</i> dans Logos	88
Figure 27 — Entrée d'un dictionnaire bilingue automatique : 1 ^{re} possibilité	96
Figure 28 — Entrée d'un dictionnaire bilingue automatique : 2 ^e possibilité	97
Figure 29 — Entrée d'un dictionnaire bilingue automatique : 3 ^e possibilité	97
Figure 30 — Modification de l'entrée <i>grocery</i> dans Globalink	104
Figure 31 — Préparation de l'entrée <i>grocery</i> dans Logos	104
Figure 32 — Sélection de la catégorie pour <i>grocery</i> dans Logos et boîte de dialogue	105
Figure 33 — Modification de l'entrée <i>groceries</i> dans Globalink	106
Figure 34 — Préparation de l'entrée <i>groceries</i> dans Logos	107
Figure 35 — Ajout de la locution <i>to do the groceries</i> dans l'entrée <i>groceries</i> de Globalink ..	108
Figure 36 — Préparation de l'entrée <i>to do the groceries</i> dans Logos	109
Figure 37 — Information sur <i>faire</i> dans Logos	110
Figure 38 — Ajout de la locution <i>grocery bill</i> dans l'entrée <i>grocery</i> de Globalink	111
Figure 39 — Préparation de l'entrée <i>grocery bill</i> dans Logos	111
Figure 40 — Information sur <i>épicerie</i> dans Logos	112
Figure 41 — Modification de la locution <i>grocery store</i> dans l'entrée <i>grocery</i> de Globalink ..	113
Figure 42 — Préparation de l'entrée <i>grocery store</i> dans Logos	114

INTRODUCTION

0.1 Présentation du sujet

La traduction automatique (TA) est un domaine relativement récent dans l'histoire de la traduction, qui allie linguistique et informatique. C'est justement le grand progrès technologique du XX^e siècle qui fait que son évolution a été si rapide.

Du point de vue linguistique, une des composantes les plus importantes de la plupart des systèmes de TA¹ est sans aucun doute les dictionnaires. Selon Arnold *et al.* :

- Dictionaries are the largest components of an MT system in terms of the amount of information they hold. If they are more than [sic] simple word lists (and they should be, if a system is to perform well), then they may well be the most expensive components to construct.
- More than any other component, the size and quality of the dictionary limits the scope and coverage of a system, and the quality of translation that can be expected.
- The dictionaries are where the end user can expect to be able to contribute most to a system — in fact, an end user can expect to have to make some additions to system dictionaries to make a system really useful. While MT suppliers rarely make it possible for users to modify other components, they normally expect them to make additions to the dictionary. Thus, from the point of view of a user, a basic understanding of dictionary construction and sensitivity to the issues involved in 'describing words' is an important asset (Arnold *et al.* 1994 : 87).

Puisque nous travaillons depuis deux ans comme assistante de recherche au projet de Dictionnaire canadien bilingue (DCB), la lexicographie, surtout bilingue, est, en quelque sorte, notre domaine de spécialisation. Étant donné notre grand intérêt pour la TA, cette thèse portera sur les dictionnaires, principalement les dictionnaires bilingues, en TA.

¹ On commence à développer certains systèmes (ex. systèmes basés sur l'exemple) qui ne contiennent pas de dictionnaires.

0.2 Problématique

Les dictionnaires bilingues automatiques sont différents des dictionnaires bilingues traditionnels (imprimés ou sur CD-ROM), car les premiers font partie intégrante d'un ensemble. Au contraire du traducteur humain, qui consulte le dictionnaire bilingue à son gré pour répondre à ses besoins variés, le système de TA dicte le moment de consultation du dictionnaire et le type de renseignements recherchés. En outre, le type de système de TA influence le contenu et l'utilisation du dictionnaire bilingue, car dans certains systèmes, il existe également des dictionnaires unilingues.

Étant donné la place grandissante que prend la TA dans la société d'aujourd'hui, il semble important de faire une comparaison entre les dictionnaires bilingues traditionnels et les dictionnaires bilingues automatiques de certains systèmes de TA afin d'établir les différences essentielles et de déterminer ce qu'il faut pour adapter un dictionnaire bilingue traditionnel pour pouvoir l'utiliser dans un système de TA au lieu de créer un dictionnaire automatique de toutes pièces.

0.3 Survol de la documentation

En effet, un survol de la documentation sur la TA montre clairement que les besoins en dictionnaires automatiques sont grands et qu'il y a plusieurs tentatives d'exploiter des dictionnaires traditionnels pour combler ces besoins.

0.3.1 Besoins en dictionnaires automatiques

En fait, il y a un consensus général disant que le développement des dictionnaires et des autres bases de données est essentiel pour la réussite dans le domaine du traitement automatique des langues naturelles (TALN) (Nirenburg 1994, Klavans 1990, Atkins *et al.* 1994, Klavans 1996 et Hongyan et McKeown 1998). Les systèmes actuels de TALN, qui fonctionnent principalement en mode de démonstration, possèdent souvent des grammaires de grande taille, mais rarement de grands dictionnaires contenant de l'information sur le sens (Nirenburg 1994 : 314). Selon Atkins *et al.* (1994 : 21), les systèmes de TALN se heurtent à ce que Byrd (1989) a appelé le 'lexical bottleneck', c'est-à-dire des limites dans le rendement des systèmes de TALN à cause du manque de dictionnaires de taille suffisante. Selon Klavans (1996 : 2), il faut non seulement de grands dictionnaires, mais aussi des dictionnaires qui fournissent une grande variété d'informations : syntaxique, sémantique, pragmatique, morphologique et phonologique.

Même s'il existe un consensus général sur les besoins en dictionnaires automatiques, il y a certains problèmes à construire de tels dictionnaires. Par exemple, il peut y avoir des problèmes dans l'architecture du dictionnaire, puis des problèmes à lui fournir l'information nécessaire et suffisante (Klavans 1996).

It has become of paramount importance to find sources of data and to develop extraction and analysis techniques that allow the building of effective large lexicons with the minimum of effort and expense (Atkins *et al.* 1994 : 21).

0.3.2 Dictionnaires imprimés et dictionnaires automatiques faits manuellement

Comme l'indiquent Calzolari et Picchi (1994), la création d'un dictionnaire automatique peut se faire de deux façons différentes : d'abord par l'entrée des données par un lexicographe, ensuite par l'acquisition des données d'une base de données existante (ex. dictionnaire électronique).

L'entrée des données par un lexicographe est souvent nécessaire parce qu'un grand nombre de sources d'information lexicale ne sont pas disponibles sous format électronique (Klavans 1990 : 268). Même dans les années 1990, la création manuelle des dictionnaires est encore un phénomène courant, comme l'indiquent Grishman *et al.* (1994 : 269) en parlant de Comlex Syntax, un lexique computationnel donnant de l'information syntaxique détaillée pour environ 38 000 mots-vedettes anglais. «Our basic approach has been to create an initial lexicon manually and then to use a variety of resources, both commercial and corpus-derived, to refine this lexicon» (1994 : 269).

La création manuelle des entrées est souvent fondée sur l'information contenue dans des dictionnaires imprimés, comme le signalent Calzolari et Picchi (1994) dans le cadre de leur base de données «Italian LDB».

Selon Hongyan et McKeown (1998), «Most generation systems are still supported by a small system lexicon, with limited entries and hand-coded knowledge» (1998 : 607). Mais, comme ils l'indiquent, l'entrée manuelle des entrées demande beaucoup de temps et d'efforts, et il risque d'y avoir des erreurs.

Le but ultime est d'extraire automatiquement le maximum d'information des sources électroniques (Klavans 1990).

0.3.3 Dictionnaires électroniques et dictionnaires automatiques

Comme le signalent Atkins *et al.* (1994 : 27), «The advent of the machine-readable dictionary added a new and powerful dimension to lexical research. Computational linguists found themselves in possession of a resource which allowed them to focus as never before upon the lexicon» (voir aussi Nirenburg 1994).

En effet, un nombre grandissant de projets tentent d'exploiter l'information dans des sources électroniques. Parmi ceux-ci, Klavans (1990 et 1996) cite Michiels (1982), Alshawi (1985), Byrd *et al.* (1987), Fox *et al.* (1988), Calzolari (1983), Calzolari et Picchi (1988), Wilks (1988), Boguraev *et al.* (1989), Boguraev (1991) et Klavans, Chodorow et Wacholder (1990).

Atkins *et al.* (1994 : 26) donnent la liste des dictionnaires électroniques unilingues anglais et bilingues qui sont le plus souvent exploités. On y trouve aussi, par exemple, les dictionnaires *Webster's Seventh* et *Ninth New Collegiate*, (Gove 1969; Mish 1986), le *Collins English Dictionary* (Hanks 1986) et *The Oxford English Dictionary* 2^e édition (1989); parmi les dictionnaires pour apprenants, le *OALD* (Hornby 1974), le *LDOCE* (Procter *et al.* 1978) et le *Cobuild* (Sinclair *et al.* 1987) figurent sur la liste; enfin, des dictionnaires bilingues d'éditeurs tels que le *British Collins*, le *Dutch Van Dale* et autres semblent aussi populaires.

Malgré l'effort considérable consacré à l'exploitation de l'information contenue dans ces

dictionnaires électroniques, Nirenburg (1994 : 314) ne croit pas que l'utilisation directe des dictionnaires électroniques pour le TALN soit une option viable, parce que l'identification systématique de l'information ainsi que son extraction représentent un travail difficile et complexe (voir aussi Meijs 1992). En fait, encore peu de résultats pratiques ont été engendrés par les tentatives entreprises pour extraire les sens des dictionnaires électroniques et pour les représenter dans un format utilisable dans des programmes d'application (Nirenburg 1994 : 314). C'est en grande partie parce que les dictionnaires existants n'indiquent pas clairement comment les différents sens présentés dans une entrée pour un mot donné sont reliés et comment le mot lui-même s'insère dans la structure de la langue (Atkins et Levin 1991 : 235).

Afin de pallier les lacunes des dictionnaires individuels, les recherches portent vers la combinaison automatique des dictionnaires pour créer une base de données lexicales plus complète. Toutefois, ce n'est pas une tâche facile non plus, parce qu'il est rare que deux dictionnaires présentent des sens identiques pour le même mot. Il n'y a donc pas de garantie que le résultat de la combinaison de plusieurs dictionnaires serait nettement supérieur à un seul dictionnaire (Atkins et Levin 1991 : 235).

Bien que l'information contenue dans les dictionnaires électroniques soit déjà utilisée dans certaines applications en langues naturelles, comme la TA, et bien qu'on puisse espérer de meilleurs résultats de l'exploitation de tels dictionnaires dans l'avenir, Nirenburg (1994 : 315) signale clairement les limites des dictionnaires électroniques dans la production des dictionnaires automatiques. «[...] it does not seem realistic to expect all the problems of lexicon acquisition for

computational applications to be solved through research in transforming MRDs serving humans into machine-tractable dictionaries (MTDs in Wilks's terminology) serving computer programs».

En effet, pour compenser les lacunes des dictionnaires électroniques, qui, selon Matsumoto (1998 : 195), ne reflètent pas suffisamment les usages différents des mots selon les domaines et selon le but de la communication, on commence maintenant à utiliser l'information tirée des corpus pour la création de dictionnaires automatiques (voir aussi Klavans 1990 et Klavans 1996). Ainsi, par exemple, le système BICORD, dont les résultats peuvent être utilisés directement par les systèmes de transfert (en TA), combine les entrées du dictionnaire Collins anglais-français/français-anglais avec un grand corpus bilingue anglais-français (Klavans 1996).

0.3.4 Conclusion

En somme, la réutilisation de l'information contenue dans des dictionnaires est devenue une question importante en lexicographie computationnelle. Comme le souligne Sanfilippo (1994 : 273), «While this trend was essentially set by the need to maximize cost-effectiveness in building large scale Lexical Knowledge Bases for NLP (LKBs), there is a clear sense in which the construction of such knowledge bases also caters to the demand for better dictionaries».

0.4 Objectifs

Notre thèse, dont le but pratique est également la réutilisation de l'information contenue dans un dictionnaire traditionnel, vise trois objectifs :

(a) examiner les dictionnaires dans un certain nombre de systèmes de TA;

- (b) examiner le rôle des dictionnaires bilingues en TA par rapport à d'autres parties des systèmes (y compris les dictionnaires unilingues);
- (c) faire une comparaison entre les dictionnaires bilingues traditionnels et les dictionnaires bilingues automatiques.

0.5 Méthodologie

La méthodologie utilisée pour répondre aux objectifs de la thèse comprend :

- (a) une analyse de la documentation sur les systèmes de TA et leurs dictionnaires afin d'avoir un aperçu de ce qui a été accompli dans le domaine de la TA, de l'étendue des progrès depuis un certain nombre d'années et des nouvelles tendances;
- (b) une expérience faite avec les dictionnaires bilingues de deux systèmes de TA, qui consiste à adapter des constituants des dictionnaires bilingues traditionnels pour les dictionnaires bilingues automatiques. Cette expérience, qui est surtout effectuée d'un point de vue lexicographique, vise principalement à établir quel genre d'information les systèmes permettent à l'utilisateur d'entrer ou de modifier dans les dictionnaires et de quelle façon les résultats de la traduction en sont affectés.

0.6 Plan de la thèse

Cette thèse est divisée en six chapitres.

Le chapitre 1 donne de l'information générale sur les systèmes de TA. Cette information comprend une présentation des différents types de systèmes de TA, une présentation des trois

phases par lesquelles beaucoup de systèmes passent lors de la traduction d'un texte et une présentation des deux modes de traduction possibles dans les systèmes de TA actuellement sur le marché.

Le chapitre 2 étudie les dictionnaires unilingues et bilingues dans quelques systèmes de TA, ainsi que le rôle général des dictionnaires unilingues et bilingues dans la TA. Cette étude porte d'abord sur les composantes et le contenu des dictionnaires des systèmes de TA les plus connus ainsi que sur leur fonctionnement à l'intérieur de ces mêmes systèmes lors du processus de traduction, puis montre le rôle des dictionnaires unilingues et des dictionnaires bilingues par rapport aux autres parties d'un système de TA.

Le chapitre 3 a comme objectif l'examen d'entrées de dictionnaires bilingues dans deux systèmes de TA différents. Ce chapitre illustre le type d'information lexicographique requise par le système et la façon dont cette information est entrée dans le dictionnaire. Certaines entrées sont donc modifiées ou préparées. De plus, on y trouve une comparaison des traductions obtenues à l'aide de ces entrées modifiées ou préparées avec celles qui sont obtenues sans ces entrées supplémentaires.

Le chapitre 4 établit une comparaison entre les dictionnaires bilingues traditionnels et les dictionnaires bilingues automatiques. Après avoir présenté les constituants d'un dictionnaire bilingue traditionnel et les constituants d'un dictionnaire bilingue automatique, nous allons présenter l'expérience dont il est question dans la méthodologie. Cette expérience consiste à

adapter des constituants des dictionnaires bilingues traditionnels pour les dictionnaires bilingues automatiques de deux systèmes.

Enfin, le dernier chapitre discute des conclusions tirées de l'étude des dictionnaires bilingues dans les systèmes de TA.

0.7 Concepts de base

Cette thèse est fondée sur plusieurs concepts de base tirés de la linguistique et de la traduction. Ils sont présentés brièvement ci-dessous, groupés selon leur affiliation.

0.7.1 Morphologie

La morphologie traite de la structure interne des mots et de leur formation. Elle comporte trois procédés importants : la flexion, la dérivation et la composition.

0.7.1.1 Flexion

La flexion consiste à ajouter des terminaisons à une racine sans en changer le sens. Le mot ainsi formé garde la même catégorie grammaticale (ex. *love* (verbe) → *loved* (verbe)).

0.7.1.2 Dérivation

La dérivation consiste à créer de nouveaux mots à partir de racines existantes en y ajoutant des affixes. Le mot ainsi formé change, le plus souvent, de catégorie grammaticale (ex. *grammatical* (adj) → *grammatically* (adv)).

0.7.1.3 Composition

La composition consiste à associer des mots pour former de nouveaux mots (ex. *casser + tête* → *casse-tête*).

0.7.2 *Lexique et sémantique*

0.7.2.1 Lexique

Le lexique d'une langue comprend la liste de toutes les unités lexicales de cette langue.

0.7.2.2 Dictionnaire

Le dictionnaire, qu'il soit pour utilisation humaine ou pour la TA, est un ouvrage enregistrant une certaine description du lexique d'une langue ou de plusieurs langues mises en parallèle.

0.7.2.3 Lemme

Le lemme est la forme canonique (ex. le masculin singulier pour un nom ou un adjectif en français, l'infinitif pour un verbe) d'un mot variable. En lexicographie, et principalement en lexicographie automatique, on distingue généralement deux étapes dans la lemmatisation :

1. le regroupement des formes fléchies sous la forme type leur servant d'adresse lexicale².
2. la séparation des formes servant d'adresses lexicales quand elles sont homographes (ex. *voile.nm* et *voile.nf*).

² L'adresse lexicale est le mot mis en vedette au début de chaque article d'un dictionnaire.

0.7.2.4 Entrées

Les entrées, qui sont de forme et de contenu variés, dépendent du type de dictionnaires (unilingue vs bilingue, traditionnel vs automatique, etc.). Les entrées des dictionnaires unilingues traditionnels indiquent la prononciation, la catégorie grammaticale et les formes fléchies des lemmes et donnent les définitions de leurs sens ainsi que l'information sur leur usage. Pour leur part, les entrées des dictionnaires bilingues traditionnels fournissent, outre certains des renseignements signalés ci-dessus, des équivalents de ces lemmes. Les entrées des dictionnaires bilingues automatiques, quant à elles, indiquent toutes la catégorie grammaticale des lemmes et donnent des équivalents. Toutefois, elles ne donnent pas de définitions, si brèves soit-elles, ni d'exemples d'utilisation (contrairement aux entrées des dictionnaires traditionnels), même si certains dictionnaires bilingues automatiques contiennent d'autres éléments (par exemple, les propriétés sémantiques).

0.7.2.5 Polysémie

Il est question de polysémie lorsqu'une unité lexicale a plusieurs sens. Par exemple : *chaton* (nom masculin) a le sens d'«inflorescence en épi souple. *Chatons de coudrier, de noyer, de saule*» (Le Nouveau Petit Robert 1994 : 353) et celui de «petits amas de poussière d'aspect cotonneux qui s'accumulent sous les meubles» (Le Nouveau Petit Robert 1994 : 353).

0.7.2.6 Homographie

Certains lemmes sont des homographes. Hutchins et Somers (1992 : 87) distinguent deux types d'homographie :

- Le cas où des unités lexicales apparentées, qui s'écrivent de la même façon, sont de catégories grammaticales différentes. Par exemple : *run* qui peut être un verbe (*to run*) ou un nom (*a run*).
- Le cas où des unités lexicales ne sont pas apparentées, mais s'écrivent de la même façon et sont de même catégorie grammaticale. Un bon exemple de ce type d'homographe est le nom anglais *bank* qui peut désigner la rive d'une rivière ou une institution financière.

0.7.2.7 Propriétés sémantiques

Les propriétés sémantiques sont des attributs, tels que 'concret', 'abstrait', 'humain', 'animé', etc., associés aux mots du dictionnaire.

Such features indicate not only the potential range of extra-linguistic objects to which they may refer (i.e. assuming a matching of semantic features and real-world attributes), but also the appropriate conjunction of words in texts (sentences), e.g. *girl* and *dress*, *chair* and *sit* (Hutchins et Somers 1992 : 19).

0.7.2.8 Sous-catégorisation

La sous-catégorisation indique les environnements syntaxiques dans lesquels un mot peut apparaître. L'information typique sur la sous-catégorisation consiste à déterminer, par exemple, si un verbe est transitif ou intransitif, s'il nécessite un seul objet, un sujet et deux objets, etc. La sous-catégorisation s'applique également aux noms et aux adjectifs.

0.7.2.9 Restrictions de sélection

Les restrictions de sélection décrivent les propriétés sémantiques de l'environnement grammatical.

Subcategorization information indicates that, for example, the verb *button* occurs with a noun phrase OBJECT. In fact, we know much more about the verb than this — the OBJECT, or in terms of semantic roles, the PATIENT, of the verb has to be a 'buttonable' thing, such as a piece of clothing, and that the SUBJECT (more precisely AGENT) of the verb is normally animate. Such information is commonly referred to as the **selectional restrictions** that words place on items that appear in constructions where they are the HEAD (Arnold *et al.* 1994 : 93-94).

0.7.2.10 Décalage

Les *décalages* sont des «phénomènes [qui] se définissent par rapport à des éléments de la description sémantique» (Heid 1993 : 173) : la langue de départ (LD) et la langue d'arrivée (LA) ne disposent pas toujours d'une unité lexicale qui exprime les mêmes informations sémantiques, parce que chaque langue découpe différemment la réalité. On peut distinguer trois sous-cas différents de décalage :

- «la langue cible fait moins de distinctions sémantiques que la langue source, ou bien certaines de ces distinctions sont exprimées par le même lexème» (Heid 1993 : 173). Pour régler ce problème, on a recours à la traduction hyperonymique (ex. *poil* ou *cheveux* ↔ *hair*);
- «la langue cible fait davantage de distinctions sémantiques ou des distinctions plus fines que la langue source» (Heid 1993 : 173). Pour régler ce problème, qui est plus difficile à résoudre que le précédent, on a recours à la traduction

hyponymique (ex. Jean boit un thé chaud ↔ *Jean trinkt einen warmen Tee* \ *Jean trinkt einen heissen Tee* (Heid 1993 : 175));

- «la langue cible lexicalise d'autres distinctions sémantiques, c'est-à-dire que les critères de découpage ne sont pas comparables» (Heid 1993 : 173).

Le champ lexical des verbes de mouvement en est un bon exemple : le français utilise des séries de verbes distincts pour le type de mouvement (par exemple, *rampier, nager, courir*, etc.) et pour le déplacement orienté dans l'espace (*entrer, sortir, partir, traverser*, etc.). L'allemand, au contraire, possède des verbes à préfixe (*hinauskriechen, durchschwimmen, herauslaufen, weglaufen*, etc.) qui couvrent les deux paramètres et regroupent des éléments des deux séries, ainsi que des verbes simples qui expriment un déplacement et qui prennent des compléments directionnels pour le déplacement orienté (par exemple, *kriechen, schwimmen, laufen*). Or, les deux paramètres de classification — le type de déplacement dans l'espace et la modalité de déplacement — ne sont pas comparables et ne donnent pas de hiérarchie immédiate.

Il sort du trou en rampant.	↔	<i>Er kriecht aus dem Loch.</i>
Il entre dans la salle en courant.	↔	<i>Er rennt in den Saal.</i>
Il traverse le fleuve en nageant.	↔	<i>Er schwimmt durch den Fluß</i>

(Heid 1993 : 175).

0.7.3 Grammaire

0.7.3.1 Grammaires formelles

D'un point de vue mathématique, une grammaire est un système formel défini comme un ensemble de règles parfaitement explicites, applicables de façon mécanique, qui transforment une certaine chaîne de symboles (la chaîne d'entrée ou *input*) en une autre chaîne de symboles (la chaîne de sortie ou *output*). Ces chaînes de symboles correspondent à des chaînes de caractères pris dans un alphabet particulier (Wehrli, 1997 : 25-26).

En somme, les grammaires formelles ne contiennent que des symboles, par opposition aux autres types de grammaires qui sont formées d'unités lexicales, et sont constituées de règles qui suivent rigoureusement une logique mathématique.

0.7.3.2 Analyse syntagmatique

L'analyse syntagmatique, aussi connue sous le nom d'analyse locale, est restreinte aux groupes de mots comme les groupes-sujets, les prédicats, les groupes-compléments, etc.

0.7.3.3 Analyse syntaxique

L'analyse syntaxique est l'analyse grammaticale de la phrase complète, c'est-à-dire l'analyse de l'interdépendance entre les différentes unités syntagmatiques d'une phrase.

0.7.3.4 Grammaire de cas

La grammaire de cas analyse la phrase selon les fonctions sémantiques pertinentes des entités impliquées dans un procès (ex. agent, instrument, objet).

0.7.3.5 Divergence

Les *divergences*³ sont des phénomènes où «les informations dénotationnelles sont (grosso modo) identiques dans les deux langues, mais d'autres propriétés sont distinctes; ces différences peuvent concerner la catégorie, la sous-catégorisation syntaxique, la pragmatique, le potentiel morphologique, etc. des lexèmes en question ou encore la relation entre la structure argumentale

³ Le cas des divergences est beaucoup mieux documenté que celui des décalages dans les ouvrages consacrés à la traduction automatique .

et la construction syntaxique» (Heid 1993 : 173). Voici, pour les illustrer, un tableau des types de divergences discutés par Dorr (1992) (cité dans Heid 1993 : 179) et commentés par Heid (1993 : 179).

Divergence type	Translation Example	Commentaires
Categorial	EN: <i>I am hungry</i> ↕ DE: <i>Ich habe Hunger</i> « <i>I Have hunger</i> »	Changement de catégorie. [...]
Demotional	EN: <i>I like eating</i> ↕ DE: <i>ich esse gern</i> « <i>I eat likingly</i> »	Réalisation par des verbes ou adverbes. [...]
Promotional	EN: <i>John usually goes home</i> ↕ ES: <i>Juan suele ir a casa</i> « <i>John tends to go (to) home</i> »	Réalisation par des verbes ou des adverbes. [...]
Conflational	EN: <i>I stabbed John</i> ↕ ES: <i>Yo le di puñaladas a Juan</i> « <i>I gave knife-wounds to John</i> »	Argument incorporé vs ouvertement réalisé. [...]
Lexical	EN: <i>John broke into the room</i> ↕ ES: <i>Juan forzó la entrada al cuarto</i> « <i>John forced entry to the room</i> »	Distribution différente des éléments sémantiques sur les lexèmes. [...]
Thematic	EN: <i>I like Mary</i> ↕ ES: <i>Me gusta Maria</i> « <i>Mary pleases me</i> »	Relation entre structure argumentale et structure syntaxique. [...]

0.7.4 Langage et TA

0.7.4.1 Langage contrôlé

L'utilisation d'un langage contrôlé vise à adapter des textes de départ afin que le système de TA puisse reconnaître et traiter le plus efficacement possible les constructions et le vocabulaire contenus dans ces textes. Les constructions ambiguës et les termes polysémiques sont donc évités lorsque les rédacteurs rédigent des textes. Le principal avantage du langage contrôlé est que les textes d'arrivée nécessiteraient moins de postédition que ceux produits à partir de textes de départ sans langage contrôlé.

The use of controlled language with MT systems must be distinguished from the 'sublanguage'⁴ approach to MT ([...]). In the latter case, the system itself is designed to deal with the vocabulary and typical constructions of a specific subject area and/or document type; but there need not be any restrictions on writers or on the texts input to the system. Controlled language is not limited to a sublanguage, it may range over all the subject areas covered by a particular user; and the MT system itself is not designed to deal only with texts in the controlled forms, it can deal with uncontrolled input (even if less successfully) (Hutchins et Somers 1992 : 152).

0.7.4.2 Sous-langage

Un sous-langage consiste en une terminologie et des structures de phrases associées à un domaine particulier.

[...] a sublanguage is not simply an arbitrary subset of the set of sentences of a language. Factors which help to characterize a sublanguage include (i) limited subject matter, (ii) lexical, syntactic and semantic restrictions, (iii) "deviant" rules of grammar, (iv) high frequency of certain constructions, (v) text structure, (vi) use of special symbols.

⁴ Le sous-langage est défini dans la sous-section suivante (0.7.4.2).

(iii) refers to rules describing sentences which, though quite normal in a given sublanguage, are considered ungrammatical in the standard language. Such sentences must be considered grammatical in the sublanguage. (iii) also refers to rules describing cooccurrence restrictions within a sublanguage that do not exist in the standard language (John Lehrberger 1982 : 102).

0.7.5 Traduction et automatisation

0.7.5.1 Traduction humaine assistée par ordinateur (THAO)

La traduction humaine assistée par ordinateur est la traduction effectuée par un humain qui fait un usage limité de l'ordinateur. En d'autres mots, la composante informatique permet au traducteur d'avoir accès à des dictionnaires ou à des banques de données terminologiques, et de faire de l'édition de texte, de la gestion de glossaires, de la constitution de concordances, etc.

0.7.5.2 Traduction automatique assistée par l'homme (TAAH)

On parle de traduction automatique assistée par l'homme lorsque l'ordinateur fait la traduction à proprement parler, mais qu'il y a néanmoins une intervention humaine. Cette intervention humaine peut s'effectuer à trois niveaux : par la préédition (intervention avant traduction), par la traduction interactive (intervention pendant la traduction) ou par la postédition (intervention après traduction)⁵. Il faut toutefois noter que ces catégories ne sont pas exclusives, c'est-à-dire que, généralement, s'il y a préédition ou interaction, il y aura aussi postédition, etc.

⁵ Ces trois types d'intervention humaine seront expliqués plus en détails à la section 1.3 (chapitre 1).

0.7.5.3 Traduction entièrement automatique (TEA)

Contrairement aux deux derniers niveaux de TA, la traduction entièrement automatique ne requiert pas d'intervention humaine entre le texte de départ et le texte d'arrivée. En d'autres mots, la traduction entièrement automatique est conçue comme un système non interactif.

0.8 Conclusion

Les concepts de base présentés ci-dessus sont nécessaires pour l'étude des dictionnaires automatiques et pour leur comparaison avec les dictionnaires traditionnels.

Chapitre 1 : INFORMATION GÉNÉRALE SUR LES SYSTÈMES DE TA

1.0 Introduction

Ce chapitre donne de l'information générale sur les systèmes de TA telle que les types de systèmes, les phases du processus de traduction en TA et les modes de traduction (différé ou interactif). Cette information est essentielle pour bien comprendre comment ces systèmes ont été conçus et structurés. De plus, elle permet d'avoir un premier aperçu de la place qu'occupent les dictionnaires dans les systèmes de TA.

1.1 Différents types de systèmes

Il existe plusieurs types de systèmes de TA, dont les systèmes directs, les systèmes de transfert et les systèmes interlangues, ainsi que les systèmes basés sur l'exemple ou sur les statistiques. Étant donné que le dictionnaire ne joue pas le même rôle central dans ces deux derniers types de systèmes, nous ne les avons pas inclus dans notre thèse.

1.1.1 *Systèmes directs*

Les systèmes directs sont conçus pour une seule paire de langues et sont généralement unidirectionnels (ex. anglais → français). Il existe toutefois certaines exceptions, notamment Globalink (voir section 2.1.1). Ces systèmes ne font appel à aucune théorie linguistique générale; ils dépendent plutôt des dictionnaires bilingues, de l'analyse morphologique et des logiciels de traitement de textes pour produire des traductions.

En général, les systèmes directs contiennent :

- (i) un (ou des) dictionnaire(s) bilingue(s) donnant les équivalents possibles en LA pour chaque unité lexicale en LD;
- (ii) des règles de désambiguïsation locale qui tentent de sélectionner le bon équivalent dans un contexte donné, en particulier dans le cas d'homographes de catégories grammaticales différentes;
- (iii) des règles d'ajustement consistant à mettre les unités lexicales en LA dans l'ordre approprié en LA, à ajouter ou à enlever des unités lexicales le cas échéant, etc.

Si nous prenons la phrase *Fresh water enters the small reservoir*, tirée de Lehrberger et Bourbeau (1988 : 12), et que nous dressons une liste de tous les équivalents possibles pouvant être donnés dans un dictionnaire bilingue automatique pour chaque unité lexicale dans la phrase, nous obtenons les résultats suivants (Lehrberger et Bourbeau 1988 : 12) :

fresh	water	enters
ADJ frais [m, sg, pl] fraîche [f, sg] fraîches [f, pl]	NOUN eau [f, sg] VERB [transitive] arroser [infinitive] arrose [1, 3, sg, pres] arroses [2, sg, pres] arrosions [1, pl, pres] arrosez [2, pl, pres] arrosent [3, pl, pres]	VERB [transitive] entre [3, sg, pres] pénètre [3, sg, pres] monte [3, sg, pres]
the	small	reservoir
DEF ARTICLE le [m, sg] la [f, sg] les [m, f, pl]	ADJ [prenominal] petit [m, sg] petite [f, sg] petits [m, pl] petites [f, pl]	NOUN réservoir [m, sg]

Étant donné que l'unité lexicale *water* peut être aussi bien un nom qu'un verbe, le système de TA doit faire appel à des règles grammaticales (ex. ADJ + VERB + VERB + DEF ART + ADJ + NOUN ≠ phrase cohérente, alors que ADJ + NOUN + VERB + DEF ART + ADJ + NOUN = phrase cohérente, donc phrase à traduire) afin de sélectionner l'entrée de dictionnaire appropriée au contexte. Les systèmes directs renferment tous une forme ou une autre d'analyse grammaticale de la phrase en LD.

La figure suivante (Hutchins 1986 : 54) illustre le processus de traduction dans un système direct (avec l'étape de consultation du (ou des) dictionnaire(s) bilingue (s)).

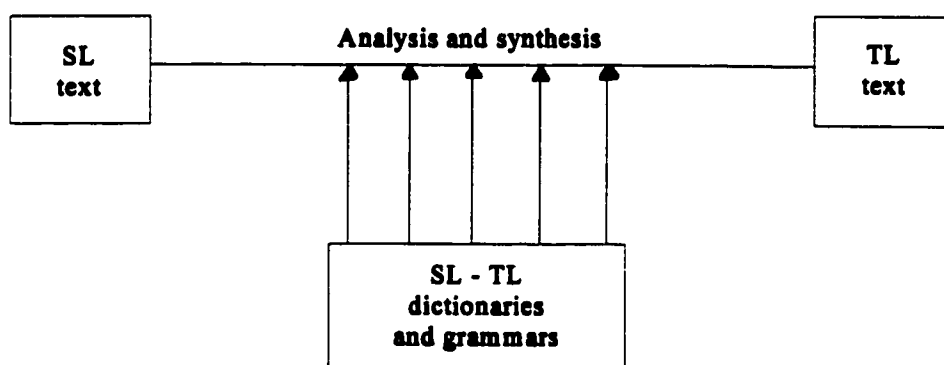


Figure 1 Système direct

1.1.2 Systèmes de transfert

Les systèmes de transfert, comprenant des représentations syntaxiques pour la LD et la LA, sont conçus en trois étapes. La première étape consiste à convertir des textes en représentations intermédiaires (quelquefois appelées «structures d'interface» (SI)), où les ambiguïtés sont traitées sans tenir compte de la LA. La deuxième étape consiste à convertir ces représentations en représentations équivalentes dans la LA. Enfin, la troisième étape consiste à générer les textes finaux.

Dans un système de transfert, les modules d'analyse et de génération sont indépendants les uns des autres en plus d'être indépendants des modules de transfert. C'est-à-dire que le même module d'analyse peut être utilisé avec différents modules de transfert, et de même avec les modules de génération. Par exemple (voir figure 2 — Pierre Isabelle, communication personnelle), un système traduisant du français vers l'anglais utilisera le même module d'analyse en français qu'un système traduisant du français vers l'espagnol ou l'allemand parce que le module d'analyse est indépendant du module de transfert (Pierre Isabelle — communication personnelle).

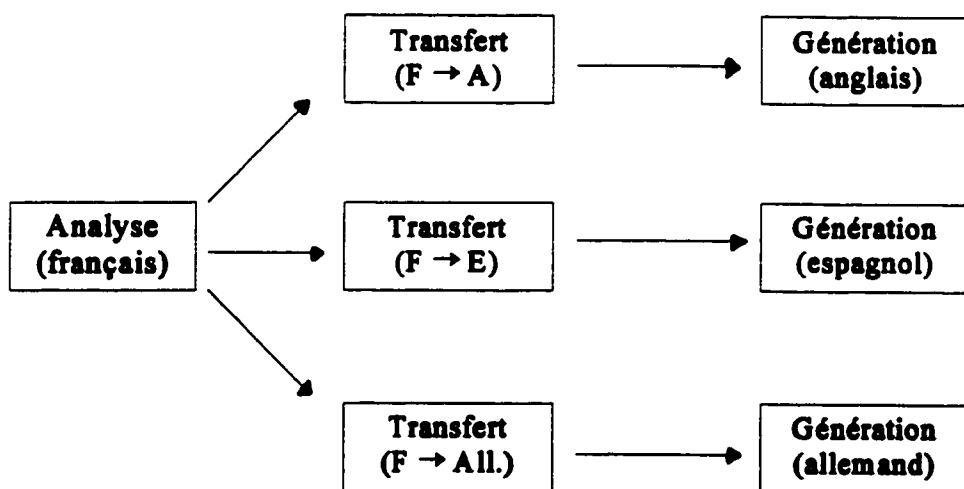


Figure 2: Modules des systèmes de transfert

Les différences de vocabulaire et de structure entre les langues sont donc traitées dans le programme de transfert intermédiaire, qui établit les équivalences pour une paire de langues donnée. Selon les systèmes, les modules de transfert sont unidirectionnels ou réversibles.

Ces systèmes possèdent donc des dictionnaires unilingues pour la LD (dictionnaire d'analyse) et la LA (dictionnaire de génération), ainsi qu'un dictionnaire de transfert bilingue.

Selon Arnold (1993 : 69), les types de représentations que les différents systèmes de transfert adoptent pour les SI peuvent être classés de trois façons :

- *formellement* : la représentation standard de la SI est un arbre (ou son équivalent) dont les noeuds sont étiquetés avec des ensembles d'attributs et de valeurs;
- en fonction de la *théorie linguistique* : la plupart des systèmes basés sur le transfert utilisent une sorte de représentation de dépendance, [...]. Cependant, on trouve également des représentations qui ont le type de structure de phrase [grammaire syntagmatique] (comme dans le système METAL, [...]). En général, il n'y a aucune raison pour qu'un système basé sur le transfert ne puisse pas être basé sur une théorie linguistique qui aurait le degré approprié de précision et de couverture et permettrait une interprétation informatique raisonnable;
- en fonction du *niveau linguistique* : certains systèmes de transfert utilisent des représentations relativement superficielles [...]. D'autres sont beaucoup plus abstraites. Comme les phénomènes considérés comme «superficiels» sont aussi spécifiques à une langue, plus de représentations abstraites tendent davantage vers l'interlangue. [...] un but commun (quoique non universel) est de définir un niveau de SI où seules les feuilles («mots») sont spécifiques aux langues sources et cibles et où la géométrie des arbres et toutes les autres propriétés sont interlangues. [...] (Arnold 1993 : 69).

La figure suivante (Hutchins 1986 : 55) illustre le processus de traduction dans un système de transfert (avec les étapes de consultation des dictionnaires).

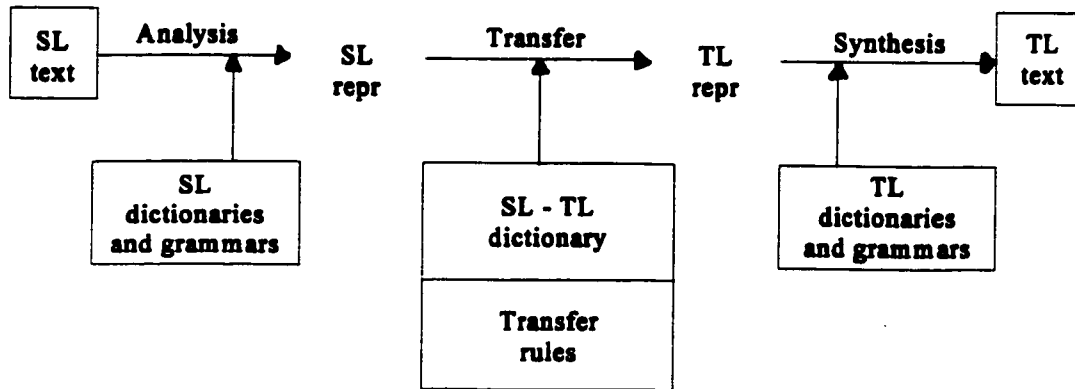


Figure 3 Système de transfert

1.1.3 Systèmes interlangues

Les systèmes interlangues sont conçus pour permettre de convertir des textes en représentations conceptuelles communes à plus d'une langue et vice-versa. La traduction consiste donc en deux étapes : de la LD à l'interlangue, puis de l'interlangue à la LA. L'interlangue utilisée par le système est obtenue en recourant le plus souvent à des primitives sémantiques ou logiques ou à un modèle du domaine, une représentation du monde, ou encore à une langue artificielle telle que l'espéranto.

Contrairement aux systèmes de transfert, les systèmes interlangues disent ne contenir aucun module de transfert. Toutefois, selon Pierre Isabelle (communication personnelle), ce type de système serait surtout basé sur ce qu'il appelle le double transfert. C'est-à-dire que le premier transfert, soit de la LD à l'interlangue, serait compris dans l'étape d'analyse et que le deuxième transfert, soit de l'interlangue à la LA, ferait partie de l'étape de génération. Ce qui fait que les

deux seules étapes apparentes sont l'analyse et la génération et qu'il semble donc ne pas avoir d'étape de transfert (voir exemple à la figure 4).

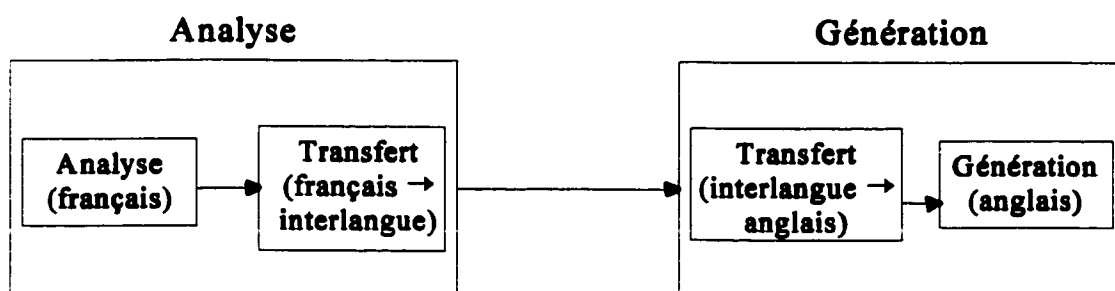


Figure 4: Modules des systèmes interlangues

Ces systèmes, tout comme les systèmes de transfert, possèdent des dictionnaires unilingues pour la LD et la LA, ainsi qu'un ou plusieurs dictionnaires «bilingues»⁶ (selon les systèmes et selon les chercheurs). Ces derniers diffèrent cependant des dictionnaires bilingues des autres types de systèmes parce qu'ils ne mettent pas directement en parallèle la LD et la LA, mais plutôt la LD et l'interlangue d'une part, et l'interlangue et la LA d'autre part.

La figure suivante illustre le processus de traduction dans un système interlangue (avec les étapes de consultation des dictionnaires) tel que vu par Pierre Isabelle (communication personnelle).

⁶ Même si nous utilisons, faute de mieux, le terme «dictionnaires «bilingues»», nous avons mis «bilingues» entre guillemets, car ces dictionnaires ne font pas la correspondance entre deux langues naturelles, mais entre une langue naturelle et une interlangue. De plus, selon Pierre Isabelle, il est tout à fait défendable de voir les systèmes interlangues comme comprenant les dictionnaires suivants : unilingue en LD, bilingue LD-interlangue, bilingue interlangue-LA et unilingue LA. Ce n'est cependant pas la position de la plupart des constructeurs de systèmes interlangues (Pierre Isabelle, communication personnelle).

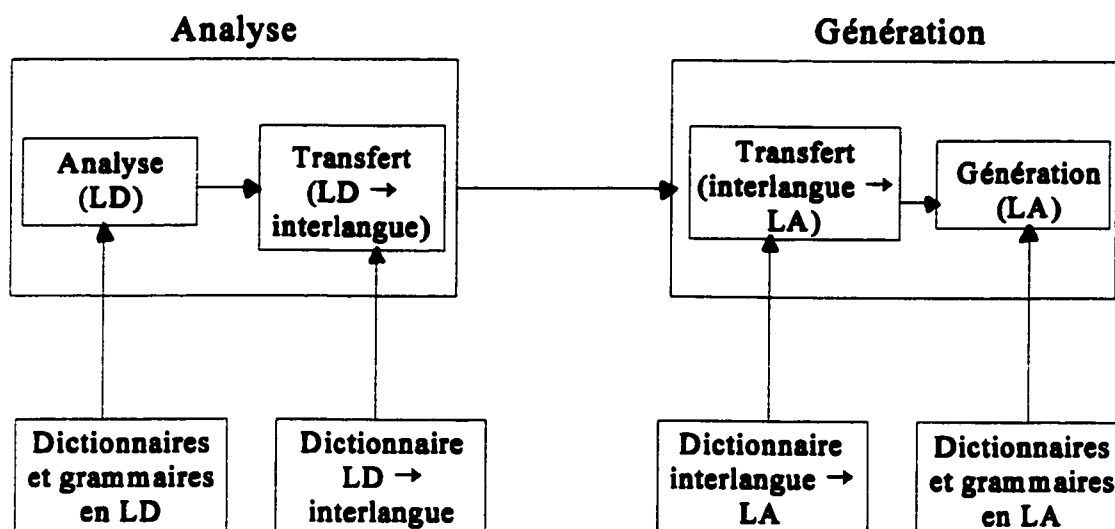


Figure 5: Système interlangue

1.1.4 Conclusion

De nos jours, la différence entre les systèmes de transfert et les systèmes interlangues s'estompe progressivement.

En résumé, la dichotomie transfert/interlangue se réduit au rôle que tient le lexique dans l'établissement des équivalences traductiques. Si le système relie les LS⁷ et LC⁸ par le lexique (ce qui ne le réduit pas au mot à mot), alors son architecture est de transfert. Si le lien entre les deux langues réfère aux objets du monde et à leurs relations, alors le système traduit par interlangue. Au delà de ce choix théorique de base, de nombreux systèmes mêlent plus ou moins les deux paradigmes (Jacqmin 1993 : 50).

En fait, selon Harold Somers (communication personnelle), très peu de chercheurs en TA parlent encore de la comparaison transfert/interlangue pour plusieurs raisons : (1) Les systèmes typiques

⁷ LS (langue source) = LD.

⁸ LC (langue cible) = LA.

de transfert fonctionnent maintenant à un niveau raisonnablement profond ('deep level') de représentation syntaxique, soit environ le même degré de «profondeur» que n'importe quel système interlangue imaginable (et fonctionnel); (2) l'ensemble des chercheurs en TA sont d'accord pour dire qu'un vrai système interlangue serait trop complexe pour être mis en oeuvre. Par conséquent, la plupart des chercheurs préfèrent un système hybride avec une sorte de représentation syntaxique interlangue et un dictionnaire du style de ceux des systèmes de transfert; (3) presque tous les systèmes qui sont opérationnels sont essentiellement des systèmes bilingues (c'est-à-dire pour une seule paire de langues et, souvent, unidirectionnels), c'est pourquoi la question de l'interlangue ne se pose même pas.

Pour ces raisons, et surtout pour la troisième citée ci-dessus, la prochaine section traite plus particulièrement des systèmes de transfert.

1.2 Trois phases dans les systèmes de transfert

Dans l'ensemble, les systèmes de transfert traduisent en trois phases : analyse, transfert et génération. Règle générale, des dictionnaires unilingues accompagnent les phases d'analyse et de génération, alors qu'un ou plusieurs dictionnaires bilingues (selon le système) accompagnent la phase de transfert.

1.2.1 Analyse

La tâche de l'analyse consiste à déterminer les relations linguistiques entre les unités lexicales d'une phrase et à inclure dans la représentation structurale de cette phrase tous les éléments

pertinents à la traduction. «La phase d'analyse prépare la phrase pour la traduction en la représentant sous une forme hiérarchisée [arbre] dans laquelle les unités lexicales sont accompagnées d'une description des phénomènes susceptibles de conditionner leur traduction» (Chevalier *et al.* 1981 : 36-37).

Pour passer d'une phrase à une représentation en arbre, l'ordinateur consulte d'abord le (ou les) dictionnaire(s) d'analyse. L'information contenue dans ces dictionnaires indique les phrases acceptables pour le système en déterminant le type d'entité qu'une unité lexicale peut désigner (par exemple, pour TAUM/AVIATION *tank* peut désigner un contenant mais pas un véhicule) ou le type de sujet et d'objet direct avec lesquels un verbe peut se combiner. En d'autres mots, le dictionnaire d'analyse détermine des contextes possibles d'une unité lexicale.

1.2.2 Transfert

Durant la phase de transfert, l'ordinateur consulte un dictionnaire bilingue (ou plusieurs, selon le système) et substitue aux unités lexicales en LD les équivalents en LA.

Les entrées des dictionnaires de transfert sont écrites par un linguiste ou un traducteur qui doit d'abord «examiner pour chaque unité lexicale, l'ensemble de ses contextes possibles d'après le dictionnaire d'analyse et décider, pour chaque contexte, quel est l'équivalent approprié. Il doit ensuite écrire, dans un langage de programmation spécialisé, une séquence d'instructions comportant des tests sur les contextes pertinents et des opérations de traduction [...]» (Chevalier

et al. 1981 : 39). Ces opérations de traduction vont de la simple substitution de l'équivalent approprié à la modification de l'environnement de l'unité traduite.

Si l'analyse inclut la formation d'une représentation hiérarchisée d'une structure lexicale en LD, le transfert doit obligatoirement convertir cette représentation en LD en une représentation en LA.

1.2.3 Génération

La phase de génération consiste à reproduire la phrase selon les règles de la syntaxe et la morphologie de la LA. En général, la génération n'est pas aussi complexe que l'analyse. Elle comprend presque toujours la dérivation des formes morphologiques adéquates pour les mots en LA (à moins que les dictionnaires de génération soient des dictionnaires de formes, c'est-à-dire contenant toutes les variantes morphologiques d'une unité lexicale). Les formes irrégulières, par contre, sont habituellement traitées par le dictionnaire (ex. *went* serait codé directement comme le passé de *go*).

Si l'analyse inclut la formation d'une représentation hiérarchisée d'une structure lexicale en LD et que le transfert convertit cette représentation hiérarchisée en une représentation équivalente en LA, la génération doit, quant à elle, convertir cette dernière en une structure lexicale en LA; c'est-à-dire qu'elle doit, en quelque sorte, inverser le processus d'analyse. Toutefois, cela ne signifie pas que les règles utilisées pour l'analyse des structures d'une langue (comme LD)

peuvent simplement être inversées pour obtenir des règles pour la génération de cette même langue (comme LA) dans une autre paire (Hutchins et Somers 1992).

La génération peut se faire en plus d'une étape selon les systèmes. Par exemple, certains systèmes (comme les systèmes de transfert) possèdent un module de génération syntaxique (qui a pour but de mettre en ordre les représentations de la section à générer en LA) et un module de génération morphologique (qui a pour but de convertir les représentations en mots dans la LA; ex. *dog* + pluriel devient *dogs*) (Hutchins et Somers 1992 : 133).

Toujours selon les systèmes, le processus de génération peut être accompagné d'un ou de plusieurs dictionnaires de génération.

1.3 Différents modes de traduction

Les systèmes de TA qui sont actuellement sur le marché offrent deux modes de traduction : la traduction en mode différé, c'est-à-dire sans intervention humaine pendant la traduction, et la traduction interactive.

1.3.1 Traduction en mode différé

La traduction en mode différé consiste en une traduction faite par le système sans qu'il y ait de possibilité d'intervention humaine durant le processus de traduction. Les seules interventions humaines possibles pour ce mode de traduction sont la préédition, la postédition et la mise à jour des dictionnaires.

La préédition consiste à adapter la forme du texte (isoler les éléments à ne pas traduire, corriger l'orthographe) ou son contenu de manière à rendre le texte conforme à ce que peut traiter le système (par exemple en utilisant un langage contrôlé). Durant cette étape, les prééditeurs utilisent des outils tels que les traitements de texte, les correcteurs d'orthographe, etc.

La postédition, quant à elle, consiste en une révision humaine des textes traduits par les systèmes de TA, pour laquelle les systèmes offrent une série d'outils similaires à ceux utilisés par les prééditeurs : traitements de texte spécialisés avec accès intégré à des dictionnaires bilingues, des dictionnaires de synonymes, etc.

Parmi les systèmes qui fonctionnent en mode différé sont Météo, METAL (maintenant appelé T1⁹), Systran et Logos.

1.3.2 Traduction interactive

La traduction interactive, bien moins commune que la traduction en mode différé, consiste en une intervention humaine durant le processus de traduction, c'est-à-dire que le système de TA fait appel à l'expertise humaine pour résoudre des problèmes d'ambiguïté durant la traduction.

Dans ce mode de traduction, le système tente d'exploiter l'expertise humaine par diverses méthodes :

⁹ Pour faciliter la compréhension, nous allons garder l'ancien nom 'METAL', au lieu de T1, dans le reste de la thèse parce qu'il est mieux connu.

- en présentant à l'utilisateur un choix de paraphrases :
 «école de cuisine lyonnaise» signifie ici :
 - 1) école lyonnaise (de cuisine)
 - 2) école de cuisine (lyonnaise)
- en présentant à l'utilisateur un choix de sens pour un mot donné :
 «glace» signifie :
 - 1) eau congelée
 - 2) vitre
 - 3) crème glacée...
- en présentant à l'utilisateur un choix de référents pour un mot donné :
 «she» refers to :
 - 1) *Cathy*
 - 2) *my mother*
 - 3) *the sailboat...*
- par un dialogue sur les intentions de communication (voir Nagao et Sato [...])
 (Jacqmin 1993 : 53).

Parmi les systèmes qui fonctionnent en mode interactif sont Globalink et Ariane (GETA). Il est à noter, cependant, que certains systèmes de TA (ex. Globalink) offrent la possibilité de traduire en mode différé ou en mode interactif selon le choix de l'utilisateur.

1.4 Conclusion

Puisque le nombre et le type de dictionnaires varient selon le type de système de TA, et puisque le traducteur voit ou non le contenu des dictionnaires selon le mode de traduction, l'information donnée dans ce chapitre est nécessaire pour bien comprendre le rôle des dictionnaires dans les différents systèmes actuellement sur le marché.

Chapitre 2 : ÉTUDE DES DICTIONNAIRES UNILINGUES ET BILINGUES DANS QUELQUES SYSTÈMES DE TA

2.0 Introduction

Dans le chapitre précédent, nous avons mentionné certains systèmes de TA. Il s'agit maintenant d'analyser les dictionnaires intégrés à quelques-uns de ces systèmes et, à partir de là, d'examiner le rôle général des dictionnaires unilingues et bilingues dans les systèmes de TA.

2.1 Analyse des dictionnaires dans quelques systèmes de TA

Dans le but de rendre l'étude des dictionnaires des systèmes choisis plus facile, nous les avons classés selon le type de système de TA auquel ils appartiennent : direct, de transfert ou interlangue. Dans chaque catégorie, nous analysons d'abord les systèmes opérationnels (s'il y en a), puis les systèmes expérimentaux (encore une fois, s'il y en a).

2.1.1 *Systèmes directs*

Ce qui caractérise les systèmes directs, c'est qu'ils ne comprennent que des dictionnaires bilingues, qui sont, en général, unidirectionnels.

Les trois systèmes que nous analyserons dans cette section, soit Globalink, Systran et Météo, sont présentement sur le marché. Contrairement à Globalink, les systèmes Systran et Météo ne sont pas des systèmes directs purs et durs. Selon Whitelock et Kilby (1995), ces systèmes sont en quelque sorte hybrides parce qu'ils présentent certaines facettes qui les rapprochent des systèmes de transfert (pour Systran) et interlangues (pour Météo), sans toutefois perdre leur caractéristique première qui veut qu'ils ne contiennent que des dictionnaires bilingues.

En parlant de Systran, Whitelock et Kilby précisent :

Translation programs fall into three types: SL specific, language-pair specific and TL specific. The system therefore has the nature of a transfer system, with separation, in principle, between SL analysis and TL synthesis (Whitelock et Kilby 1995 : 41).

Au sujet de Météo, ils signalent :

It is a difficult matter to locate MÉTÉO at a precise point on the spectrum of strategies for MT, i.e. direct—transfer—interlingual. On the one hand, the process of translation is highly language-pair specific, with target language lexical items being inserted prior to analysis, and no specific stage which could be termed “transfer”. On the other hand, the operations of analysis and synthesis are notionally separated; the structure which mediates between the two could be considered a syntactic interlingua, representing an abstraction of structural features from the limited syntax source and target languages, and using target language lexical items as a matter of convenience (i.e. a limited subset of French vocabulary is the lexicon of this interlingua) (Whitelock et Kilby 1995 : 67-68).

2.1.1.1 Globalink

Le logiciel *Power Translator Pro* de Globalink, qui fonctionne sur PC (dans Windows 3.1 ou Windows 95), est l'un des rares systèmes directs¹⁰ qui utilisent un traitement bidirectionnel pour chaque paire de langues.

La version de Globalink qui est utilisée dans un environnement Windows 3.1 est constituée de deux grandes classes de dictionnaires : les dictionnaires de mots simples et les dictionnaires d'unités sémantiques. Chacune de ces deux classes de dictionnaires est divisée en deux types de

¹⁰ Bien que l'architecture du système Globalink ne soit pas connue publiquement, certaines caractéristiques nous portent à croire qu'il est un système direct. Par exemple le fait qu'il ne possède qu'un seul dictionnaire bilingue et que les traductions qu'il produit montrent clairement une pauvre syntaxe indiquent des caractéristiques propres aux systèmes directs (communication personnelle de Pierre Isabelle).

dictionnaires : le dictionnaire standard et les dictionnaires spécialisés (qui peuvent être achetés séparément du logiciel de traduction ou construits par l'utilisateur).

Il est possible d'avoir accès aux entrées des différents dictionnaires de Globalink. L'information accessible et ajoutable comprend, entre autres, des informations grammaticales supplémentaires, des équivalents additionnels (autres que l'équivalent principal), et les synonymes du mot recherché.

Lorsque l'utilisateur ajoute un mot au dictionnaire, il est important qu'il n'entre que la racine du mot et non ses formes fléchies (qui sont générées automatiquement par le système). Puisque Globalink est bidirectionnel, si l'utilisateur ajoute un mot en anglais (ex. *psycholinguistics*) dont l'équivalent en français (ex. **psycholinguistique**) n'apparaît pas dans le dictionnaire français-anglais, il doit créer l'entrée en français (**psycholinguistique**) avec son équivalent anglais (*psycholinguistics*).

En ce qui a trait à la mise à jour et à la modification des dictionnaires d'unités sémantiques, il faut noter que toutes les informations contenues dans l'enregistrement d'une unité sémantique sont modifiables. Cependant, les mots qui composent l'unité sémantique doivent déjà figurer dans un des dictionnaires de mots simples. Si ce n'est pas le cas, il faut les y ajouter avant de créer une entrée à l'un des dictionnaires d'unités sémantiques. Parallèlement, si l'équivalent en LA n'apparaît pas comme mot-vedette dans le dictionnaire inverse, il faut créer cette entrée. Si l'équivalent d'une unité sémantique est un mot simple, il faudra tout de même entrer cet

équivalent (mot simple) comme mot-vedette principal dans le dictionnaire d'unités sémantiques inverse (et non dans le dictionnaire de mots simples, parce que l'équivalent est considéré comme une unité sémantique).

Que ce soit dans les dictionnaires de mots simples ou dans les dictionnaires d'unités sémantiques, l'utilisateur ne peut supprimer que les mots qu'il a lui-même créés ou modifiés (et non ceux qui se trouvaient à l'origine dans les dictionnaires du système).

La version de Globalink fonctionnant dans un environnement Windows 95 s'utilise de la même façon que celle qui fonctionne dans l'environnement Windows 3.1, excepté que les dictionnaires de mots simples et les dictionnaires d'unités sémantiques sont regroupés en un seul dictionnaire pour chaque type de dictionnaire (général ou spécialisé¹¹).

Selon Allison Sutherland (communication personnelle), la consultation des dictionnaires a lieu pendant la première des cinq étapes de l'analyse linguistique. À cette étape, les formes canoniques des mots sont déterminées, tandis qu'à l'étape suivante — étape de désambiguïsation —, leur partie du discours est décidée.

¹¹ Il n'y a qu'un dictionnaire général, mais il est possible de créer plusieurs dictionnaires spécialisés en fonction des domaines ou des utilisateurs.

2.1.1.2 Systran

Les bases de données lexicales de Systran, qui est le système opérationnel le plus connu, consistent en deux dictionnaires bilingues qui peuvent être mis à jour régulièrement : le dictionnaire de radicaux, contenant comme mots-vedettes des mots simples, et le dictionnaire d'unités lexicales complexes. Une série de dictionnaires de traduction sont créés automatiquement à partir de ces deux dictionnaires principaux : un seul à partir du dictionnaire de radicaux et cinq à partir du dictionnaire d'unités lexicales complexes. Ce sont ces dictionnaires, créés à partir des deux dictionnaires principaux, que le système consulte durant le processus de traduction.

Le dictionnaire «Main Stem», dérivé du dictionnaire de radicaux, donne une description morphologique, syntaxique et sémantique complète pour chaque mot en LD : catégorie grammaticale, rection, valence, accord, transitivité, type de nom (ex. animé, abstrait, etc.) et marques sémantiques (ex. propriété physique, contenant, aliment, etc.). De plus, il donne un équivalent de la forme de base avec l'information grammaticale et syntaxique nécessaire pour la génération. Il est important de noter que chaque mot-vedette dans ce dictionnaire ne possède qu'un seul équivalent, soit l'équivalent donné par défaut. Donc, malgré sa richesse, ce dictionnaire seul ne pourrait fournir que des traductions mot à mot. Ce sont les dictionnaires contextuels (présentés ci-dessous) qui permettent de varier la traduction de chaque mot en fonction du contexte. Le traitement des homographes est aussi divisé entre le dictionnaire «Main Stem» et les dictionnaires contextuels : les homographes dont les catégories grammaticales sont différentes sont entrés séparément dans le dictionnaire «Main Stem» et font l'objet d'un renvoi à

l'entrée de base; quant aux homographes de même catégorie grammaticale, ils sont traités comme des polysèmes dans les dictionnaires contextuels.

Les dictionnaires créés à partir du dictionnaire d'unités lexicales complexes et appelés «dictionnaires contextuels» fournissent les données nécessaires pour modifier les analyses et les traductions en fonction du contexte et forment un ensemble de dictionnaires qui sont utilisés à différents stades de l'analyse et de la traduction. De plus, comme il a été mentionné dans le paragraphe précédent, ce sont ces dictionnaires qui traitent les polysèmes en permettant d'entrer plusieurs équivalents, qui sont distingués l'un de l'autre par des codes de domaine (TG). Il y a cinq types de dictionnaires contextuels (Hutchins et Somers 1992 : 179) :

- (a) Le dictionnaire idiomatique traite des expressions idiomatiques (ex. *on the one hand, in order to*), qui peuvent, dans certains cas, être traduites par un seul mot en LA;
- (b) Le dictionnaire des groupes nominaux présente des syntagmes nominaux (ex. *hydraulic brake*) qui ne forment pas une unité sémantique fermée, ainsi que des syntagmes nominaux qui constituent une seule unité sémantique (ex. *pomme de terre*). Dans certaines versions de Systran, ces deux types d'unités sont placés dans des dictionnaires différents;
- (c) Le dictionnaire homographique donne l'information syntaxique contextuelle nécessaire à la désambiguïsation de certains homographes;
- (d) Les dictionnaires analytiques contiennent les exceptions aux règles syntaxiques générales qui s'appliquent à des mots particuliers (surtout les mots grammaticaux). Ces dictionnaires peuvent être utilisés à divers stades de l'analyse;

- (e) Le dictionnaire conditionnel intervient à l'étape du transfert pour faire la sélection finale de l'équivalent lexical. Il comprend l'information syntaxique et sémantique nécessaire pour faire la distinction entre les équivalents possibles. Dans certains cas, il présente un grand nombre de caractéristiques contextuelles. Par exemple, 400 entrées permettent de faire la distinction entre les traductions *huile* et *pétrole* pour le mot anglais *oil*.

L'utilisateur ne peut modifier, de quelque façon que ce soit, les dictionnaires fournis dans le système. Toutefois, il a la possibilité d'ajouter des entrées dans un dictionnaire spécial qui s'appelle le 'Customer Specific Dictionary' (CSD). Ce dictionnaire permet à l'utilisateur d'entrer des unités lexicales (simples ou complexes) qui ne se trouvent pas dans les dictionnaires du système ou de changer les équivalents des unités qui s'y trouvent.

La figure suivante (Hutchins et Somers 1992 : 181) montre le processus de traduction de Systran et l'utilisation des dictionnaires par le système :

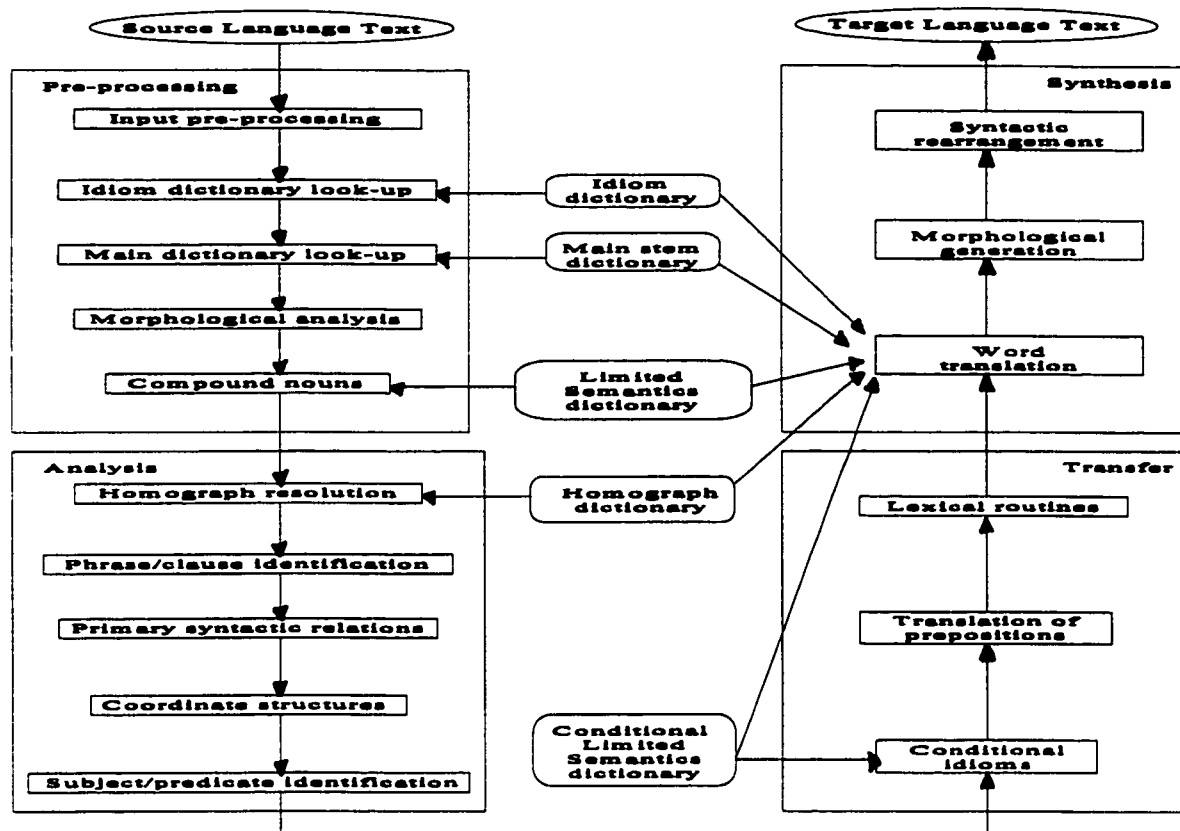


Figure 6 Processus de traduction de Systran

2.1.1.3 Météo

La version originale de Météo¹², système opérationnel depuis 1977, comprend trois dictionnaires bilingues anglais-français : un dictionnaire général, un dictionnaire d'idiomes et un dictionnaire de noms de lieux (Chevalier *et al.* 1978 : 25; Hutchins et Somers 1992 : 209).

¹² Puisque la dernière version de Météo n'est pas accessible, et parce qu'elle a considérablement changé, nous avons décidé de présenter la version originale.

Selon Hutchins et Somers (1992 : 210-211), le dictionnaire principal, soit le dictionnaire général, est un dictionnaire de formes. Chaque entrée pour un mot anglais donne l'équivalent français (parfois plus d'un seul), la catégorie grammaticale, les traits sémantiques, ainsi que l'information morphologique et syntaxique de la LA (ex. AMOUNT = N((F,MSR),QUANTITÉ); où N indique le nom, F, le féminin et MSR, un nom de mesure). Pour les noms, l'information morphologique est limitée au genre (féminin ou non) et au nombre (pluriel ou non). L'indication des terminaisons au pluriel des noms français n'est pas nécessaire pour les mots-vedettes au singulier puisque toutes les formes anglaises (aussi bien au singulier qu'au pluriel) sont entrées dans les dictionnaires avec leurs équivalents respectifs (ex. *area* → **région** et *areas* → **régions**). Cette façon de procéder permet de traiter pareillement les différences de nombre entre les deux langues (*skies* → **ciel**). En ce qui a trait aux adjectifs français, l'information syntaxique vise simplement à déterminer si l'adjectif doit précéder le nom français ou non, et l'information morphologique indique les modifications nécessaires pour produire le féminin et le pluriel. L'information syntaxique pour les verbes et les adverbes français est minime. Les entrées pour les adverbes indiquent s'ils sont attachés aux adjectifs, aux verbes ou aux prépositions, et celles des verbes, simplement s'ils sont transitifs ou intransitifs. Les traits sémantiques quant à eux sont attachés aux noms, aux adjectifs, aux adverbes et aux prépositions.

Le dictionnaire d'idiomes contient les traductions des idiomes (qui sont définis comme n'importe quelle suite de plus d'un mot devant être traitée comme un seul mot) appartenant au domaine de la météorologie (ex. *blowing snow* → **poudrerie**) et des règles de normalisation de l'anglais (ex. *kilometers per hour* ou *kilometres per hour* → **KMH**) (Chevalier et al. 1978 : 25).

Enfin, le dictionnaire des noms de lieux contient seulement les noms qui sont différents en anglais et en français, ceux qui sont composés de plus d'un mot (ex. *Greater Vancouver* → **Vancouver et banlieue**), et ceux qui doivent contenir de l'information linguistique particulière afin d'obtenir la bonne traduction (ex. les noms qui sont au pluriel, qui ont un article défini ou qui sont au féminin). Tous les autres noms de lieux ne sont pas traités, car ils sont considérés comme des noms propres et ne sont, par conséquent, pas traduits.

Le processus de traduction de Météo comprend trois modules de traitement : un pour l'analyse syntaxique de l'anglais, un pour la génération syntaxique du français et un pour la génération morphologique du français. Puisque le style télégraphique des bulletins météorologiques en français suit la même structure que celui des bulletins en anglais, le système ne requiert pas de module de transfert. De plus, la simplicité du vocabulaire et la variation morphologique limitée de l'anglais ont permis d'omettre un module pour l'analyse morphologique de l'anglais. C'est pourquoi toutes les variantes morphologiques sont entrées dans les dictionnaires du système. La sélection de l'équivalent approprié (qui se distingue par ses traits sémantiques) ne se fait qu'aux derniers stades de l'analyse syntaxique.

Il existe, outre la version de Météo décrite ci-dessus, d'autres versions de ce système. Par exemple, en octobre 1984, une nouvelle version, appelée Météo-2, a été installée pour fonctionner sur micro-ordinateurs. Alors que la version originale était écrite en Systèmes-Q, Météo-2 est écrite dans le langage de programmation GramR, mis au point par le groupe John Chandioux Consultants Inc. Enfin, un autre point tournant dans l'histoire de Météo est

l'installation, en 1989, d'un système de traduction, allant du français vers l'anglais, de bulletins météorologiques du Centre météorologique du Québec (Hutchins et Somers 1992 : 208).

2.1.2 *Systèmes de transfert*

Comme nous l'avons vu dans le chapitre précédent, les systèmes de transfert possèdent deux types de dictionnaires : unilingues pour l'analyse et la génération et bilingues pour le transfert.

Les systèmes METAL, Logos, Tovna, qui sont des systèmes opérationnels, ainsi que SUSY et Ariane (GETA), qui sont des systèmes expérimentaux, font partie des systèmes de transfert.

2.1.2.1 METAL

METAL, qui a récemment été racheté par la société GMS, elle-même devenue une filiale de Lernout & Hauspie, et qui a été renommé 'T1', est un système de transfert opérationnel qui comprend trois dictionnaires par paire de langues : un dictionnaire unilingue en LD, un dictionnaire unilingue en LA et un dictionnaire bilingue. Les dictionnaires unilingues en LD et en LA contiennent l'information morphologique, syntaxique et sémantique de base.

Les entrées des dictionnaires unilingues donnent le radical, la catégorie grammaticale, les variantes morphologiques, le nombre, la personne, etc. Elles comprennent aussi une valeur de préférence qui permet de choisir entre les différentes analyses possibles, les indications sur les collocations lexicales, et l'indication des domaines d'emploi.

Exemple¹³ : (THIS CAT (DET)

ALO (this) (these)
NU (SG) (PL) \$
PCL (WI WF)

•
•
•

)

Les entrées pour les noms indiquent leurs classes flexionnelles et ajoutent un trait sémantique (ex. entité, commodité, etc.) servant à restreindre les cooccurrences. Les entrées pour les verbes comprennent les rôles sémantiques (ex. agent, bénéficiaire, etc.) avec les traits qui restreignent les valeurs sémantiques des arguments, le type des arguments (ex. syntagme nominal, syntagme prépositionnel, préposition, etc.) et la fonction syntaxique de surface des arguments (ex. sujet, objet, etc.); enfin, les entrées pour les verbes comprennent aussi une spécification de leur mode de valence : intransitif avec un seul argument (agent), intransitif avec deux arguments (agent et emplacement), transitif avec deux arguments (agent et cible), etc.

Les dictionnaires unilingues sont conçus pour être neutres et indépendants, et pour être utilisés autant pour l'analyse en LD que pour la génération en LA, peu importe l'autre langue de la paire. Au contraire, les dictionnaires de transfert (bilingues) sont conçus pour une paire de langues donnée dans une seule direction.

METAL a hiérarchisé ses vocabulaires bilingues : il existe des modules lexicaux de base, qui peuvent s'appliquer indépendamment du domaine (un module pour les mots outils, un pour le

¹³ Cet exemple est tiré de Whitelock et Kilby (1995 : 175).

vocabulaire général et un pour le vocabulaire technique commun). Les utilisateurs peuvent ajouter à ces modules autant de glossaires spécialisés qu'ils le désirent et spécifier l'ordre dans lequel ils devraient être consultés. Il est possible non seulement de créer des glossaires pour des domaines particuliers, mais aussi pour des pays donnés (ex. pour que l'allemand *Lastwagen* se traduise par **truck** aux États-Unis et par **lorry** en Grande-Bretagne).

Les entrées du dictionnaire bilingue sont assez simples en général (ex¹⁴. (**give (geben) VST (CAT VST))**). Toutefois, on trouve des entrées plus complexes lorsqu'il y a plus d'un équivalent possible pour le mot en LD. L'équivalent sera alors déterminé en fonction des restrictions (ex. traits sémantiques propres au mot, domaines d'emploi, etc.) qui s'appliquent au mot en LD.

Exemple¹⁵ :

(be-lost	(gehen) VST	(CAT VST) (PX VERLOREN))
(go	(gehen) VST	(CAT VST) (PX NIL)
		(PF FIN INF PAPL))
(outgo	(gehen) VST	(CAT VST) (PX NIL)
		(PF PRPL))

L'information dans les dictionnaires bilingues peut comprendre des rôles sémantiques particuliers, la présence d'arguments d'un certain type sémantique, les changements de structures actives en constructions impersonnelles, l'ajout et le retrait d'éléments, etc.

La création des entrées lexicales est facilitée par l'Intercodeur qui agit de façon interactive avec l'utilisateur pour obtenir de l'information grammaticale et traductionnelle. L'Intercodeur

¹⁴ Cet exemple est tiré de Whitelock et Kilby (1995 : 194).

¹⁵ Cet exemple est tiré de Whitelock et Kilby (1995 : 194).

comprend un programme lexical par défaut qui accepte un minimum d'information (ex. le radical et la catégorie grammaticale), puis génère automatiquement des variantes morphologiques et encode des traits et des valeurs syntaxiques. L'Intercodeur maintient la cohérence et l'intégrité du dictionnaire à l'aide de programmes de validation automatique qui identifient les erreurs de forme ou de syntaxe. Cet outil peut être utilisé à la fois par les créateurs du système et par les utilisateurs qui créent leurs propres dictionnaires, tant pour les dictionnaires unilingues que bilingues.

Le processus de traduction et l'étape de consultation des dictionnaires sont illustrés dans la figure 7 (Hutchins et Somers 1992 : 261).

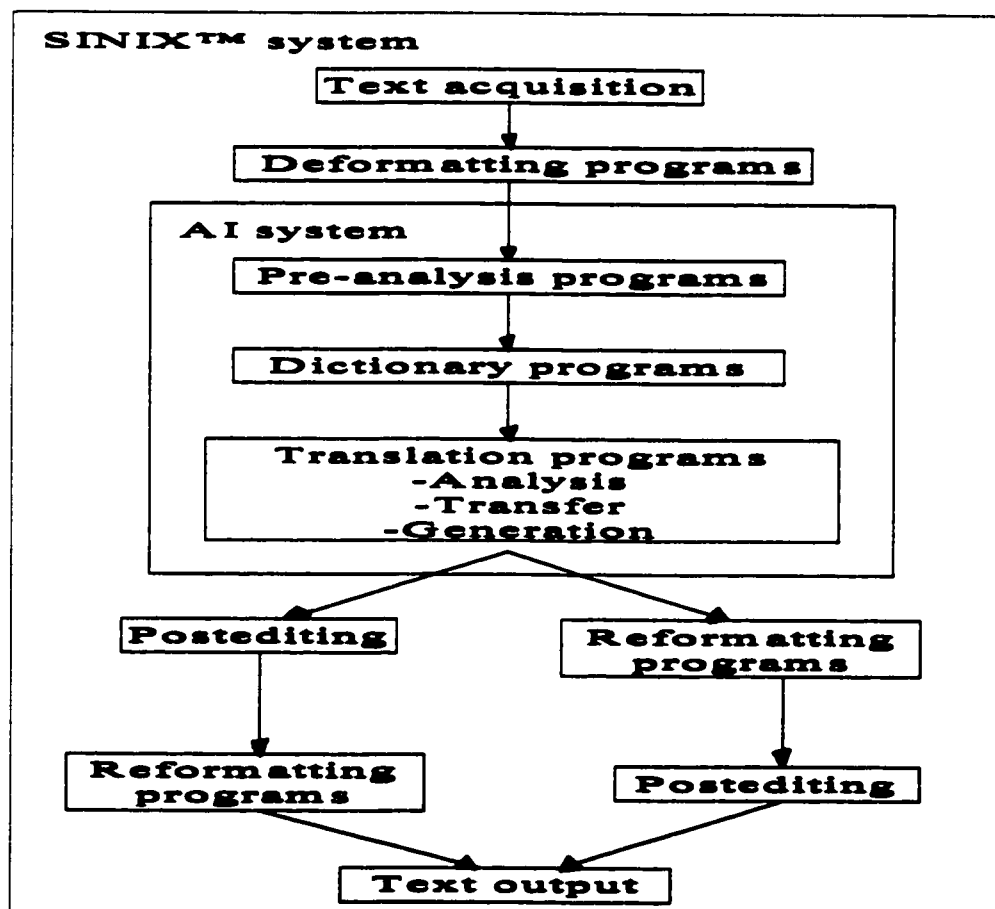


Figure 7 Processus de traduction de Metal

2.1.2.2 Logos

Logos est un système de transfert opérationnel différent de ceux que nous présentons dans cette section en ce sens qu'il ne contient qu'un seul dictionnaire de base, soit un dictionnaire bilingue¹⁶. Ce dictionnaire bilingue de base est subdivisé en un certain nombre de dictionnaires bilingues spécialisés qui sont eux-mêmes classés selon les codes de domaine et selon les codes d'entreprise (autrement dit les codes d'utilisateur).

Actuellement, pour un domaine ou un utilisateur (ou une entreprise) spécifique, un mot-vedette en LD ne peut avoir qu'un seul équivalent en LA. Cela ne signifie pas pour autant qu'il n'y ait qu'une seule entrée possible pour un mot donné. Par exemple (communication personnelle de W. Scott Bennett), le mot anglais *table* compte environ huit entrées (pour des domaines ou des utilisateurs différents) donnant différents équivalents en LA. Donc, si un utilisateur veut ajouter un équivalent additionnel pour un domaine spécifique, il n'a qu'à ajouter une neuvième entrée et ainsi de suite. La nouvelle base de données relationnelle (communication personnelle de W. Scott Bennett) élimine cependant ces restrictions : les utilisateurs peuvent maintenant créer autant d'entrées qu'ils le veulent pour un même mot en LD pour le même domaine ou le même utilisateur.

En somme, le rôle de Logos est d'offrir à l'utilisateur un dictionnaire bilingue de base contenant du vocabulaire général, tout en lui fournissant un outil de construction de dictionnaires. Cet outil,

¹⁶ Il semblerait qu'étant donné la richesse de ce dictionnaire bilingue (spécialement pour les entrées en LD), le système ne requiert pas de dictionnaires unilingues (communication personnelle de W. Scott Bennett).

nommé Alex ('Automatic Lexicographer'), permet à l'utilisateur d'entrer des mots ou des groupes de mots nouveaux dans le dictionnaire d'utilisateur, ou tout simplement de modifier l'équivalent des entrées qu'il a préparées. Il faut cependant noter qu'il est impossible de changer quoi que ce soit au dictionnaire d'origine; on peut seulement créer de nouveaux dictionnaires spécialisés.

Lors de la création d'entrées, Alex requiert les informations suivantes : le genre des noms (masculin, féminin ou neutre —dans le cas de l'allemand), le nombre (singulier seulement, pluriel seulement, ou les deux), le code d'entreprise (si plusieurs personnes utilisent Logos en même temps, il est plus facile de retrouver la terminologie de l'utilisateur par le code d'entreprise), les propriétés sémantiques du concept (ex. mot concret, abstrait, de masse, etc.), et ainsi de suite. Puisque l'analyse sémantique de Logos est assez bien développée (le système est basé sur une grammaire de cas simplifiée), l'utilisateur doit donner une certaine quantité d'information sémantique dans les entrées de dictionnaire.

L'outil qui sert à la création de règles sémantiques se nomme Samantha. Selon W. Scott Bennett (communication personnelle), les règles sémantiques sont en fait des «mini-grammar rules which the user writes to effect a certain result in the translation. [...] they are rules which allow the user to code specific translations for particular strings when the system would do something else». Cet outil permet donc d'entrer des groupes de mots comme des collocations et des expressions idiomatiques qui exigent une traduction particulière dans le dictionnaire. Tout comme Alex, Samantha est un outil interactif qui offre plusieurs choix à l'utilisateur (ex. adjectif + nom, etc.).

Lorsque tous les champs sont remplis, le groupe de mots est automatiquement ajouté à la base de données de l'utilisateur.

Le processus de traduction de Logos comprend plusieurs étapes. La première étape est en fait la consultation du dictionnaire de base ainsi que des dictionnaires spécialisés pour y trouver toutes les entrées pertinentes au texte de départ. Ensuite, le système passe à l'étape de désambiguïsation qui consiste à résoudre les ambiguïtés qui se trouvent tant au niveau des mots eux-mêmes qu'au niveau des groupes de mots et des propositions. Les étapes suivantes, appelées TRAN1, TRAN2 et TRAN3, sont en fait les étapes d'analyse de la LD avec stockage de l'information nécessaire pour le transfert : TRAN1 fonctionne au niveau des mots et des groupes de mots; TRAN2 fonctionne à un niveau plus élevé dans la phrase; et TRAN3 fonctionne aux niveaux les plus avancés dans la phrase. L'étape finale, TRAN4, effectue à la fois le transfert et la génération en LA (W. Scott Bennett, communication personnelle).

2.1.2.3 Tovna

Tovna, un système de transfert opérationnel¹⁷ fondé sur des techniques d'intelligence artificielle, est le premier système «apprenant». Non seulement l'utilisateur peut-il ajouter de nouveaux mots ou groupes de mots au dictionnaire, mais le système peut lui-même dériver certaines règles à partir des groupes de mots. L'«apprentissage» se fait principalement par un processus interactif¹⁸ de diagnostic (c'est-à-dire que le système relève, à chaque étape du processus de traduction,

¹⁷ Il semblerait que ce système n'existe plus, mais nous n'avons pas pu avoir confirmation de ce fait.

¹⁸ Tovna peut être utilisé en mode différé ou en mode interactif.

toutes les possibilités de traduction de cette étape, puis les fait passer à l'étape suivante, et ainsi de suite), d'enseignement (c'est-à-dire que l'utilisateur corrige certaines erreurs faites par le système) et de retraduction. Le système mémorise (sur des fiches) chaque étape de l'analyse d'une phrase. De cette façon, à la postédition, si l'utilisateur relève des erreurs, il peut consulter les fiches mémorisées pour trouver la cause de l'erreur et ainsi la corriger. Plus le système est utilisé, meilleure est sa performance.

L'étape du diagnostic permet à l'utilisateur de voir quels modèles de la table bilingue d'exemples¹⁹ ont servi durant le transfert. Il peut ensuite choisir un autre modèle ou en créer un nouveau, dont il peut déterminer le degré de spécificité, en entrant une séquence de mots typique comme exemple.

Après avoir «enseigné» au système, l'utilisateur peut retraduire la phrase de départ et, si nécessaire, répéter le processus d'apprentissage jusqu'à ce que la traduction soit satisfaisante. Il faut noter que si le fait de pouvoir enseigner au système donne un certain contrôle à l'utilisateur, ce dernier doit toutefois posséder un certain niveau de connaissances et d'habileté.

Malgré ces innovations et les principes de séparation et de modularité auxquels Tovna prétend adhérer, ce système semble n'offrir qu'une piètre analyse linguistique qui se limite à une analyse

¹⁹ Ces tables bilingues d'exemples donnent la règle de traduction pour chaque modèle de groupe de mots dans la LD.

au niveau des groupes de mots. En ce qui concerne l'analyse sémantique, elle semble comprendre, néanmoins, un peu plus que les restrictions de sélection (Segal 1991 : 83).

La base de données linguistiques de Tovna est constituée de cinq types d'éléments : un dictionnaire syntaxique unilingue pour chaque langue (ces dictionnaires contiennent la liste de tous les attributs grammaticaux des mots ou expressions idiomatiques d'une langue donnée), un dictionnaire bilingue pour chaque paire de langues (ces dictionnaires donnent les équivalents à partir du contexte pour chaque mot-vedette — mot ou expression idiomatique), une table bilingue d'exemples pour chaque paire de langues (voir note 19), un modèle de passage pour chaque langue (ces modèles de passage sont utilisés durant l'analyse des structures syntaxiques possibles du texte de départ) et divers fichiers pour chaque langue (ces fichiers décrivent les règles typographiques et morphologiques, les listes d'attributs et de rôles, et les automates).

Le processus de traduction de Tovna est composé de cinq étapes : la typographie, la morphologie, le passage, le transfert et la génération.

La typographie est l'étape au cours de laquelle le système divise le texte en mots, groupes de mots, phrases et paragraphes et détermine les attributs typographiques (ex. si on trouve *ordi-* à la fin d'une ligne et *nateur* au début de la ligne suivante, le système se demande s'il s'agit d'un seul mot — *ordinateur* — ou s'il s'agit d'un groupe de mots — *ordi - nateur*).

On appelle morphologie l'étape de consultation du dictionnaire syntaxique et des règles morphologiques de la LD afin de déterminer les attributs morphologiques des mots et des expressions idiomatiques (ex. *joue* (verbe) ou *joue* (nom)).

Le passage est l'étape d'analyse des phrases de la LD pour déterminer le rôle des mots et des groupes de mots dans chaque phrase. Il se divise en deux phases : la première est syntaxique et la seconde, sémantique (ex. *He saw the personal computer manual on the table* — est-ce que l'adjectif *personal* modifie le mot *manual* ou le mot *computer*?).

Lors de l'étape du transfert, les mots et les groupes de mots en LA sont substitués aux mots et aux groupes de mots en LD. À cette étape, le bon équivalent pour chaque mot ou groupe de mots est déterminé selon le contexte (ex. dans l'exemple précédent, *table* = un meuble ou une figure?).

Enfin, la génération est l'étape de construction d'un texte en LA, à partir des éléments produits par le transfert. Le texte en LA doit être correct sur le plan grammatical et avoir du sens.

2.1.2.4 SUSY

SUSY est un système expérimental qui possède des dictionnaires unilingues pour la LD et la LA qui contiennent l'information syntaxique et morphologique pour chaque unité lexicale. En général, chaque langue du système a des dictionnaires distincts pour l'analyse et pour la génération, bien qu'il existe certaines exceptions où les mêmes dictionnaires sont utilisés à la fois pour l'analyse et la génération.

Les dictionnaires unilingues en LD sont consultés par le système pour aider à désambiguïser les unités lexicales pour faciliter le transfert. Il y a un dictionnaire morpho-syntaxique contenant des radicaux et leur information grammaticale, ainsi qu'un dictionnaire syntaxico-sémantique contenant des marqueurs sémantiques, des routines pour la désambiguïisation, pour le traitement des expressions idiomatiques et pour la création de représentations de transfert. La désambiguïisation se fait surtout au niveau syntaxique. Par exemple, *to break* peut avoir les significations suivantes : *break* (sans objet), *break* (avec un objet non spécifié), *break (into a house)*, *break (the bank)*, etc. La composante sémantique, quant à elle, emploie des traits sémantiques comme +/- animé, +/- humain, etc.

Le dictionnaire de transfert est utilisé pour la conversion des représentations intermédiaires en LD en représentations intermédiaires équivalentes en LA.

Enfin, SUSY possède, tout comme pour la LD, deux dictionnaires unilingues en LA : un dictionnaire syntactico-sémantique contenant des routines pour convertir les représentations de transfert en unités lexicales, et un dictionnaire morpho-syntaxique contenant l'information syntaxique et flexionnelle nécessaire à la génération du texte d'arrivée.

La figure 8 (Hutchins et Somers 1992 : 193) illustre le processus de traduction de ce système, où les étapes²⁰ de consultation des dictionnaires sont clairement indiquées.

²⁰ WOBUSU (*Wörterbuchsuche* 'dictionary look-up') est l'étape durant laquelle seuls les mots marqués comme noms propres pendant la prédiction ne sont pas traités.

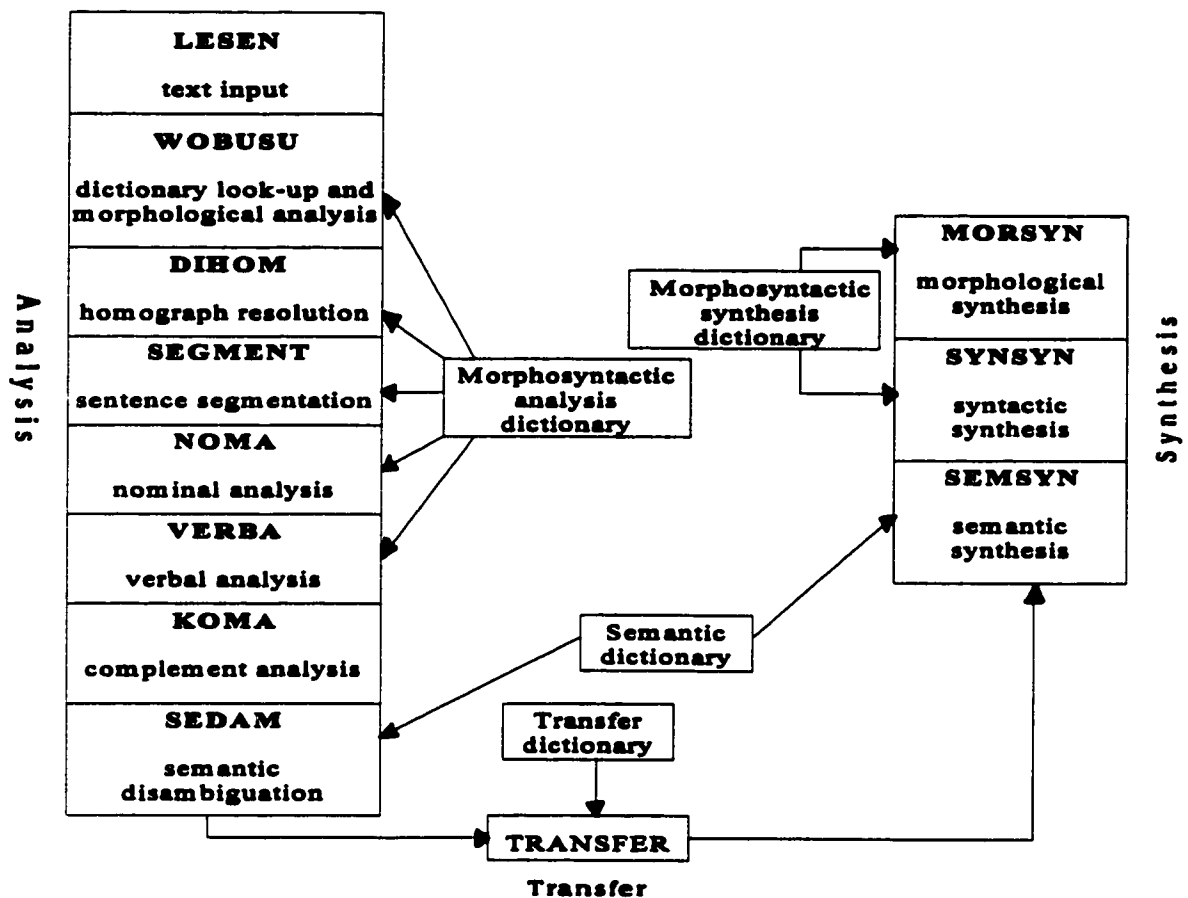


Figure 8 Processus de traduction dans SUSY

2.1.2.5 Ariane (GETA)

Ariane est un système de transfert expérimental dont l'analyse et la génération sont divisées en modules morphologiques et syntaxiques. L'étape de transfert a aussi deux phases : le transfert lexical et le transfert structurel. Ce système comprend des dictionnaires en LD, un dictionnaire bilingue et des dictionnaires en LA (Hutchins et Somers 1992 : 222).

Les dictionnaires en LD sont situés dans l'ATEF (*Analyse de textes en états finis*), dont la tâche est de convertir des séquences de texte en LD en une représentation en arbre de la LD. L'ATEF

permet à l'utilisateur de choisir un maximum de six dictionnaires plus un dictionnaire d'expressions idiomatiques. Cela comprend au moins un dictionnaire de radicaux (qui inclut les composés) et un dictionnaire d'affixes. Le nombre et l'arrangement des autres dictionnaires est laissé au choix de l'utilisateur.

Un seul dictionnaire suffit pour le transfert lexical puisque les radicaux (incluant les composés) ont été réduits à leur forme canonique et que les flexions ont été incluses comme valeurs dans les étiquettes de noeud. Le transfert structurel (qui traite des différences syntaxiques) se fait sans consultation de dictionnaires.

Enfin, un maximum de huit dictionnaires peuvent être utilisés durant la génération. Chacun de ces dictionnaires est consulté selon les valeurs des variables. Le premier dictionnaire est un dictionnaire d'unités lexicales en LA et il est consulté selon les valeurs des unités lexicales. Les autres dictionnaires peuvent être consultés en utilisant les valeurs d'autres variables. Par exemple, un dictionnaire de flexions de verbes peut être accédé par les valeurs des variables qui indiquent la personne, le nombre ou le temps.

Le processus de traduction d'Ariane se fait selon les étapes illustrées par la figure 9 (Hutchins et Somers 1992 : 223) (où les étapes de consultation des dictionnaires ne sont malheureusement pas clairement indiquées).

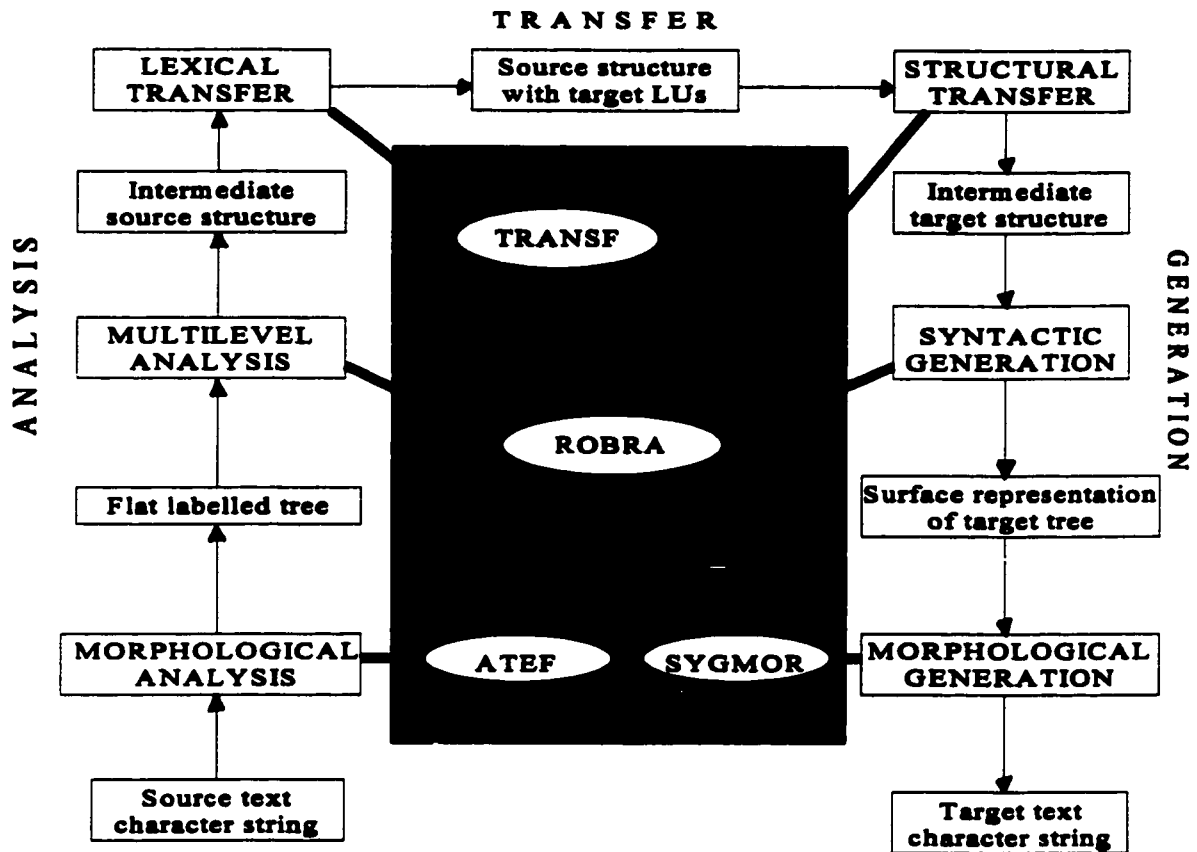


Figure 2 Configuration du système Ariane

2.1.3 Systèmes interlangues

Comme nous l'avons mentionné dans le chapitre précédent, les systèmes interlangues, tout comme les systèmes de transfert, possèdent deux types de dictionnaires : unilingues et bilingues, bien que ces derniers ne jumellent pas directement la LD et la LA. Les systèmes Rosetta et DLT reposent sur le concept d'interlangue qui veut que le transfert se fasse à l'aide d'une langue pivot (ou langue intermédiaire).

2.1.3.1 Rosetta

Rosetta est un système interlangue expérimental fondé sur la grammaire Montague et qui utilise des arbres de dérivation sémantique comme représentation interlangue. Ce système comprend plusieurs dictionnaires qui peuvent être divisés en deux types : les lexiques (qui décrivent les propriétés d'un objet) et les dictionnaires (qui indiquent les relations entre les objets de deux lexiques différents). Les lexiques et les dictionnaires peuvent être utilisés aussi bien pendant la génération que pendant l'analyse. Le système Rosetta comprend les dictionnaires et lexiques suivants : le MORPH-DICT, le S-LEX, l'IL²¹-DICT, l'IL-LEX, l>ID-DICT et le CP-DICT (Odijk 1994 : 120-122).

Le '**MORPH-DICT**' relie un radical à une clé syntaxique, qui est le nom donné à un ensemble de propriétés morphosyntaxiques. Il est possible qu'un mot corresponde à plusieurs clés syntaxiques (ex. le mot *attempt* correspond au nom *attempt* et au verbe *attempt*) ou que plusieurs mots correspondent à une seule clé (c'est le cas des variantes orthographiques). De plus, il peut y avoir des clés syntaxiques auxquelles aucun mot n'est associé, mais il ne peut pas y avoir de mots pour lesquels il n'existe pas de clé syntaxique. La relation entre les mots et les clés syntaxiques peut être influencée par les conditions contextuelles.

Le **S-LEX** indique les propriétés morphosyntaxiques pour chaque clé syntaxique. L'attribut principal est la catégorie syntaxique. D'autres propriétés morphosyntaxiques (par exemple, comment les radicaux fléchissent, etc.) sont indiquées par un ensemble fixe de paires attribut-

²¹ IL signifie 'interlingua' (interlangue).

valeur qui dépendent de la valeur de l'attribut de la catégorie. Les propriétés syntaxiques sont aussi indiquées par un nombre fixe de paires attribut-valeur qui dépendent de la catégorie syntaxique (par exemple, pour les verbes, il y a un attribut pour indiquer le nombre d'arguments et leur nature, etc.).

L'IL-DICT relie les clés syntaxiques aux clés sémantiques de la langue intermédiaire. Une clé syntaxique peut être reliée à plusieurs clés sémantiques (dans le cas de mots ambigus) et plusieurs clés syntaxiques peuvent être reliées à une seule clé sémantique (dans le cas de synonymes). Une clé sémantique donne le sens d'un mot ou d'une expression idiomatique. Ce dictionnaire comprend, pour chaque paire <clé syntaxique, clé sémantique>, une courte description de sens, qui est utilisée pour demander à l'utilisateur quel est le sens voulu lorsque le système ne peut résoudre une ambiguïté.

L'IL-LEX est conçu pour indiquer les propriétés sémantiques de chaque clé sémantique, par exemple le type sémantique du sens et les types sémantiques des arguments du sens.

L>ID-DICT est un dictionnaire d'expressions idiomatiques (ex. *to kick the bucket*, *to pull someone's leg*). Dans le système Rosetta, les expressions idiomatiques sont représentées par des arbres de dérivation, mais, dans les dictionnaires, elles sont entrées en indiquant la classe de l'expression, les clés syntaxiques des mots qui forment l'expression, dans un ordre donné, et la clé syntaxique de l'expression en entier. La classe de l'expression définit la structure syntaxique interne de l'expression, comprenant les positions des arguments libres. La classe de l'expression

et les clés (dans l'ordre) des mots de l'expression fournissent l'information nécessaire au système pour construire un arbre syntaxique de l'expression.

Enfin, le **CP-DICT** est le dictionnaire des prédicats complexes, qui sont, entre autres, des expressions telles que *to have a bath* ou *to make a decision*, c'est-à-dire des expressions constituées de certains verbes (ex. *have, get, be, make, do, etc.*) et d'objets qui déterminent le sens de ces verbes. Ce dictionnaire indique quelles paires de mots sont des quasi-expressions.

Le processus de traduction de ce système est illustré par la figure 10 (Hutchins et Somers 1992 : 284). Toutefois, cette figure ne permet pas de voir les étapes de consultation des dictionnaires.

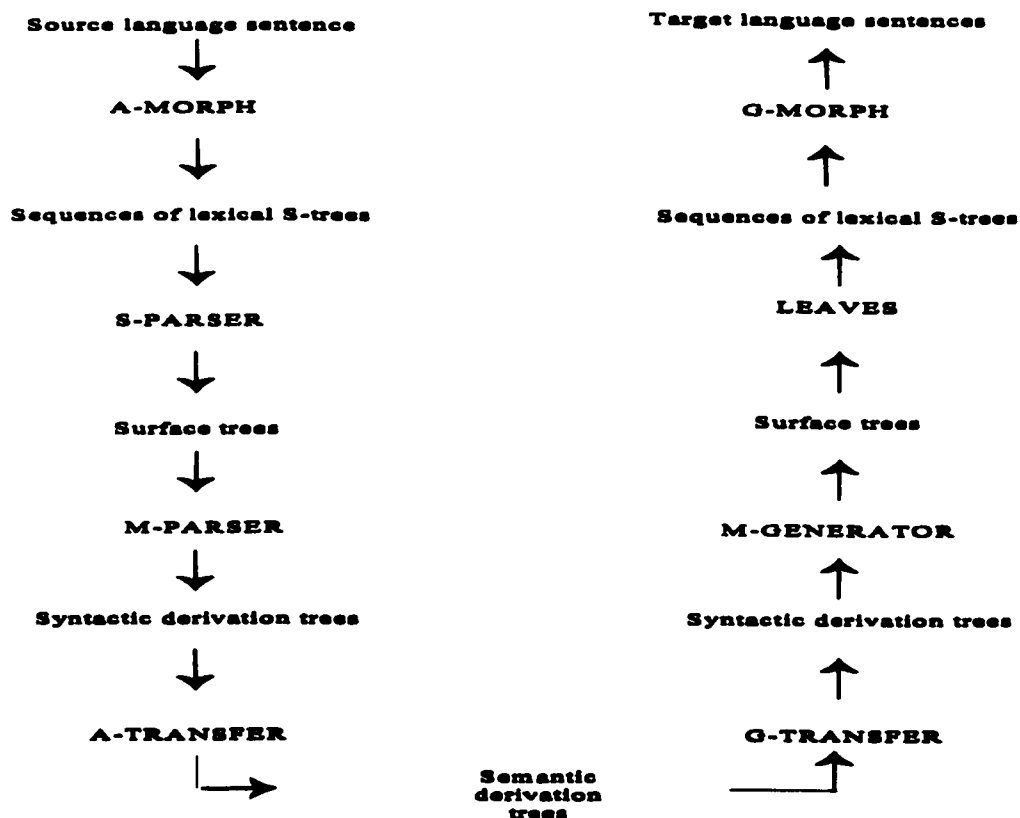


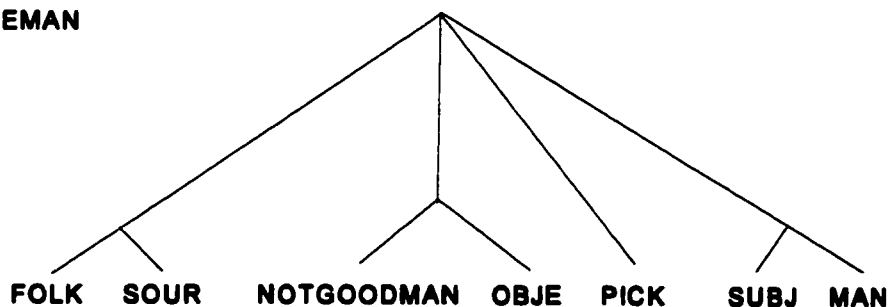
Figure 10 Processus de traduction de Rosetta

2.1.3.2 DLT

DLT ('Distributed Language Translation') est un système interlangue expérimental fondé sur l'espéranto. Puisque l'interlangue (espéranto modifié) n'est pas une représentation abstraite, mais une langue régularisée, le système est composé de deux dictionnaires bilingues : un dictionnaire LD-espéranto et un dictionnaire espéranto-LA, ainsi que d'un dictionnaire d'analyse unilingue pour la LD.

La première étape du processus de traduction consiste à désambiguïser le texte de départ. Cette désambiguïstation commence par la consultation du dictionnaire d'analyse en LD. Le dictionnaire assigne une formule, qui est une représentation en arbre correspondant au sens d'un mot, à chaque unité lexicale. Par exemple, voici la formule de *policeman* (Laffling 1991 : 42) :

POLICEMAN



(I.e. a person who selects bad persons out of the body (SOUR) of the people (FOLK))

La consultation de ce dictionnaire se fait en deux étapes. La première vise à trouver les têtes des formules (c'est-à-dire la branche qui exprime la catégorie fondamentale à laquelle la formule appartient) et à les placer dans l'ordre suivant : AGENT-ACTION-OBJET (ex. MAN FORCE MAN, MAN FORCE THING, etc.). Ensuite, des représentations en arbre sont construites en ayant recours aux formules entières applicables au mot, et non seulement aux têtes.

L'étape suivante est l'étape des transformations des arbres en LD en représentations interlangues. Ensuite, le système doit choisir, parmi les représentations possibles, l'arbre en interlangue (espéranto) qui correspond le mieux, dans le contexte donné, à celui de la LD. Cette sélection se fait à l'aide d'un dictionnaire bilingue et d'une base de données lexicales appelée 'Lexical Knowledge Bank' (LKB). Cette étape est répétée pour passer de l'espéranto à la représentation en LA et, de même que pour le transfert de la LD à l'interlangue, le transfert de l'interlangue à la LA peut générer plusieurs arbres en LA pour un seul arbre en espéranto.

Une des caractéristiques qui distingue ce système est qu'il ne fait pas usage d'un dictionnaire unilingue en LA. Pour générer le texte d'arrivée à partir de l'arbre en LA, DLT ne se sert que de règles de morphologie et de linéarisation.

Le dictionnaire d'analyse est le seul dictionnaire qui peut être mis à jour par l'utilisateur. Il est constitué d'une collection de fiches consistant en une entrée principale suivie d'une liste de sous-entrées (pouvant aller jusqu'à 39 sous-entrées) liées sémantiquement. Le programme de mise à jour demande d'abord à l'utilisateur d'entrer le mot-vedette d'une entrée principale. Ensuite, il

fouille sa base de données à la recherche de fiches déjà existantes. Si le programme n'en trouve aucune, l'utilisateur peut alors créer une nouvelle entrée. S'il existe déjà des fiches, la première que le programme repère apparaît à l'écran. L'utilisateur a alors quatre options : (1) modifier une ou plusieurs des sous-entrées, (2) effacer l'entrée lexicale au complet, (3) demander que les autres fiches de la même entrée principale (s'il en existe) apparaissent à l'écran, ou (4) demander la création d'une nouvelle entrée lexicale. Toutes les entrées principales apparaissent sous leur forme canonique. Enfin, le dictionnaire peut être imprimé au complet (en ordre alphabétique) à la demande de l'utilisateur (Laffling 1991 : 65).

Le processus de traduction de ce système ainsi que les étapes de consultation des dictionnaires sont très bien illustrés par la figure 11 (Hutchins et Somers 1992 : 302).

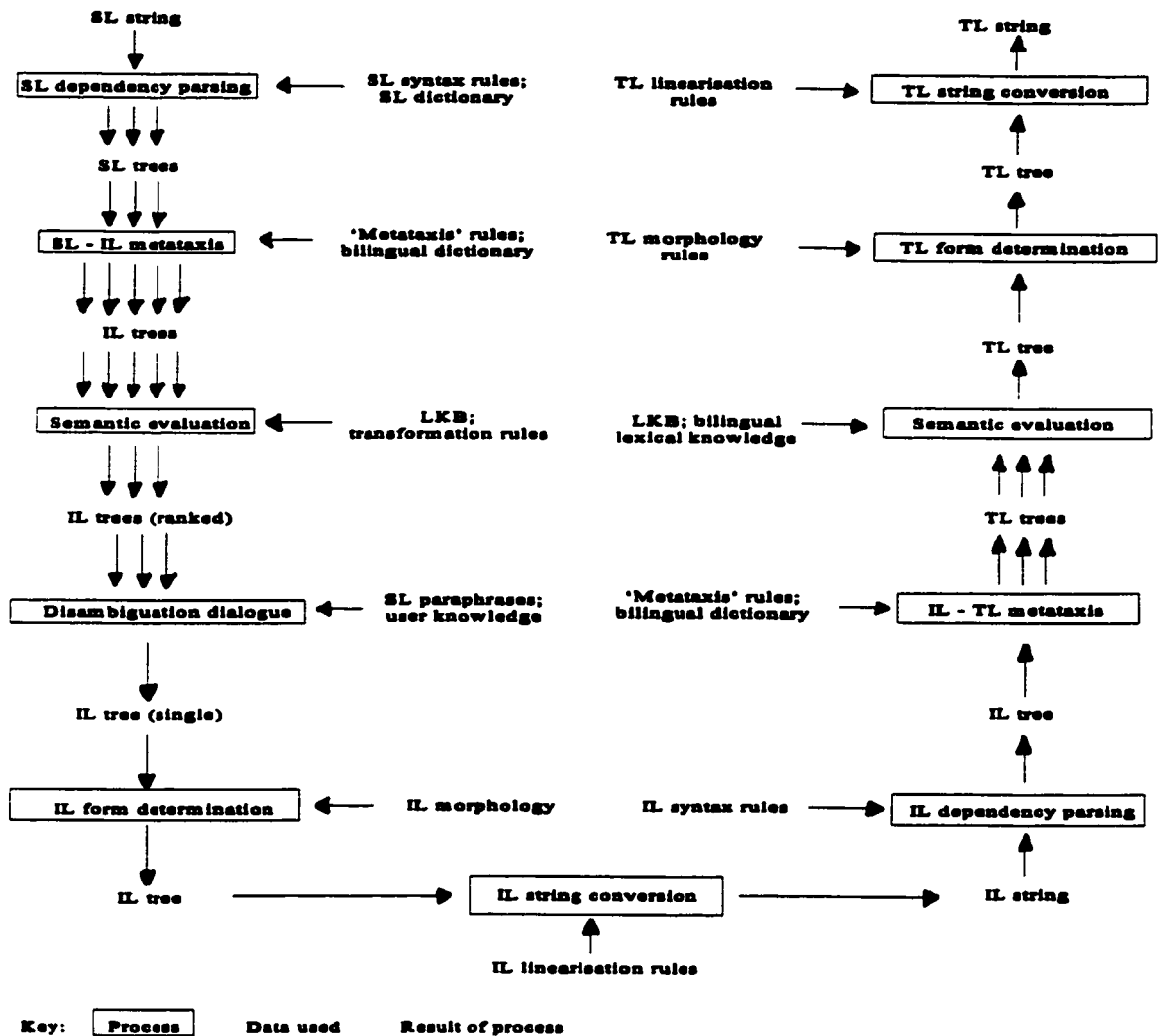


Figure 1.1 Processus de traduction de DLT

2.2 Rôle des dictionnaires dans les systèmes de TA

Les dictionnaires dans les systèmes de TA et leur fonctionnement par rapport aux autres parties des systèmes dépendent, en grande partie, du type de système dans lequel se trouvent ces dictionnaires. De plus, comme nous l'avons constaté dans la section précédente, pour un même

type de système, les dictionnaires peuvent être différents. Il est néanmoins possible de cerner le rôle général des dictionnaires (unilingues et bilingues) à l'intérieur des systèmes de TA.

Nous allons d'abord analyser le rôle des dictionnaires unilingues, puis celui des dictionnaires bilingues.

2.2.1 *Rôle des dictionnaires unilingues*

Les dictionnaires unilingues n'existent que dans les systèmes de transfert et interlangues, qui, au contraire des systèmes directs, séparent nettement les diverses étapes de la traduction. Par conséquent, leur rôle sera défini en fonction des systèmes de transfert et interlangues.

Comme nous l'avons mentionné dans le chapitre 1, les systèmes de transfert présentent trois modules différents : l'analyse, le transfert et la génération. L'analyse consiste à convertir un énoncé en LD en une représentation intermédiaire qui décrit les propriétés linguistiques de cet énoncé. Le transfert fait une traduction entre deux représentations intermédiaires, c'est-à-dire entre celle de la LD et celle d'un énoncé équivalent en LA. Enfin, la tâche de la génération consiste à convertir la représentation en LA en un énoncé en LA. Dans un tel type de système, l'analyse et la génération nécessitent de l'information morphologique, syntaxique, sémantique, etc. pour permettre le passage d'un énoncé en LD à une représentation intermédiaire en LD et d'une représentation intermédiaire en LA à un énoncé en LA. Le système aura accès à ce type d'information (morphologique, syntaxique, etc.) dans les dictionnaires unilingues (dictionnaire d'analyse et dictionnaire de génération).

L'idée sous-jacente des systèmes interlangues, dans lesquels il n'y a que deux modules (l'analyse et la génération), est d'utiliser une même représentation intermédiaire pour toutes les langues traitées par le système. Tout comme pour les systèmes de transfert, l'analyse des systèmes interlangues permet de convertir un énoncé en LD en une représentation intermédiaire; toutefois, la génération de ces systèmes vise à convertir cette même représentation en un énoncé en LA (c'est-à-dire qu'une même représentation intermédiaire peut générer des énoncés dans plusieurs langues). Malgré cette différence, les besoins lexicaux des modules d'analyse et de génération sont similaires à ceux des systèmes de transfert, c'est-à-dire que les dictionnaires unilingues doivent fournir de l'information morphologique, syntaxique, sémantique, etc. pour permettre le passage de l'énoncé en LD à la représentation intermédiaire en LD et de cette représentation intermédiaire à l'énoncé en LA.

De manière générale, les dictionnaires unilingues en LD (ou dictionnaires d'analyse) servent à l'analyse structurelle et à la désambiguïsation du texte de départ. Dans ces dictionnaires, les homographes et les mots polysémiques font souvent l'objet d'entrées séparées pour chaque sens.

Les dictionnaires unilingues en LA (ou dictionnaires de génération), quant à eux, sont souvent moins détaillés que les dictionnaires d'analyse puisque l'étape de la désambiguïsation n'est pas présente lors de la génération. En fait, selon Pierre Isabelle (communication personnelle), la plupart des dictionnaires de génération, sinon tous, contiennent toute l'information requise pour la flexion des mots (ex. classe de conjugaison des verbes). De plus, un bon dictionnaire de génération devrait contenir, par exemple, des informations sur les prépositions fortement régies

par un verbe (ex. le verbe français *dépendre* prend un complément en *de*) ou des informations sur la sous-catégorisation des verbes (ex. le verbe *savoir* prend un complément phrastique — *savoir que* + *PHRASE* —, alors que le verbe *connaître* prend plutôt un groupe nominal). Toujours selon Pierre Isabelle (communication personnelle), beaucoup de systèmes traitent le deuxième exemple (sous-catégorisation des verbes) dans le dictionnaire de transfert plutôt que dans le dictionnaire de génération. Pour lui, cette pratique, qui fait que le dictionnaire de génération est moins détaillé que le dictionnaire d'analyse, est injustifiée, car il ne voit aucune raison pour que les dictionnaires de génération contiennent des renseignements différents des dictionnaires d'analyse. En fait, beaucoup de systèmes traditionnels demandent au transfert d'exécuter des tâches qui devraient, en principe, relever du processus de génération.

2.2.2 Rôle des dictionnaires bilingues

Ce qui distingue la traduction automatique des autres applications de l'analyse et de la génération automatiques, c'est l'aspect contrastif. La contrastivité, c'est-à-dire la comparaison linguistique de deux langues, n'est pas nécessaire dans d'autres applications, même pas dans des applications «multilingues», comme l'interrogation de bases de données en plusieurs langues naturelles ou la génération de textes multilingues. Les dictionnaires (et les grammaires) contrastifs sont donc en quelque sorte les modules les plus spécifiques et les plus caractéristiques parmi les sources de connaissances linguistiques des systèmes de TA (Heid 1993 : 170).

Les dictionnaires bilingues se trouvent dans tous les types de système de TA. Dans les systèmes directs, les équivalences traductiques se font plus ou moins mot à mot. Ces systèmes ne contiennent que des dictionnaires bilingues combinant des unités lexicales en LD avec leurs

équivalents en LA. En général, de tels dictionnaires sont très complexes, car ils ne comprennent qu'une seule liste de mots avec toutes les informations nécessaires à la traduction.

Dans les systèmes de transfert, qui comprennent aussi des dictionnaires unilingues, les dictionnaires bilingues sont habituellement plus simples que les dictionnaires unilingues, car ils ne servent qu'à donner les correspondances lexicales en LA, avec un minimum d'information grammaticale semblable à l'information donnée dans les dictionnaires unilingues (d'analyse et de génération). Heid, dans son article «Le lexique : Quelques problèmes de description et de représentation lexicale», résume bien le rôle du dictionnaire bilingue dans un système de transfert :

Le dictionnaire bilingue part d'un lemme de la langue de départ, en donne une analyse sémantique (et/ou pragmatique, communicative, etc., selon le cas), met en parallèle les résultats de l'analyse sémantique avec une description sémantique en langue cible et donne enfin la forme des lexèmes de la langue cible qui correspondent à la description sémantique en question (1993 : 171).

Comme nous l'avons constaté précédemment, les systèmes interlangues tendent vers l'utilisation d'une même représentation intermédiaire pour toutes les langues traitées par le système. Par conséquent, les dictionnaires «bilingues» de ce type de système vont de la LD à l'interlangue et de l'interlangue à la LA. Comme pour les systèmes de transfert, les dictionnaires bilingues des systèmes interlangues semblent généralement plus simples que les dictionnaires unilingues, car ils ne servent qu'à donner la correspondance entre les composantes de l'énoncé en LD et l'interlangue, d'une part, et entre l'interlangue et les composantes de l'énoncé en LA, d'autre part. Il est toutefois difficile de généraliser parce que les dictionnaires bilingues varient selon l'interlangue utilisée.

2.3 Conclusion

Dans l'ensemble, les dictionnaires unilingues, que l'on retrouve dans les systèmes de transfert et interlangues, contiennent plus d'information lexicale que les dictionnaires bilingues de ces mêmes types de systèmes. Dans ces systèmes, le rôle des dictionnaires bilingues consiste principalement à convertir une représentation intermédiaire en LD en une représentation intermédiaire en LA pour les systèmes de transfert, et à convertir un énoncé en LD en une représentation interlangue et une représentation interlangue en un énoncé en LA pour les systèmes interlangues.

Au contraire des systèmes de transfert et interlangues, les systèmes directs possèdent des dictionnaires bilingues très complexes, car ces dictionnaires doivent fournir toutes les informations nécessaires à la traduction pour compenser l'absence de dictionnaires unilingues. Toutefois, l'analyse linguistique est beaucoup moins élaborée que celle que l'on retrouve dans les dictionnaires unilingues des systèmes de transfert et interlangues.

Chapitre 3 : LES ENTRÉES DE DICTIONNAIRE BILINGUE AUTOMATIQUE

3.0 Introduction

Après avoir présenté un certain nombre de systèmes de TA dans le chapitre 2 et avoir cerné le rôle des dictionnaires unilingues et bilingues dans les systèmes de TA en général, nous allons maintenant aborder une section un peu plus pratique : l'examen, la modification et la préparation d'entrées de dictionnaire bilingue dans deux systèmes différents, soit Globalink et Logos. Le but de cet exercice est de voir de façon concrète les éléments linguistiques fournis par les dictionnaires d'un système donné, ceux qui sont exigés pour créer une entrée dictionnaire et ceux qu'on peut y ajouter si on le juge nécessaire. Ce chapitre se relie donc à notre premier objectif qui est celui d'examiner les dictionnaires dans un certain nombre de systèmes de TA.

Il est important de noter que l'utilisateur peut avoir accès aux entrées du dictionnaire de base de Globalink (version fonctionnant dans un environnement Windows 95) alors qu'il ne peut absolument pas consulter celles du dictionnaire de base de Logos. Cependant, à l'aide d'Alex (outil lexicographique offert à l'utilisateur), nous pouvons créer de nouvelles entrées pour des mots non existants dans le dictionnaire de base de Logos.

3.1 Méthodologie

Au lieu d'examiner, de modifier ou de préparer des entrées choisies au hasard, nous avons décidé de fonder notre étude lexicographique sur un texte qui est soumis à la traduction automatique. Après l'évaluation de la traduction de ce texte, nous avons choisi, selon des critères variés, un certain nombre d'entrées dictionnaires à examiner, à modifier ou à ajouter. Notre analyse ne

portera pas sur la traduction entière du texte, mais sur les équivalents lexicaux choisis par les systèmes. Nous avons donc choisi comme texte de départ un extrait d'un texte économique anglais tiré d'un article de presse (août ou septembre 1995), qui est relativement simple du point de vue syntaxique, mais qui est intéressant du point de vue lexical, car il contient des termes spécialisés (ex. *excess capacity*), des collocations (ex. *high + unemployment*), ainsi que plusieurs mots polysémiques (ex. *potential*).

3.1.1 *Texte pour Globalink et Logos*

ECONOMIC PLANNING ASSUMPTIONS

Introduction

For several years, the Canadian economy has been performing below potential.

Until recently interest rates have been high, the recovery from the recession has been painfully slow and unemployment remains unacceptably high.

This poor economic performance has created substantial excess capacity in output and labour markets. This, combined with good productivity performance, has pushed inflation and cost pressures down to the lowest levels in a generation.

3.2 Globalink

3.2.1 Traduction du texte en LD avec les dictionnaires du système

Texte anglais

ECONOMIC PLANNING ASSUMPTIONS

Introduction

For several years, the Canadian economy has been performing below potential. Until recently interest rates have been high, the recovery from the recession has been painfully slow and unemployment remains unacceptably high.

This poor economic performance has created substantial excess capacity in output and labour markets. This, combined with good productivity performance, has pushed inflation and cost pressures down to the lowest levels in a generation.

Traduction française

les suppositions de l'organisation économiques

La présentation

Pour les plusieurs années, l'économie Canadienne a exécuté en dessous capacité. Jusqu'à récemment taux de l'intérêt a été haut, la récupération de la récession a été lente douloureusement et le chômage reste haut inadmissible.

Cette performance économique pauvre a créé la capacité de l'excès substantielle dans production et marchés du labour. Cela, combiné avec bonne performance de la productivité, a poussé inflation et pressions du coût jusqu'à les niveaux les plus bas dans une génération.

3.2.2 Entrées à examiner

Après avoir évalué cette traduction, nous avons sélectionné cinq unités lexicales dans le texte de départ dont nous avons voulu examiner les entrées dans le dictionnaire de base : *Canadian*, *potential*, *interest rate*, *performance* et *labour market*. Les entrées *Canadian*, *interest rate*, *performance* et *labour market* ont été choisies à cause des erreurs de traduction; celle de *potential*, à cause d'une bonne traduction.

Étant donné que le mot-vedette principal dans le dictionnaire de Globalink ne peut être qu'un mot simple et que les unités lexicales de plus d'un mot comme *interest rate* et *labour market* ne peuvent figurer que comme locutions dans l'entrée du mot simple, nous avons dû aller aux entrées *interest* et *labour* pour trouver les composés qui nous intéressaient. Par ailleurs, n'ayant pu trouver d'entrée pour *labour*, nous avons cherché celle pour sa variante orthographique *labor*.

Un examen préliminaire des entrées de Globalink pour les mots choisis a confirmé leur intérêt parce qu'elles sont différentes les unes des autres : *Canadian* est présenté à la fois comme adjectif et comme nom, avec un seul équivalent pour chacun; *potential* est aussi présenté comme adjectif et comme nom, mais avec deux équivalents proposés pour le nom; *interest*, qui est également présenté comme nom et comme verbe, contient des locutions; *performance*, qui ne peut être qu'un nom, est traduit par cinq équivalents que l'utilisateur peut réordonner; enfin, comme nous l'avons déjà signalé, *labour* n'apparaît pas dans le dictionnaire de base et nécessite la création d'une nouvelle entrée.

3.2.2.1 Canadian

Il y a une seule entrée pour le mot *Canadian* comme adjectif et comme nom. Selon l'information contenue dans l'entrée déjà existante du dictionnaire de base (figure 12), l'équivalent s'écrit avec une minuscule lorsqu'il est adjectif (*canadien*) et avec une majuscule lorsqu'il est nom (*Canadien*). Donc, l'erreur typographique *Canadienne*, dans le texte traduit, ne semble pas découler du dictionnaire.

Rechercher: canadian	Mots entiers: Canada Canadian canoe canalization canalize
---	--

☐ **Canadian**

- ◆ **Adjectif**
 -
- ◆ **Nom**
 -

Figure 12 Entrée *Canadian* dans Globalink

3.2.2.2 potential

Comme dans le cas de *Canadian*, le mot *potential*, qui peut être un adjectif ou un nom, n'a qu'une seule entrée dans le dictionnaire de Globalink (figure 13). Lorsqu'il est un nom, il peut se traduire, selon cette entrée, de deux façons. Par défaut, le système va toujours chercher le premier équivalent qui apparaît dans le dictionnaire²². Il faut noter, cependant, qu'il est possible de changer l'ordre des équivalents de façon à ce que l'équivalent que l'utilisateur préfère soit celui donné par défaut lors des traductions.

Rechercher: potential	Mots entiers: potestate potential potentially potentially potul
--	--

☐ **potential**

- ◆ **Adjectif**
 -
- ◆ **Nom**
 -
 -

Figure 13 Entrée *potential* dans Globalink

²²

C'est le cas pour la plupart des mots. Cependant, pour les verbes, il est possible de leur donner des attributs (verbe pronominal, verbe suivi d'un objet direct ou verbe suivi d'un objet indirect) qui seront reconnus par le système qui sélectionnera alors l'équivalent correspondant à ces attributs.

3.2.2.3 interest

L'entrée pour le mot *interest* contient les formes nominale et verbale, ainsi que certaines locutions (figure 14). Dans le cas qui nous intéresse, il est à noter qu'*interest rate* ne fait pas partie des locutions comprises dans l'entrée *interest*. Nous l'ajouterons par la suite (voir section 3.2.3.1).

The screenshot shows a software interface with a search bar at the top left containing the word "interest". To its right is a list of "Mots entiers:" (whole words) including "interest", "interest-bearing", "interest-free", and "interestingly". Below the search bar, the results for "interest" are displayed in a tree-like structure. Under the "Nom:" (Noun) section, "intérêt" is listed. Under the "Verbe:" (Verb) section, "intéresser", "s'intéresser", "concerner", and "éveiller l'intérêt de" are listed. Under the "Locutions:" (Phrases) section, "vested interest : capital" invest" and "controlling interest : participation" majoritaire" are listed.

Figure 14 Entrée *interest* dans Globalink

3.2.2.4 performance

Le mot *performance* est un nom auquel le dictionnaire de base de Globalink a attribué cinq équivalents (figure 15). Comme nous l'avons indiqué précédemment, par défaut, le système ira toujours chercher le premier équivalent qui apparaît dans le dictionnaire. Toutefois, l'utilisateur peut changer l'ordre des équivalents (ce que nous ferons dans la section 3.2.3.2).

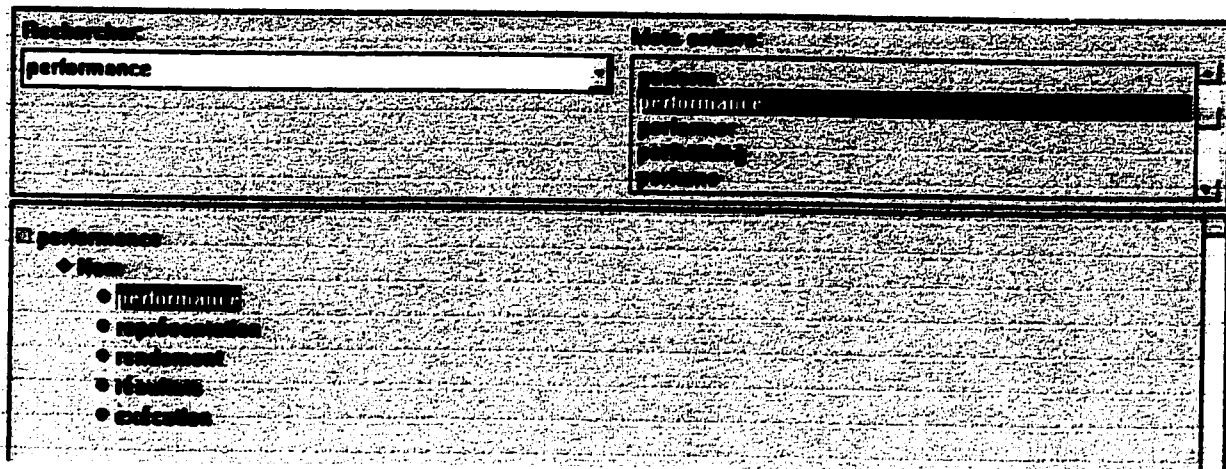


Figure 15 Entrée *performance* dans Globalink

3.2.2.5 labor

Puisque l'entrée *labour* ne se trouve pas dans le dictionnaire de base de Globalink, nous avons choisi d'examiner l'entrée de sa variante orthographique *labor* pour avoir une idée des informations minimales à inclure dans l'entrée *labour* lorsque nous la créerons pour y mettre la locution *labour market* (voir section 3.2.3.3). La figure 16 montre que l'entrée *labor* présente le nom et le verbe, ainsi que quatre locutions nominales (dont *labor market* ne fait pas partie).

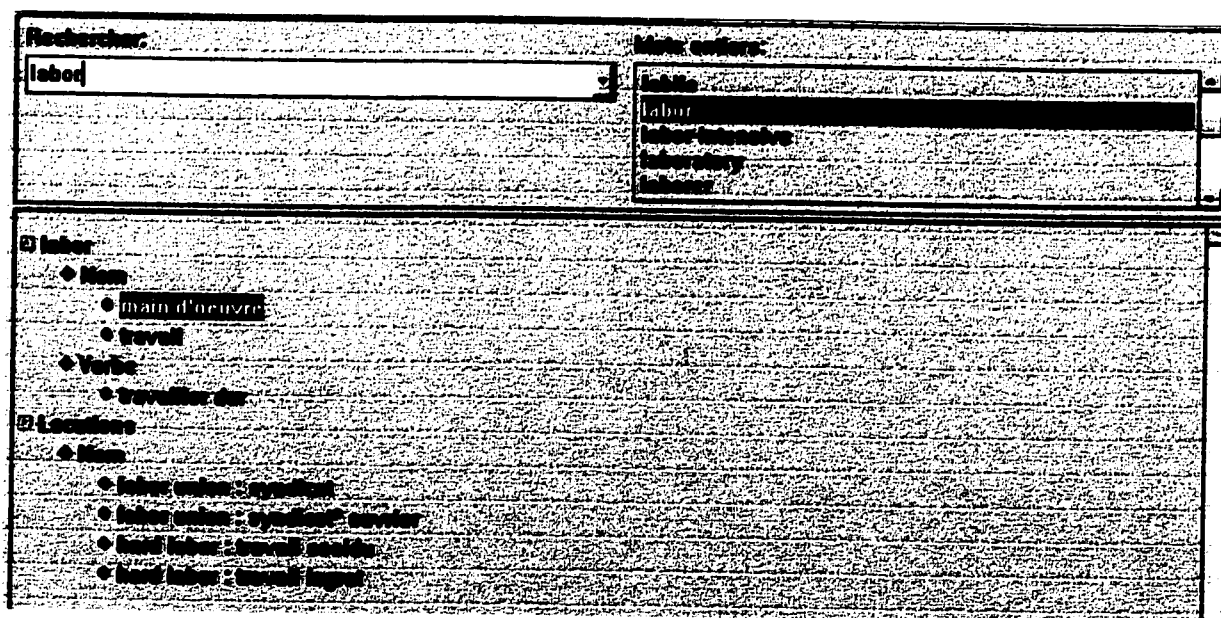


Figure 16 Entrée *labor* dans Globalink

3.2.3 Entrées modifiées ou préparées

Comme il a été mentionné précédemment, nous avons modifié les entrées *interest* et *performance* et nous avons créé une nouvelle entrée pour *labour* sous laquelle nous avons placé la locution *labour market*.

3.2.3.1 interest rate

Sous l'entrée *interest*, nous avons ajouté la locution *interest rate*, et son équivalent, dans la section «locutions» (figure 17). Les astérisques indiquent les mots qui peuvent fléchir, par exemple, les noms qui s'accordent au pluriel et les verbes qui se conjuguent. Lorsque l'utilisateur modifie une entrée déjà existante ou crée une nouvelle entrée, il ne doit pas oublier d'indiquer les mots qui peuvent fléchir afin d'améliorer les traductions de ces mots.

The screenshot displays the Globalink software interface. At the top, there is a search bar labeled 'Rechercher:' containing the text 'Interest'. To the right, a list of suggestions is shown under the heading 'Mots entrés:', including 'interest', 'interest bearing', 'interest free', and 'interestingly'. Below the search bar are two buttons: 'Accepter' and 'Rejeter'. The main area of the interface shows a detailed view of the 'Interest' entry. It includes a section for 'Nom' (Noun) with sub-entries: 'intérêt', 'interest rate* : taux* d'intérêt', 'vested interest* : capital* investi*', and 'controlling interest* : participation* majoritaire*'. There is also a section for 'Verbe' (Verb) with sub-entries: 'intéresser', 's'intéresser', 'concerner', and 'avoir une part d'intérêt de'. The interface also features a 'Change' button and a 'GEN' button.

Figure 17 Entrée modifiée *interest* dans Globalink

3.2.3.2 performance

La modification de l'entrée *performance* ne consiste qu'à changer l'ordre des équivalents de manière à placer l'équivalent voulu (*rendement*) en premier. Les flèches allant vers le haut et vers le bas, placées devant l'équivalent, servent à le déplacer (figure 18).

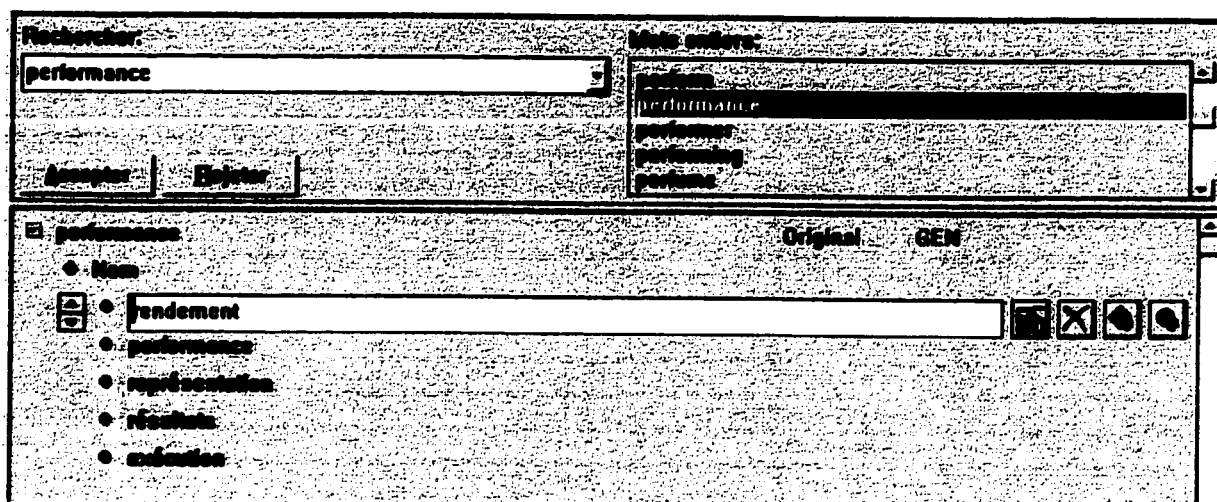


Figure 18 Entrée modifiée *performance* dans Globalink

3.2.3.3 labour market

Lors de la création d'une nouvelle entrée, le système demande d'abord à l'utilisateur d'indiquer dans quel dictionnaire ajouter l'entrée (dictionnaire général de base ou dictionnaire spécialisé créé par l'utilisateur, qui s'ajoute au dictionnaire de base) (figure 19).

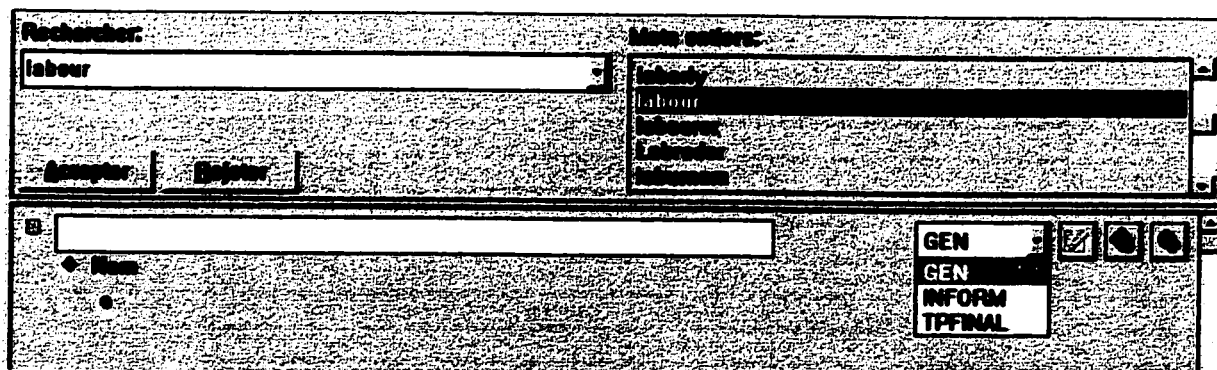


Figure 19 Sélection du dictionnaire dans Globalink

Puis le système demande qu'on lui donne de l'information sur les flexions (figure 20) et sur la catégorie grammaticale du mot-vedette.

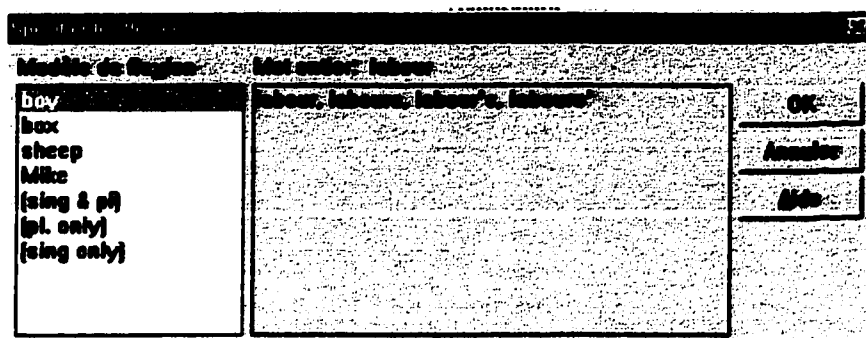


Figure 20 Information sur les flexions dans Globalink

L'utilisateur peut aussi spécifier les attributs du mot-vedette (figure 21) pour aider au traitement de ce mot lors de la traduction. Cependant, cette spécification est optionnelle et nous n'avons donc rien précisé pour *labour*.

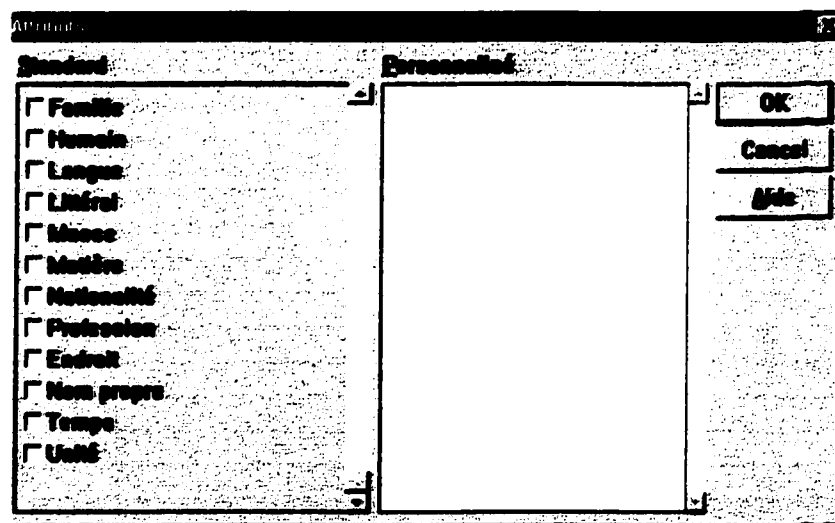


Figure 21 Sélection des attributs des noms dans Globalink

Enfin, l'utilisateur doit entrer le (ou les) équivalent(s) (figure 22). Il faut noter que les informations sur le genre et le nombre des équivalents français ne peuvent être entrées que dans l'entrée correspondante du dictionnaire français-anglais et non dans le dictionnaire anglais-français. L'utilisateur a cependant la possibilité d'indiquer les attributs des équivalents (comme pour les mots-vedettes).

Une fois que le mot-vedette et son (ou ses) équivalent(s) sont entrés avec leurs informations (flexions et attributs), l'utilisateur peut ajouter différentes locutions, comme *labour market*, avec leurs équivalents respectifs (figure 22).

The screenshot displays the Globalink interface for entering a word entry. At the top, there are two input fields: 'Rechercher:' (Search) containing 'labour' and 'Mot: entrée:' (Word: entry) containing 'labour'. Below these, a list of suggestions is shown: 'labour', 'labour', 'labour', 'labour', and 'labour'. The main area of the interface is divided into two sections. The first section, titled 'labour', has a 'Nom' (Name) field containing 'travail'. The second section, titled 'Locutions' (Collocations), has a 'Nom' field containing 'labour market : marché du travail'.

Figure 22 Entrée préparée pour *labour* dans Globalink

3.2.4 Traduction du texte en LD avec les modifications apportées au dictionnaire

Après ces quelques modifications et ajouts, nous avons encore soumis le texte de départ à la traduction avec les résultats qui suivent.

Texte anglais

ECONOMIC PLANNING ASSUMPTIONS

Introduction

For several years, the Canadian economy has been performing below potential. Until recently interest rates have been high, the recovery from the recession has been painfully slow and unemployment remains unacceptably high.

This poor economic performance has created substantial excess capacity in output and labour markets. This, combined with good productivity performance, has pushed inflation and cost pressures down to the lowest levels in a generation.

Traduction française

les suppositions de l'organisation économiques

La présentation

Pour les plusieurs années, l'économie Canadienne a exécuté en dessous capacité. Jusqu'à récemment taux d'intérêt a été haut, la récupération de la récession a été lente douloureusement et le chômage reste haut inadmissible.

Ce rendement économique pauvre a créé la capacité de l'excès substantielle dans production et marchés du travail. Cela, combiné avec bon rendement de la productivité, a poussé inflation et pressions du coût jusqu'à les niveaux les plus bas dans une génération.

3.2.5 Conclusions sur le dictionnaire de Globalink

Les modifications apportées aux entrées du dictionnaire de base améliorent la traduction dans le sens où l'utilisateur peut choisir les équivalents qu'il préfère. De plus, puisque le système reconnaît l'équivalent *marché du travail* comme unité lexicale dans laquelle *marché* peut

s'accorder en nombre (marqué par l'astérisque), la traduction produite est nettement meilleure que celle du début (*marché du labour*).

Cependant, le fait que l'utilisateur doive toujours aller dans le dictionnaire français-anglais lorsqu'il veut donner de l'information sur le genre et le nombre de l'équivalent français (dans le dictionnaire anglais-français) est fastidieux, de même qu'avoir à sélectionner l'ordre des équivalents afin d'obtenir l'équivalent voulu.

3.3 Logos

Logos est un système de TA qui ne permet pas à l'utilisateur d'accéder directement aux entrées du dictionnaire de base et encore moins de les modifier. Toutefois, il offre à l'utilisateur un outil, nommé Alex, servant à la création de dictionnaires spécifiques à une entreprise ou à un utilisateur donné. Même si Alex interdit parfois la création d'entrées pour des mots réservés, en général, le processus de création d'entrées se déroule assez bien. Il est important d'indiquer qu'Alex ne permet à l'utilisateur de créer que les dictionnaires suivants : anglais-français, anglais-allemand, anglais-italien et anglais-espagnol.

Avant de procéder à la création d'entrées, cependant, nous avons soumis le texte de départ à la traduction automatique.

3.3.1 Traduction du texte en LD avec les dictionnaires du système

Texte anglais

ECONOMIC PLANNING ASSUMPTIONS

Introduction

For several years, the Canadian economy has been performing below potential. Until recently interest rates have been high, the recovery from the recession has been painfully slow and unemployment remains unacceptably high.

This poor economic performance has created substantial excess capacity in output and labour markets. This, combined with good productivity performance, has pushed inflation and cost pressures down to the lowest levels in a generation.

Traduction française

SUPPOSITIONS DE PRÉVISION ÉCONOMIQUES

Introduction

Pendant plusieurs années, l'économie canadienne a accompli au-dessous du potentiel. Jusqu'à ce que récemment les taux d'intérêt aient été hauts, la reprise de la récession a resté de manière inadmissible haute lente et de chômage.

Ce rendement économique pauvre a créé la capacité d'excès substantielle dans la sortie et les marchés du travail. Ceci, avec avec le bon rendement de productivité, a fait baisser les pressions d'inflation et de coût aux niveaux les plus bas dans une génération.

3.3.2 Entrées préparées

Les collocations ou les mots juxtaposés en anglais ne sont pas faciles à traduire pour un système de TA. C'est pourquoi nous avons choisi de préparer des entrées pour les groupes de mots *excess capacity* et *cost pressures* qui ont été mal traduits par Logos et qui ne font évidemment pas partie du dictionnaire de base du système. Contrairement à Globalink, qui ne permet d'entrer des groupes de mots que comme locutions sous un mot-vedette d'un seul mot, Alex permet de préparer des entrées dont le mot-vedette est composé de plusieurs mots.

Lorsqu'il crée une nouvelle entrée, l'utilisateur est invité à entrer le mot-vedette (si ce dernier compte plus d'un mot, le mot principal doit être surligné), son équivalent (un seul est permis dans Alex), le nombre du mot-vedette (singulier seulement, pluriel seulement ou singulier et pluriel), la catégorie sous laquelle il est classé, ainsi que le genre (masculin, féminin ou neutre²³) et le nombre (mêmes possibilités que pour le mot-vedette) de l'équivalent.

3.3.2.1 excess capacity

Nous avons procédé ainsi à la création de l'entrée pour *excess capacity*. Étant donné qu'il s'agit d'un groupe de mots, il fallait surligner le mot principal. En fait, le système surligne déjà un mot principal (généralement le premier), mais l'utilisateur a la possibilité de changer le mot principal. Dans ce cas, nous avons retenu comme mot principal *excess* qui était déjà surligné par le système, parce que dans l'équivalent français, le mot important est *excédent* (figure 23).

Figure 23 Entrée préparée *excess capacity* dans Logos

²³ Ce genre ne s'applique que pour l'allemand.

Comme nous l'avons mentionné précédemment, l'utilisateur doit obligatoirement donner de l'information pour la catégorie sous laquelle est classé le mot-vedette (du moins, le mot principal). La figure 24 illustre le choix de catégories pour le mot principal *excess*.

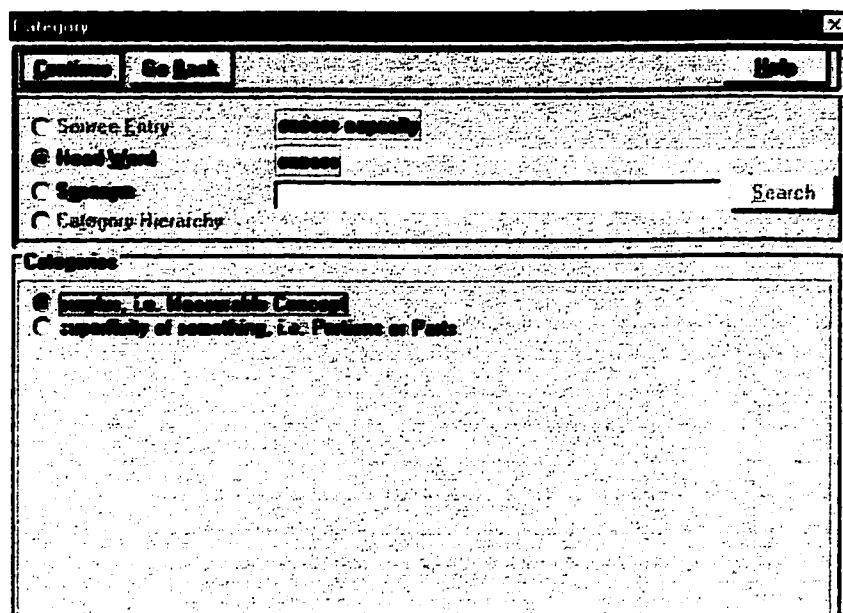


Figure 24 Sélection de la catégorie pour *excess capacity* dans Logos

Enfin, lorsque l'équivalent a plus d'un mot et que le mot principal est surligné, comme pour le mot-vedette, l'utilisateur est invité à fournir de l'information (catégorie grammaticale, genre et nombre, etc.) sur les autres mots de l'équivalent (voir figure 25).

Target Phrase Information

Continue Go Back Help

Click on each highlighted word and set its attributes.

Transfer:

excédant de capacité

Part of Speech:	Noun
Gender:	Feminine
Number:	Singular

Figure 25 Information sur *capacité* dans Logos

3.3.2.2 cost pressures

La figure 26 illustre l'entrée que nous avons préparée pour le groupe de mots *cost pressures*.

Dans ce cas, nous avons considéré *pressure* comme le mot principal parce que dans l'équivalent français *tension de coûts*, le mot important est *tension*. Étant donné que le système avait surligné *cost* et *coûts* comme mots principaux, il a fallu corriger le surlignement dans ces cas. Les étapes concernant le choix de la catégorie et l'ajout de l'information sur les autres mots accompagnant le mot principal de l'équivalent ont été les mêmes que pour *excess capacity* (voir les figures 24 et 25).

File Edit Window Help

Add Delete Previous Next

Source - English: cost pressure
Target - French: tension de coûts

Source	Target
Entry Type: Word/Phrase	Entry Type: Word/Phrase
Number: Singular & Plural	Number: Singular & Plural
Category: Mass	Gender: Feminine
	Inflection: Variable

Alternate:

Part of Speech: Noun
Subject Matter Dictionary: General
Company Dictionary: ISA

OK Cancel

Figure 26 Entrée préparée *cost pressure* dans Logos

3.3.3 Traduction du texte en LD avec le dictionnaire d'utilisateur

Après ces deux ajouts au dictionnaire, le système a produit la traduction qui suit.

Texte anglais

ECONOMIC PLANNING ASSUMPTIONS

Introduction

For several years, the Canadian economy has been performing below potential. Until recently interest rates have been high, the recovery from the recession has been painfully slow and unemployment remains unacceptably high.

This poor economic performance has created substantial excess capacity in output and labour markets. This, combined with good productivity performance, has pushed inflation and cost pressures down to the lowest levels in a generation.

Traduction française

SUPPOSITIONS DE PRÉVISION ÉCONOMIQUES

Introduction

Pendant plusieurs années, l'économie canadienne a accompli au-dessous du potentiel. Jusqu'à ce que récemment les taux d'intérêt aient été hauts, la reprise de la récession a resté été de manière inadmissible haute lente et de chômage.

Ce rendement économique pauvre a créé l'excédent de capacité substantiel dans la sortie et les marchés du travail. Ceci, avec avec le bon rendement de productivité, a fait baisser l'inflation et les tensions de coûts aux niveaux les plus bas dans une génération.

3.3.4 Conclusions sur le dictionnaire de Logos

Même si Alex (téléchargé de l'Internet) est un bon outil lexicographique permettant d'ajouter des entrées qui améliorent la traduction, il est limité quant aux types de mots (catégories grammaticales) qui peuvent être entrés dans le dictionnaire d'utilisateur, c'est-à-dire qu'il ne permet d'entrer que des substantifs. D'autre part, mais à un degré moindre, Alex est aussi limité au point de vue des paires de langues possibles. En fait, il ne permet de créer des dictionnaires que de l'anglais vers le français, l'espagnol, l'italien et l'allemand.

3.4 Conclusion

Bien que Globalink et Logos n'appartiennent pas au même type de système (Globalink étant un système direct et Logos, un système de transfert), ils possèdent néanmoins certains points en commun.

Premièrement, ils offrent tous les deux la possibilité de créer de nouvelles entrées. Dans le cas de Globalink, ces entrées peuvent être ajoutées au dictionnaire de base ou tout simplement faire partie d'un nouveau dictionnaire alors que dans le cas de Logos, il n'est possible que de créer de nouveaux dictionnaires.

Ensuite, ces deux systèmes permettent à l'utilisateur (ou lexicographe) de donner des attributs (propriétés sémantiques) aux unités lexicales entrées dans les dictionnaires. Cependant, alors que le choix d'une catégorie sémantique est obligatoire dans Logos (sinon le système ne permet pas de

sauvegarder l'entrée), il ne l'est pas dans Globalink. En fait, dans Globalink, il semblerait qu'il faille modifier les règles (ou du moins les activer) pour que les attributs donnés aux unités lexicales par l'utilisateur soient pris en considération durant la traduction. Dans Logos, il n'est pas nécessaire de toucher aux règles pour que le système tienne compte de la catégorie attribuée à l'unité lexicale.

Même si Globalink et Logos ont quelques points communs, leurs différences sont bien plus frappantes. Les plus évidentes sont l'accès (ou non) au dictionnaire de base et la possibilité de ne donner qu'un seul équivalent (pour Globalink) ou plus d'un équivalent (pour Logos) avec l'information s'y rapportant (genre, nombre).

Même si Logos est plus limité que Globalink en ce qui a trait aux catégories de mots que l'utilisateur peut entrer dans le dictionnaire, ce système permet de préparer des entrées séparées pour des unités lexicales complexes (ex. *excess capacity*), ce qui est impossible dans Globalink (où on ne peut les entrer que comme locutions). De plus, la section 'catégorie' d'Alex offre davantage de choix à l'utilisateur que la section 'attribut' de Globalink.

Dans l'ensemble, Alex offre la possibilité d'entrer un peu plus d'éléments linguistiques que Globalink. Toutefois, pour être vraiment utile au système, le dictionnaire devrait servir à mieux traduire et, malheureusement, cela dépend surtout des règles associées au dictionnaire.

Chapitre 4 : COMPARAISON ENTRE LES DICTIONNAIRES BILINGUES TRADITIONNELS ET LES DICTIONNAIRES BILINGUES AUTOMATIQUES

4.0 Introduction

Le présent chapitre vise à examiner les constituants des dictionnaires bilingues traditionnels d'une part et ceux des dictionnaires automatiques d'autre part afin de comparer les deux types de dictionnaires bilingues et de voir ce qu'il faut pour adapter une entrée de dictionnaire bilingue traditionnel pour un dictionnaire bilingue automatique.

Étant donné que l'entrée du dictionnaire automatique reste la même pour les mots généraux et pour les mots spécialisés, nous avons limité l'étude des constituants d'un dictionnaire bilingue traditionnel à celle d'un dictionnaire général.

4.1 Constituants des dictionnaires bilingues traditionnels

Même si les dictionnaires bilingues de poche se limitent à ne présenter que les mots-vedettes et leurs différents équivalents accompagnés seulement de leur catégorie grammaticale (et, dans le cas des noms et des adjectifs français, de leur genre), les bons dictionnaires bilingues traditionnels (imprimés ou sur CD-ROM), c'est-à-dire les grands et les récents comme le *Robert & Collins Senior* (RCS) et le *Dictionnaire Hachette-Oxford* (OXHA), sont constitués d'entrées qui contiennent beaucoup d'éléments. Ceux-ci comprennent le mot-vedette (apparaissant sous une forme de base — fléchi ou non selon la langue), la catégorie grammaticale du mot-vedette, l'information grammaticale supplémentaire sur le mot-vedette (ex. le pluriel ou le féminin irréguliers ou les formes verbales irrégulières), les marques d'usage du mot-vedette (ex. les

marques de registre comme ‘informal’, les marques de domaine comme ‘Pathol’, etc.), une indication sémantique pour au moins certains sens du mot-vedette, parfois des actants²⁴ ou des référents²⁵, le (ou les) équivalent(s), l’information grammaticale sur chaque équivalent (ex. son genre, s’il est ‘non count’, etc.) et les marques d’usage qui s’appliquent à chaque équivalent. Certains dictionnaires (ex. le RCS) indiquent même la prononciation du mot-vedette.

En plus des différents sens et des différents équivalents du mot-vedette, les entrées contiennent souvent des exemples d’utilisation du mot-vedette avec leurs traductions (pour montrer l’usage du mot-vedette en contexte). Ces exemples comprennent des combinaisons libres, ainsi que des collocations, des expressions idiomatiques et des composés dans lesquels se trouve le mot-vedette. Lorsque le mot-vedette fait partie d’une collocation, d’une expression ou d’un composé pour lequel il n’est pas le mot principal, il fait l’objet d’un renvoi à l’entrée du mot principal de la collocation, de l’expression ou du composé.

Il est vrai que les constituants des entrées peuvent varier quelque peu même d’un bon dictionnaire à l’autre et d’une entrée à l’autre. Cela se voit en comparant l’article *grocery* (et son pluriel *groceries*) dans le RCS, le OXHA et le DCB.

²⁴ Selon la méthodologie du DCB (Roberts 1997 : 60), un actant est «a word which, used with the headword, determines the headword’s target language equivalents. *Example*: if the headword is a verb, the actant may be a noun serving as direct object of that verb. Thus, the verb **fournir** will have the equivalent “supply” if its direct object is *matériel*, and “produce” if the direct object is *pièce d’identité*».

²⁵ Selon la méthodologie du DCB (Roberts 1997 : 62), un référent est «[an] extension[...] of the definition. [...] a referent situates the equivalent in a general context».

a) RCS

grocery ['grəʊsəri] *n* a (*esp Brit: shop*) épicerie *f*. **he's in the ~ business** il est dans l'épicerie. **b I spent £7 on groceries** j'ai dépensé 7 livres en épicerie (*NonC*) or en provisions; **all the groceries are in this basket** toute l'épicerie est dans ce panier.

b) OHXA

groceries /grəʊsəri/ *npl* 1 (shopping) courses *fpl*; 2 (type of merchandise) épicerie *f* ϕ .

grocery /grəʊsəri/ ► 1692 *I n* (also ~ **shop** GB, ~ **store**) épicerie *f*.
II *modif* [*bill, products, sales*] d'épicerie; [*chain*] d'épiceries.

c) DCB

grocery *n* 1 **grocery** *inv* (*store*) épicerie *f*, magasin d'alimentation *m*, marché d'alimentation *m* [(*CD*)]. * **he went to the grocery to get some milk** il est allé chercher du lait à l'épicerie.

2 **groceries** (*commodities, especially food*) provisions *fpl*; (*food*) épicerie *f*, nourriture *f*. * **to do the groceries = to buy groceries = to shop for groceries** faire son épicerie [(*CD*)] = faire l'épicerie [(*CD*)] = faire le marché [(*CD*)] = faire son marché [(*CD*)] = faire ses courses; **he arrived loaded down with bags of groceries** il est arrivé à la maison les bras chargés de sacs d'épicerie [(*CD*)] = il est arrivé à la maison les bras chargés de sacs de provision; **how much do you spend on groceries?** combien dépensez-vous en épicerie?.

3 (*modifier*) [empty]. * **grocery bill** facture d'épicerie *f* [(*CD*)]; **grocery list** liste d'épicerie *f* [(*CD*)]; **grocery products** produits d'épicerie *mpl* = articles d'épicerie *mpl*; **grocery shopping** faire son épicerie [(*CD*)] = faire l'épicerie [(*CD*)] = faire le marché [(*CD*)] = faire son marché [(*CD*)] = faire ses courses.

(*cmp1*) **grocery bag** sac d'épicerie *m* [(*CD*)] (*au Québec, on ne semble pas faire la distinction entre le sac vide (sac à épicerie) et le sac plein (sac d'épicerie)*) = sac à épicerie *m* [(*CD*)] = sac à provisions *m*; **grocery cart** *a* (*in store*) chariot *m* =

chariot d'épicerie *m* = chariot de supermarché *m*; *b* (*small shopping cart*) chariot *m*;
grocery chain chaîne de magasins d'alimentation *f* = chaîne d'alimentation *f*;
grocery store épicerie *f* = magasin d'alimentation *m* = marché d'alimentation *m*
[(*CD*)]. *qv* **corner** *n*.

La différence la plus frappante dans la façon dont ces trois dictionnaires traitent *grocery* est que le OXHA fait une entrée séparée pour le pluriel **groceries** alors que le RCS le traite dans une combinaison libre et que le DCB le traite comme sous-entrée. Ensuite, la seconde différence principale est que le RCS et le OXHA, qui sont des dictionnaires européens, ne donnent pas de canadianismes [(*CD*)], contrairement au DCB. Enfin, l'entrée du DCB contient davantage de collocations et de composés que les entrées de RCS et de OXHA.

Mais comme la microstructure des articles du DCB, présentée en annexe, le montre, il est possible d'inclure une foule de détails sur chaque mot-vedette et ses équivalents.

Puisque les dictionnaires bilingues traditionnels sont des outils de référence tout faits auxquels les utilisateurs ne peuvent faire d'ajouts, ils doivent offrir un grand nombre d'entrées et suivre le plus près possible les développements de la LD et de la LA. Malheureusement, comme la création et la mise à jour d'un dictionnaire prennent plusieurs années, aucun dictionnaire ne peut se vanter d'être vraiment complet, car il ne peut inclure tous les nouveaux mots qui sont constamment créés, empruntés, utilisés dans de nouveaux sens, et formés par des procédés morphologiques normaux.

4.2 Constituants des dictionnaires bilingues automatiques

Les dictionnaires automatiques sont sensiblement différents des dictionnaires traditionnels sur plusieurs points :

- 1) à la différence des dictionnaires bilingues traditionnels, les dictionnaires bilingues automatiques ne donnent pas de définitions, si brèves soient-elles, ni d'exemples d'utilisation. Par contre, ils donnent, en principe, certaines informations syntaxiques (surtout les restrictions de sélection) et sémantiques (ex. les propriétés sémantiques) sur les mots-vedettes (ex. *know*: to know a fact or how to do something; *board*: flat surface or group of people) pour permettre au système de bien analyser le mot en LD et de le générer en LA. Le dictionnaire de base de Logos, par exemple, contient une case pour une catégorie sémantique (ex. *measurable concept*) pour le mot-vedette afin d'aider le système à mieux saisir le fonctionnement de ce mot dans une phrase;
- 2) contrairement aux dictionnaires traditionnels qui ne contiennent que les formes de base comme mots-vedettes, un petit nombre de systèmes de TA possèdent des dictionnaires de formes, c'est-à-dire contenant des entrées séparées pour toutes les formes fléchies des mots (ex. *walk*, *walks*, *walked*, etc.);
- 3) enfin, certains dictionnaires automatiques peuvent être créés, modifiés ou mis à jour par un utilisateur alors que les dictionnaires traditionnels sont créés, modifiés ou mis à jour exclusivement par des lexicographes professionnels;

Comme c'est le cas pour les dictionnaires bilingues traditionnels, il y a beaucoup de variation dans les entrées d'un dictionnaire bilingue automatique. Certains dictionnaires donnent simplement la liste, pour un mot donné, de tous les équivalents en LA avec leur catégorie grammaticale²⁶ (figure 27, Lehrberger et Bourbeau 1988 : 57).

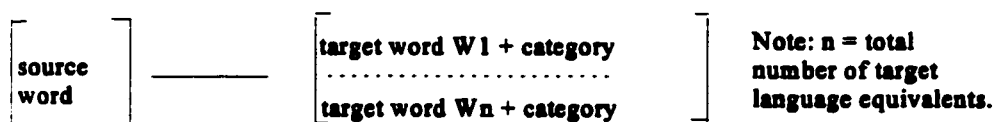


Figure 27 Entrée d'un dictionnaire bilingue automatique : 1^{ère} possibilité

D'autres donnent la liste des équivalents avec l'information non seulement sur leurs catégories grammaticales, mais aussi sur leurs sous-catégories syntaxiques ou sémantiques, sur leurs arguments (généralement les arguments des verbes), etc. (figure 28, Lehrberger et Bourbeau 1988 : 57).

²⁶ Les systèmes de TA modernes n'utilisent plus vraiment ce type de dictionnaire bilingue.

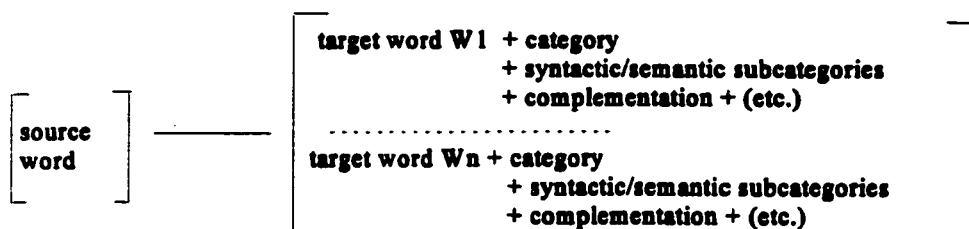


Figure 28 Entrée d'un dictionnaire bilingue automatique : 2^e possibilité

Il y en a d'autres encore qui, au lieu de simplement donner la liste de tous les équivalents, indiquent l'équivalent approprié pour des contextes différents (figure 29, Lehrberger et Bourbeau 1988 : 58), ce qui permet au système la sélection automatique d'un équivalent particulier selon le contexte.

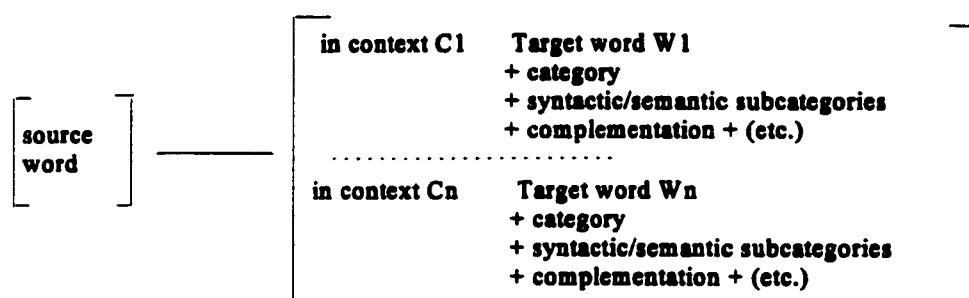


Figure 29 Entrée d'un dictionnaire bilingue automatique : 3^e possibilité

Les entrées des dictionnaires automatiques unilingues sont équivalentes, essentiellement, à des ensembles plus ou moins développés d'attributs et de valeurs. Par exemple, selon Arnold *et al.* (1994 : 94), pour le nom *button*, on pourrait avoir quelque chose du genre :

lex = button
cat = n
ntype = common
number = ²⁷
human = no
concrete = yes

L'information donnée indique que la base est *button* et que c'est un nom commun concret (contrairement à abstrait, comme *happiness* ou *sincerity*), mais non humain. Évidemment, si les catégories grammaticales sont différentes, les ensembles d'attributs seront aussi différents (ex. les verbes auront un **vtype** au lieu d'un **ntype**, etc). L'information sur la sous-catégorisation et sur les restrictions de sélection est ainsi codée à l'aide d'ensembles d'attributs ayant des valeurs telles que 'number', 'human', etc.

En ce qui concerne le traitement de l'information traductionnelle dans les dictionnaires bilingues, il est encore une fois possible d'utiliser des attributs et des valeurs (ex. *button*, cat = n, trans = bouton). Toutefois, selon Arnold *et al.* (1994 : 96), cette approche n'est pas particulièrement intéressante parce qu'elle ne permet de traduire que dans une direction (il sera difficile de créer des entrées pour le sens contraire). En général, on préférera utiliser des règles de traduction qui relient des mots-vedettes dans le dictionnaire de la LD à des mots-vedettes dans le dictionnaire de la LA (ex. *temperature* ↔ **température**).

²⁷ Le nombre n'est pas indiqué, car il n'y a rien de particulier à signaler.

Bien que les règles de traduction facilitent la tâche du système, il existe tout de même certains cas problématiques, particulièrement les cas d'homographie de même catégorie grammaticale (où l'utilisation de propriétés sémantiques et de restrictions de sélection est nécessaire) et de choix de l'équivalent.

Dans les dictionnaires bilingues automatiques, il est possible de traiter les problèmes d'homographie de même catégorie grammaticale en faisant appel à une hiérarchisation des entrées, c'est-à-dire que le dictionnaire est construit de façon à ce que le système consulte du vocabulaire général (séparé en modules distincts, consultés les uns après les autres) au vocabulaire spécifique (comme le fait le système METAL, voir chapitre 2, section 2.1.2.1). Cette approche hiérarchique permet aussi de traiter (partiellement du moins) certains problèmes de décalage.

D'autre part, l'intégration de certains dispositifs lexicaux dans les dictionnaires unilingues automatiques (constructions syntaxiques à interprétation sémantique identique, synonymes, collocations quasi synonymes avec des formes non collocationnelles, paraphrases syntaxiques, dérivés morphologiques réguliers, etc.) permet de traiter plus efficacement les différents choix possibles d'équivalents : «si la traduction repose sur une description du sens des unités lexicales de la langue de départ, il faudrait avoir, pour rendre ce sens en langue cible, le choix entre différentes lexicalisations possibles. Ceci implique que les unités lexicales de chaque langue soient décrites par rapport aux relations sémantiques qu'elles entretiennent» (Heid 1993 : 181). Le dictionnaire unilingue automatique doit décrire en détail les propriétés linguistiques des

possibilités lexicales et établir un réseau de relations sémantiques (ex. synonymiques, dérivationnelles, collocationnelles, etc.) entre les diverses possibilités.

Les différents éléments possibles dans un dictionnaire unilingue automatique, que nous venons de présenter, se retrouvent, à un degré moindre bien entendu, dans les dictionnaires bilingues des systèmes de transfert, qui donnent généralement de l'information sur les unités lexicales en LD et sur leurs équivalents, allant même jusqu'à donner, parfois, de l'information syntaxique sur la LA. Mais, comme nous l'avons signalé dans le chapitre 2, les différents systèmes de TA requièrent des dictionnaires bilingues différents (et des contenus différents). Ainsi, les dictionnaires des systèmes interlangues n'ont pas besoin de contenir d'information traductionnelle à proprement parler, mais tout simplement d'associer les mots avec les concepts interlangues appropriés.

Le grand avantage des dictionnaires bilingues automatiques de beaucoup de systèmes est que l'utilisateur a la possibilité d'ajouter des entrées soit au dictionnaire de base, soit dans un dictionnaire d'utilisateur. Cela veut dire que les dictionnaires automatiques n'ont pas besoin d'être aussi exhaustifs du point de vue de la nomenclature que les dictionnaires bilingues traditionnels.

Il a déjà été remarqué que les dictionnaires sont une des composantes les plus importantes de la plupart des systèmes de TA. En général, bien que certaines informations trouvées dans les dictionnaires traditionnels n'aient qu'une valeur limitée en TA (ex. l'information sur la

prononciation n'est utile que pour les systèmes 'speech to speech'), il existe d'autres informations (ex. les restrictions de sélection) nécessaires au système qui doivent apparaître dans les dictionnaires bilingues automatiques, y compris certaines qui sont évidentes et qui ne se trouvent donc pas dans les dictionnaires traditionnels (ex. certaines propriétés sémantiques comme 'humain', etc.).

4.3 Expérience : adaptation des constituants des dictionnaires bilingues traditionnels pour les dictionnaires bilingues automatiques

Le projet de Dictionnaire canadien bilingue, auquel nous travaillons comme assistante de recherche, prépare un dictionnaire traditionnel, qui paraîtra sous forme imprimée et sur CD-ROM. Cependant, les responsables du projet espèrent que les données du DCB pourront être utiles pour les systèmes de TA. Pour cette raison, nous avons décidé d'adapter une entrée bilingue du DCB (de l'anglais vers le français) sur laquelle nous avons travaillé (*grocery.n*) pour les dictionnaires bilingues automatiques de Globalink et de Logos.

Nous avons utilisé Globalink et Logos parce que ce sont les deux seuls systèmesq, parmi ceux qui sont à notre disposition, qui nous donnent accès aux dictionnaires. Nous avons choisi le substantif *grocery* parce qu'il appartient à la seule catégorie grammaticale pour laquelle Alex (de Logos) permet de préparer des entrées²⁸.

²⁸ Logos est l'un des systèmes de TA auxquels l'utilisateur peut avoir accès via Internet. Puisque nous avons facilement accès à Internet, nous nous sommes servie de cette méthode d'utilisation de Logos pour notre thèse. La version d'Alex que l'utilisateur peut télécharger du site Logos ne permet d'entrer que des substantifs, contrairement à la version d'Alex qui vient avec le système Logos et qui permet d'entrer des mots de toutes catégories grammaticales (mais à laquelle nous n'avons malheureusement pas accès).

Voici l'entrée du Dictionnaire canadien bilingue telle qu'elle est présentement (cette entrée est rendue à l'étape de la révision finale) :

grocery *n* 1 **grocery** *inv (store)* épicerie *f*, magasin d'alimentation *m*, marché d'alimentation *m* [(CD)]. * **he went to the grocery to get some milk** il est allé chercher du lait à l'épicerie.

2 **groceries** (*commodities, especially food*) provisions *fpl*; (*food*) épicerie *f*, nourriture *f*. * **to do the groceries = to buy groceries = to shop for groceries** faire son épicerie [(CD)] = faire l'épicerie [(CD)] = faire le marché [(CD)] = faire son marché [(CD)] = faire ses courses; **he arrived loaded down with bags of groceries** il est arrivé à la maison les bras chargés de sacs d'épicerie [(CD)] = il est arrivé à la maison les bras chargés de sacs de provision; **how much do you spend on groceries?** combien dépensez-vous en épicerie?.

3 {*modifier*} [empty]. * **grocery bill** facture d'épicerie *f* [(CD)]; **grocery list** liste d'épicerie *f* [(CD)]; **grocery products** produits d'épicerie *mpl* = articles d'épicerie *mpl*; **grocery shopping** faire son épicerie [(CD)] = faire l'épicerie [(CD)] = faire le marché [(CD)] = faire son marché [(CD)] = faire ses courses.

(*cmpl*) **grocery bag** sac d'épicerie *m* [(CD) (*au Québec, on ne semble pas faire la distinction entre le sac vide (sac à épicerie) et le sac plein (sac d'épicerie)*)] = sac à épicerie *m* [(CD)] = sac à provisions *m*; **grocery cart** *a (in store)* chariot *m* = chariot d'épicerie *m* = chariot de supermarché *m*; *b (small shopping cart)* chariot *m*; **grocery chain** chaîne de magasins d'alimentation *f* = chaîne d'alimentation *f*; **grocery store** épicerie *f* = magasin d'alimentation *m* = marché d'alimentation *m* [(CD)]. **qv corner** *n*.

Nous allons montrer que cette seule entrée exige plusieurs entrées différentes dans les dictionnaires automatiques.

4.3.1 *grocery* (division sémantique 1)

L'examen de l'article *grocery* du DCB nous montre que dans la division 1 (sens de *store*), il y a plusieurs éléments : la catégorie grammaticale (*n*), l'information grammaticale

supplémentaire (*inv*), l'indication sémantique (*store*), plusieurs équivalents, dont le dernier est accompagné d'une marque d'usage, et un exemple d'utilisation avec sa traduction. Tous ces renseignements ne peuvent figurer dans les entrées de Globalink et de Logos. Par contre, comme nous allons le constater, l'entrée de Logos exige d'autres renseignements non présentés dans l'entrée du DCB.

4.3.1.1 Globalink

Dans l'entrée *grocery* du dictionnaire de base (figure 30), Globalink présente le mot-vedette comme nom. Cependant, le dictionnaire ne permet pas d'indiquer qu'il est invariable dans le sens de *store*. Pour ce mot-vedette, le dictionnaire de base ne donne qu'un seul équivalent (*épicerie*). Puisque Globalink permet d'ajouter plusieurs équivalents dans l'ordre désiré et que dans le DCB, les équivalents sont placés par ordre de fréquence d'occurrence (les plus fréquents en premier), nous avons ajouté les deux autres équivalents (*magasin d'alimentation* et *marché d'alimentation*), dans l'ordre où ils apparaissent dans le DCB, afin que l'utilisateur puisse changer l'ordre des équivalents selon celui qu'il désire avoir dans la traduction (par exemple, pour un texte canadien, on pourrait décider de mettre le canadianisme — *marché d'alimentation* — comme premier équivalent).

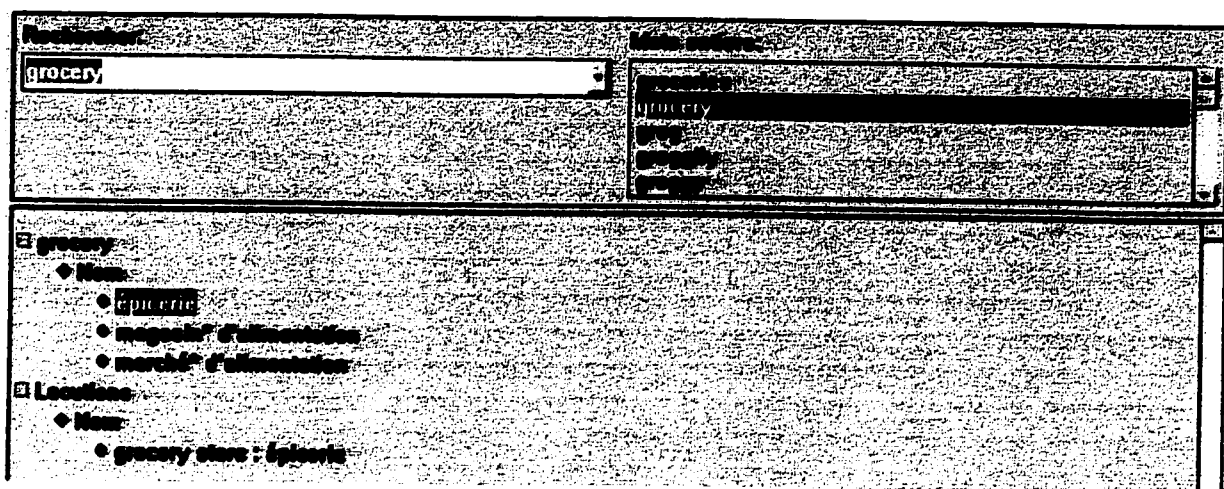


Figure 30 Modification de l'entrée *grocery* dans Globalink

4.3.1.2 Logos

Comme nous l'avons mentionné précédemment, il est impossible d'entrer plus d'un équivalent pour un mot donné dans le dictionnaire d'utilisateur de Logos (à moins de créer plusieurs dictionnaires d'utilisateur). Par conséquent, pour l'entrée *grocery*, nous n'avons donné que le premier équivalent principal donné par le DCB, *épicerie* (figure 31).

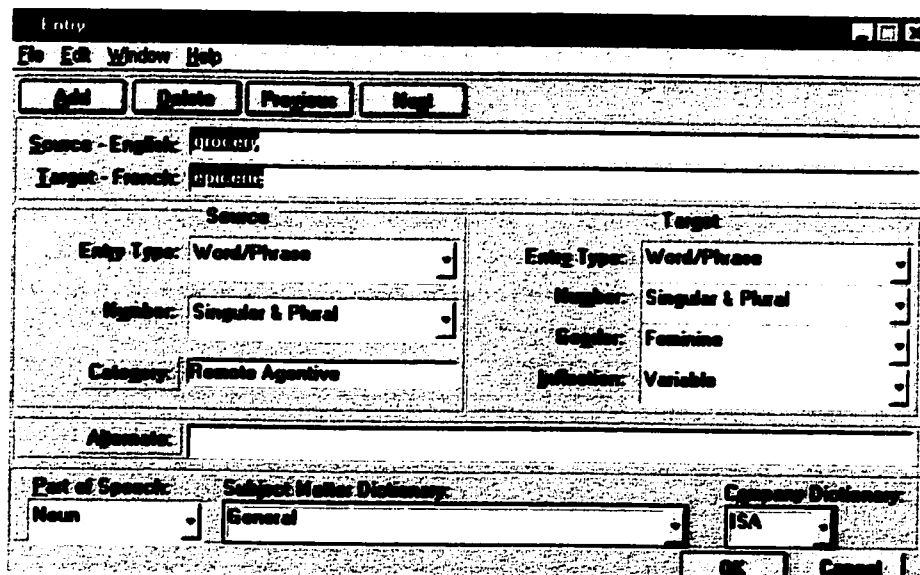


Figure 31 Préparation de l'entrée *grocery* dans Logos

Au moment de spécifier la catégorie du mot-vedette, il est possible de donner un synonyme pour faciliter la classification. Pour *grocery*, nous avons entré le synonyme *supermarket* et nous avons obtenu 'functional location' comme seule catégorie possible. Lorsque l'utilisateur sélectionne cette catégorie, une boîte de dialogue apparaît à l'écran demandant à l'utilisateur de spécifier, dans le cas présent, si le mot-vedette fonctionne parfois comme agent (figure 32).

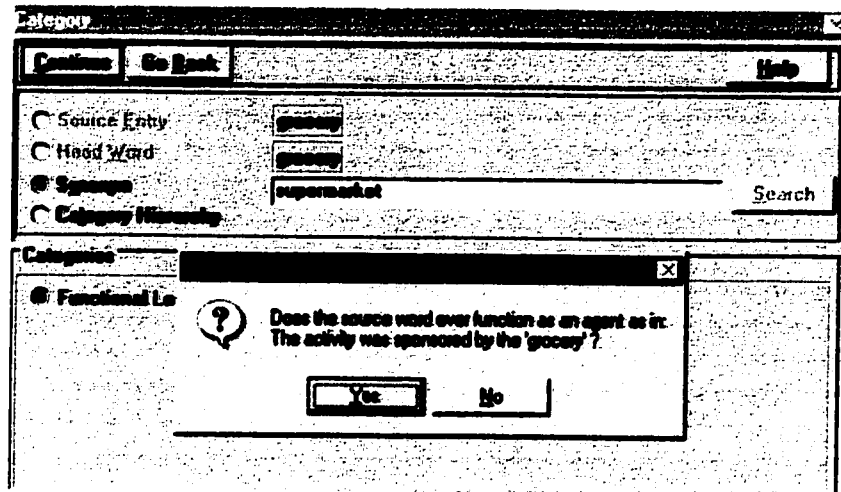


Figure 32 Sélection de la catégorie pour *grocery* dans Logos et boîte de dialogue

Puisque nous avons répondu par l'affirmative à la question posée par le système (à savoir si le mot-vedette *grocery* fonctionne comme agent dans ce sens), la catégorie (voir figure 31) est devenue 'remote agentive'.

4.3.2 groceries (division sémantique 2)

Dans l'article *grocery* du DCB, *groceries* est traité comme sous-entrée et possède, encore une fois, plusieurs équivalents. Comme pour *grocery*, nous avons modifié l'entrée *groceries* qui

existe comme entrée distincte dans le dictionnaire de base de Globalink et nous en avons créé une dans Logos.

4.3.2.1 Globalink

Dans l'entrée *groceries* du dictionnaire de base, Globalink ne donne que l'équivalent *articles d'épicerie* (qui n'est même pas donné dans le DCB). Nous avons donc ajouté les trois équivalents du DCB (*provisions*, *épicerie* et *nourriture*) dans l'ordre où ils apparaissent (figure 33). Un fait intéressant est que quand on consulte l'entrée *groceries* dans Globalink, elle apparaît avec l'entrée complète de *grocery* (avec les équivalents et les locutions qui se trouvaient déjà dans le dictionnaire de base et les équivalents que nous avons ajoutés) dans les deuxième et troisième sections de l'entrée (voir figure 33).

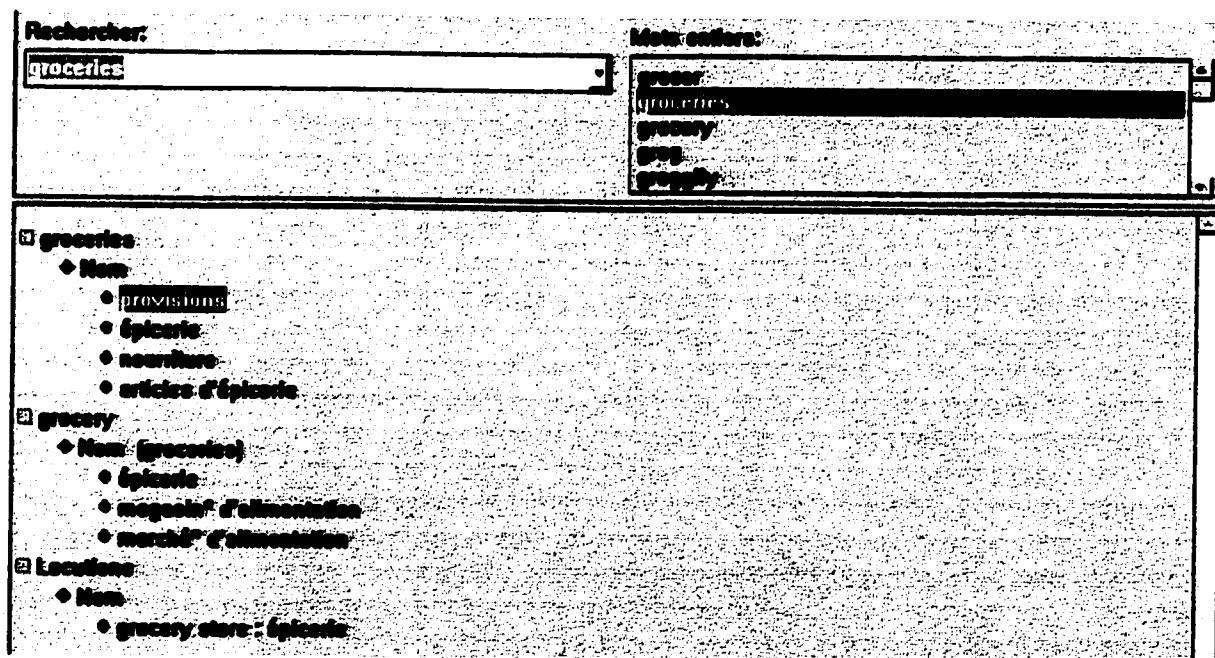


Figure 33 Modification de l'entrée *groceries* dans Globalink

4.3.2.2 Logos

Comme pour l'entrée *grocery*, nous n'avons donné que l'équivalent principal *provisions* (donné par le DCB) pour le mot-vedette *groceries* (figure 34). Pour la catégorie du mot-vedette, nous avons entré *food* comme synonyme et nous avons obtenu 'edible' comme seule catégorie possible. Contrairement à l'entrée *grocery*, aucune boîte de dialogue n'apparaît lorsque l'utilisateur sélectionne cette catégorie.

The screenshot shows the 'Entry' window in the Logos software. At the top, there are menu options: 'File', 'Edit', 'Window', and 'Help'. Below the menu is a toolbar with buttons: 'Add', 'Delete', 'Previous', and 'Next'. The main area is divided into two columns: 'Source' and 'Target'. Under 'Source', 'Entry Type' is set to 'Word/Phrase', 'Number' is 'Plural Only', and 'Category' is 'Edible'. Under 'Target', 'Entry Type' is 'Word/Phrase', 'Number' is 'Plural Only', 'Gender' is 'Feminine', and 'Inflection' is 'Invariable'. Below these columns is an 'Alternate' field. At the bottom, there are three dropdown menus: 'Part of Speech' (set to 'Noun'), 'Subject Matter Dictionary' (set to 'General'), and 'Company Dictionary' (set to 'ISA'). At the very bottom are 'OK' and 'Cancel' buttons.

Figure 34 Préparation de l'entrée *groceries* dans Logos

4.3.3 *to do the groceries*

Même si les entrées des dictionnaires bilingues automatiques ne contiennent pas d'«exemples» du genre de ceux qu'on trouve dans les dictionnaires traditionnels, nous avons voulu vérifier si nous ne pouvions pas en inclure un certain nombre. Dans l'article *grocery* du DCB, *to do the groceries* est traité comme combinaison libre dans la division sémantique 2 et c'est cet exemple que nous avons choisi d'ajouter aux dictionnaires de Globalink et de Logos. *To do the groceries*

est un exemple intéressant non seulement parce qu'il possède plusieurs équivalents, mais aussi parce qu'un utilisateur humain saurait conjuguer, le cas échéant, le verbe *to do*. Notre but étant de vérifier le traitement automatique du verbe dans cet exemple, nous n'avons pas entré tous les équivalents possibles pour Globalink, seulement *faire l'épicerie* (nous avons utilisé le même équivalent pour Logos).

4.3.3.1 Globalink

Dans l'entrée *groceries* du dictionnaire de base, Globalink ne donne pas la locution *to do the groceries*. Nous l'avons donc ajoutée avec son équivalent *faire l'épicerie*. Étant donné que Globalink permet de fléchir non seulement les noms, mais aussi les verbes, nous avons ajouté deux fois la même locution, *faire l'épicerie* (sans flexion) et *faire* l'épicerie* (avec flexion) (figure 35), afin de comparer les traductions de *to do the groceries*.

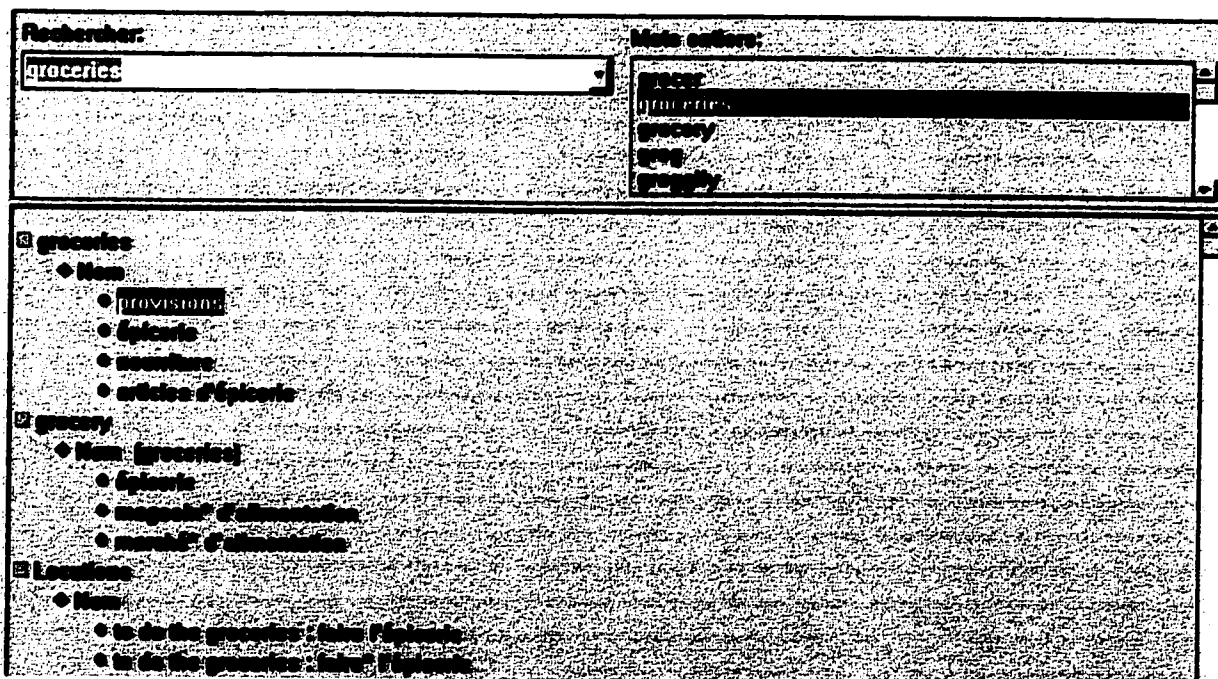


Figure 35 Ajout de la locution *to do the groceries* dans l'entrée *groceries* de Globalink

4.3.3.2 Logos

Puisque Logos permet d'entrer des mots-vedettes de plus d'un mot, nous avons fait une entrée séparée pour *to do the groceries* (figure 36).

The screenshot shows the 'Entry' window in the Logos software. The window has a menu bar (File, Edit, Window, Help) and buttons for 'Add', 'Delete', 'Previous', and 'Next'. The 'Source - English:' field contains 'to do the groceries' and the 'Target - French:' field contains 'faire l'épicerie'. Below these are two columns of settings: 'Source' and 'Target'. The 'Source' column has 'Entry Type: Word/Phrase', 'Number: Plural Only', and 'Category: Edible'. The 'Target' column has 'Entry Type: Word/Phrase', 'Number: Singular Only', 'Gender: Masculine', and 'Inflection: Invariable'. There is an 'Alternate:' field below the columns. At the bottom, there are three dropdown menus: 'Part of Speech:' (set to 'Noun'), 'Subject Matter Dictionary:' (set to 'General'), and 'Company Dictionary:' (set to 'ISA'). 'OK' and 'Cancel' buttons are at the bottom right.

Figure 36 Préparation de l'entrée *to do the groceries* dans Logos

Dans cette entrée, le mot principal de l'équivalent est *épicerie*. Étant donné que dans l'expression *faire l'épicerie*, *épicerie* reste invariable, nous avons indiqué dans la case nombre «singular only», ce qui fait automatiquement apparaître l'indication «invariable» dans la case flexion.

Comme nous l'avons expliqué au chapitre 3, Logos invite toujours l'utilisateur à fournir de l'information sur les autres mots de l'équivalent. Dans le cas présent, le seul autre mot de l'équivalent est le verbe *faire*. Étant donné qu'Alex (du moins la version que nous utilisons et qui provient de l'Internet) ne permet d'entrer que des substantifs comme mots-vedettes,

l'information supplémentaire sur les mots accompagnant l'équivalent principal peut se placer dans trois catégories grammaticales seulement ('noun', 'adjective modifying the head word' et 'other' — figure 37). Nous avons donc indiqué 'other' comme catégorie grammaticale pour le verbe *faire*, mais tout en pensant que le fait que l'utilisateur ne puisse pas indiquer clairement que *faire* est un verbe pourrait nuire à la traduction (voir les exemples de traduction dans la section 4.3.6).

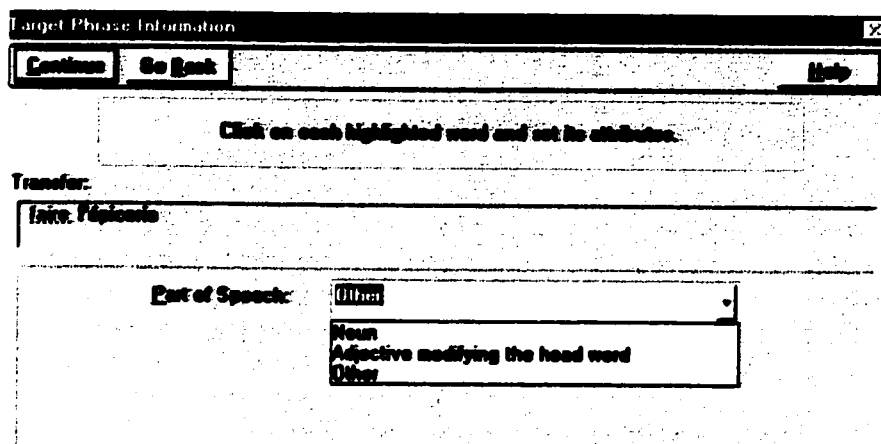


Figure 37 Information sur *faire* dans Logos

4.3.4 *grocery bill*

Un autre exemple de l'article *grocery* du DCB, que nous avons inclus dans les dictionnaires automatiques est *grocery bill*, qui est traité comme combinaison libre dans la division *modifier*.

4.3.4.1 Globalink

Puisque Globalink ne permet pas à l'utilisateur d'entrer des mots-vedettes de plus d'un mot, nous avons placé le groupe de mots *grocery bill* comme locution sous l'entrée *grocery* du dictionnaire de base et nous avons donné comme équivalent *facture d'épicerie* qui est le seul proposé dans le DCB (figure 38).

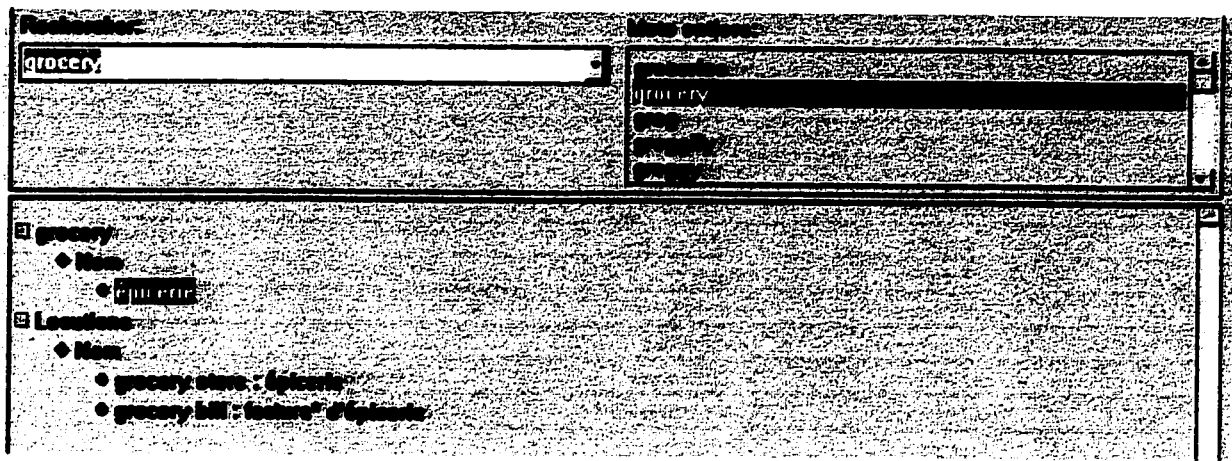


Figure 38 Ajout de la locution *grocery bill* dans l'entrée *grocery* de Globalink

4.3.4.2 Logos

Étant donné que Logos permet de préparer des entrées pour des groupes de mots, nous avons créé une entrée séparée pour *grocery bill*. Comme pour les autres entrées de plus d'un mot, le système demande à l'utilisateur de surligner le mot principal du mot-vedette et de l'équivalent qui sont, dans ce cas-ci, *bill* et *facture* (figure 39).

The screenshot shows the 'Entry' window in the Logos software. The window has a menu bar (File, Edit, Window, Help) and buttons for 'Add', 'Delete', 'Previous', and 'Next'. The 'Source - English' field contains 'grocery bill' and the 'Target - French' field contains 'facture d'épicerie'. Below these fields is a table with two columns: 'Source' and 'Target'. The table has four rows: 'Entry Type', 'Number', 'Gender', and 'Inflection'. The 'Source' column has dropdown menus for 'Word/Phrase', 'Singular & Plural', and 'Measurable Concept'. The 'Target' column has dropdown menus for 'Word/Phrase', 'Singular & Plural', 'Feminine', and 'Variable'. Below the table is an 'Alternate' field. At the bottom, there are three dropdown menus: 'Part of Speech' (set to 'Noun'), 'Subject Matter Dictionary' (set to 'General'), and 'Company Dictionary' (set to 'ISA'). The 'OK' and 'Cancel' buttons are at the bottom right.

Source		Target	
Entry Type:	Word/Phrase	Entry Type:	Word/Phrase
Number:	Singular & Plural	Number:	Singular & Plural
Category:	Measurable Concept	Gender:	Feminine
		Inflection:	Variable

Alternate:

Part of Speech: Noun Subject Matter Dictionary: General Company Dictionary: ISA

OK Cancel

Figure 39 Préparation de l'entrée *grocery bill* dans Logos

Pour l'équivalent, le système demande de l'information supplémentaire sur le mot *épicerie*.

Nous avons indiqué qu'il s'agit d'un nom féminin au singulier, car *épicerie* reste invariable même quand *facture d'épicerie* est au pluriel (figure 40).

Target Phrase Information	
Continue Go Back Help	
Click on each highlighted word and set its attributes.	
Transfer:	
facture d'épicerie	
Part of Speech:	Noun
Gender:	Feminine
Number:	Singular

Figure 40 Information sur *épicerie* dans Logos

4.3.5 *grocery store*

Le dernier exemple de l'article *grocery* du DCB que nous avons voulu entrer dans les dictionnaires automatiques est le composé *grocery store* qui possède trois équivalents (dont un *canadianisme*).

4.3.5.1 Globalink

Dans Globalink, le groupe de mots *grocery store* figure déjà dans l'entrée *grocery* comme locution avec l'équivalent *épicerie* (voir figure 30). Nous avons ajouté les équivalents *magasin d'alimentation* et *marché d'alimentation* en plaçant ce dernier comme premier équivalent, car il est canadien d'origine (figure 41).

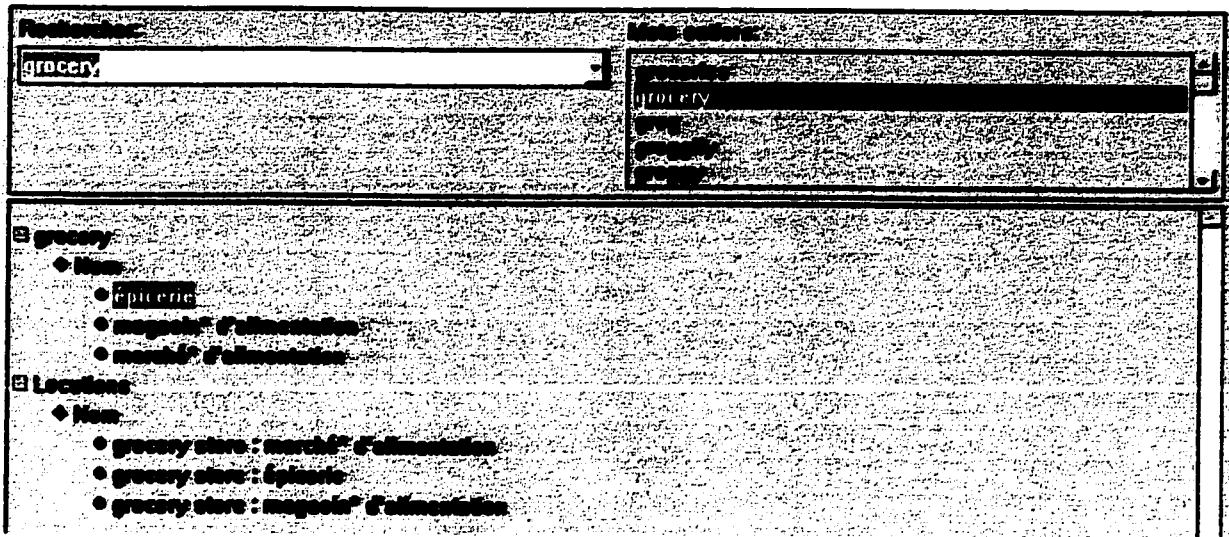


Figure 41 Modification de la locution *grocery store* dans l'entrée *grocery* de Globalink

4.3.5.2 Logos

Dans Logos, nous avons créé l'entrée *grocery store* en choisissant le canadianisme *marché d'alimentation* comme équivalent, car Logos ne permet qu'un seul équivalent. Les mots *store* et *marché* ont été surlignés par le système comme mots principaux et nous les avons gardés comme tel parce que ce sont, en effet, les mots principaux de ces groupes de mots (figure 42). Pour l'équivalent, Alex demande l'information supplémentaire sur le mot *alimentation* (nom, féminin, singulier). Comme pour l'entrée *grocery*, une boîte de dialogue demandant si le mot-vedette fonctionne parfois comme agent apparaît à l'écran lorsque l'utilisateur sélectionne la catégorie 'place' (figure 32) et, comme pour l'entrée *grocery*, nous avons répondu par l'affirmative.

The screenshot shows the 'Entry' window in the Logos software. At the top, there are menu options: 'File', 'Edit', 'Window', and 'Help'. Below the menu is a toolbar with buttons: 'Add', 'Delete', 'Previous', and 'Next'. The main area is divided into two columns: 'Source' and 'Target'.
 Source column:
 - Source - English: grocery store
 - Target - French: marché d'alimentation
 - Entry Type: Word/Phrase
 - Number: Singular & Plural
 - Category: Place
 - Alternate: (empty field)
 - Part of Speech: Noun
 - Subject Matter Dictionary: General
 - Company Dictionary: ISA
 Target column:
 - Entry Type: Word/Phrase
 - Number: Singular & Plural
 - Gender: Masculine
 - Inflection: Variable
 At the bottom right, there are 'OK' and 'Cancel' buttons.

Figure 42 Préparation de l'entrée *grocery store* dans Logos

4.3.6 Conclusion sur l'expérience

Dans l'ensemble, la microstructure du DCB permet d'entrer plus d'information sur *grocery* que le format des entrées des dictionnaires bilingues de Globalink et de Logos. Par exemple, dans l'entrée du DCB, les équivalents sont présentés en ordre de fréquence, ce que nous avons fait dans Globalink; mais dans le DCB, le bon équivalent peut être choisi par l'utilisateur selon le contexte, tandis que dans Globalink, le système donnera toujours le premier équivalent par défaut, sans tenir compte du contexte. Un autre point concernant les équivalents est l'indication du féminin, du masculin ou du pluriel. Comme le DCB, Logos donne de l'information grammaticale sur l'équivalent du mot-vedette dans l'entrée de ce dernier, alors que Globalink ne permet pas d'en ajouter dans son dictionnaire anglais-français. L'utilisateur doit donc indiquer l'information grammaticale sur les mots français dans le dictionnaire français-anglais.

Ensuite, alors que le DCB contient des exemples pour illustrer l'utilisation du mot-vedette et des équivalents en contexte, Logos ne permet d'entrer ces exemples que comme entrées distinctes. En ce qui concerne Globalink, il permet aussi d'entrer des exemples en autant qu'ils contiennent des noms, car ces exemples ne peuvent être saisis que comme locutions dans l'entrée des noms (ex. *grocery bill* sous *grocery*).

Une autre différence entre l'entrée du DCB et les entrées des dictionnaires automatiques touche les indications de sens. Tandis que le DCB permet la formulation libre d'une indication de sens précise, la seule façon de donner une quelconque indication de sens dans Globalink est au moyen des attributs, qui sont très généraux (ex. humain, matière, etc.) et très limités en nombre²⁹. Pareillement, même si Alex (de Logos) permet, en quelque sorte, d'entrer des indications de sens sous forme de catégories (ex. *grocery store* = place, *groceries* = edible, etc.), le choix des catégories reste limité.

Enfin, l'article *grocery* du DCB contient plusieurs équivalents spécifiquement canadiens, clairement identifiés par la marque géographique (*CD*), alors qu'il est impossible de marquer les équivalents dans les systèmes de TA Globalink et Logos.

²⁹ Dans Globalink, il est possible de créer des attributs personnalisés (contrairement à Logos où tout est prédéfini), mais le système ne les reconnaît pas.

Dans un autre ordre d'idées, lorsque nous avons essayé de traduire les phrases suivantes : (1)

Whose turn is it to do the groceries? et (2) *Her husband did the groceries yesterday*, dans

Globalink et Logos, nous avons obtenu les résultats ci-dessous :

Globalink : (1) À qui tour est-ce que c'est faire l'épicerie? (traduction obtenue avec la locution sans flexion)

(1) À qui tour est-ce que c'est faits l'épicerie? (traduction obtenue avec la locution avec flexion)

(2) Son mari a fait les provisions hier.

Logos : (1) À qui est de qui ce tour le faire l'épicerie?

(2) Son mari a fait les épiceries hier.

À première vue, les traductions obtenues dans Globalink semblent meilleures que celles qui ont été obtenues dans Logos. Pour traduire les phrases de départ, Globalink a utilisé l'entrée modifiée *groceries* du dictionnaire de base, où nous avons ajouté l'équivalent **provisions**, et la locution *to do the groceries* (avec et sans flexion) avec sa traduction **faire l'épicerie**. Même si le système n'a pas reconnu que *did the groceries* dans la phrase (2) n'est que la forme fléchie de *do the groceries* et qu'il traduit séparément *did* (**a fait**) et *groceries* (**les provisions**), le résultat est néanmoins acceptable. Cependant, dans la deuxième traduction de la phrase (1) (celle qui utilise la locution avec flexion), il est incompréhensible que le verbe soit accordé comme participe (*faits*) puisqu'il n'y a pas d'auxiliaire et pas de sujet au pluriel.

Au contraire de Globalink, qui a utilisé l'entrée que nous avons modifiée, Logos n'a pu utiliser l'entrée de *groceries* (dont l'équivalent est *provisions*), que nous avons créée pour sa traduction de la phrase (2). Dans la traduction de cette phrase, le système ne semble pas avoir utilisé non

plus l'entrée que nous avons créée pour *to do the groceries*, dans laquelle nous avons indiqué qu'*épicerie* reste invariable dans cette expression. Il est encore une fois incompréhensible que le système ait accordé *épicerie* au pluriel malgré les entrées que nous avons créées. La traduction de la phrase (1), quant à elle, contient bien l'équivalent donné dans l'entrée *to do the groceries*, soit **faire l'épicerie**.

4.4 Conclusion

Présentement, les dictionnaires bilingues traditionnels semblent donner, dans l'ensemble, plus d'information utile pour la traduction que les dictionnaires bilingues automatiques, même si ces derniers contiennent parfois des renseignements qui ne se trouvent pas dans les premiers (c'est-à-dire l'information sur la sous-catégorisation, sur les restrictions de sélection, etc.). En fait, ce que nous entendons par «information utile» est la présence de différents exemples (combinaisons libres, collocations, expressions et composés) dans les entrées de dictionnaires bilingues traditionnels. Ces exemples permettent à l'utilisateur humain d'observer l'utilisation du mot-vedette et des équivalents en contexte. Comme le dit Toope dans sa thèse *Examples in the Bilingual Dictionary* :

Examples — phrases or sentences that show the lemma in context and their translations — are one of the means by which bilingual lexicographers present information about a lemma. The lexicographer may first present information by such means as semantic and grammatical indicators, lists of equivalents in the target language, and typographical symbols that warn the user when a lemma has special stylistic characteristics and when it appears in special syntagmatic structures. Examples reinforce this information by showing how it all works together in a concrete context (96 : 1)

Il faut souligner que des recherches sont présentement menées afin de mettre au point des systèmes de TA basés sur l'exemple, qui sont fondés sur l'analyse d'exemples provenant de corpus spécifiques (Arnold *et al.* 1994 : 198). Toutefois, les systèmes de ce genre ne comprennent pas de dictionnaires et ne nous intéressent donc pas dans cette thèse. Ce qui attire davantage notre attention, ce sont les tentatives faites pour créer des dictionnaires automatiques plus complets, en utilisant non seulement les dictionnaires existants, mais aussi des corpus (Klavans 1990 et 1996), ce qui pourrait mener à l'élaboration des dictionnaires automatiques contenant plus d'information utile pour la traduction. Mais la sélection de l'information qui se trouve dans les sources électroniques (dictionnaires sur CD-ROM et corpus électroniques) ainsi que son extraction automatique restent toujours problématiques (Nirenburg 1994 : 314, Meijs 1992) et, tant que ces problèmes ne seront pas résolus, l'élaboration de bons dictionnaires automatiques sera très coûteuse.

CONCLUSION

Dans le cadre de cette thèse, nous nous étions fixé comme premier objectif d'examiner les dictionnaires automatiques dans un certain nombre de systèmes de TA. Nous avons d'abord décrit les différents types de systèmes de TA (direct, de transfert et interlangue) dont nous allions étudier les dictionnaires et les trois phases de traduction des systèmes de transfert. Ceci nous a permis de mieux comprendre à quelles étapes du processus de traduction les dictionnaires sont consultés par différents systèmes. Une étude plus poussée des dictionnaires de certains systèmes nous a permis de constater que leur fonctionnement dépend principalement du type de système dans lequel ils s'intègrent et du genre de système dont il s'agit (système opérationnel, expérimental, bilingue, multilingue, etc.).

Notre deuxième objectif était d'étudier le rôle du dictionnaire bilingue en TA. Encore une fois, selon le type de système auquel on fait référence, le dictionnaire bilingue ne contient pas tout à fait les mêmes informations. Ainsi, les dictionnaires bilingues des systèmes directs donnent, en général, plus d'informations nécessaires à la traduction, sans toutefois offrir une analyse linguistique aussi élaborée que les dictionnaires unilingues des systèmes de transfert et des systèmes interlangues. Quant aux dictionnaires «bilingues» des systèmes interlangues, ils ne mettent pas directement en parallèle les unités lexicales de la LD avec celles de la LA, mais associent plutôt chaque unité lexicale de la LD avec une représentation interlangue, et cette même représentation interlangue avec une unité lexicale de la LA.

Dans l'ensemble, le nombre de dictionnaires bilingues utilisés par chaque système ainsi que le contenu des entrées comme telles est différent pour chaque système étudié, même s'il s'agit de systèmes de même type. Par exemple, Globalink, un système direct, n'utilise qu'un seul dictionnaire de base qui comprend des mots-vedettes simples et les locutions qui s'y rapportent, en revanche, Systran et Météo, qui sont aussi des systèmes directs, contiennent respectivement cinq et trois dictionnaires bilingues distincts (dictionnaire général, dictionnaire d'expressions idiomatiques, etc.).

Enfin, notre troisième objectif visait à comparer les dictionnaires bilingues traditionnels et les dictionnaires bilingues automatiques. En premier lieu, nous avons examiné quelques entrées de dictionnaires bilingues automatiques des systèmes Globalink et Logos pour avoir un aperçu de la façon dont leurs entrées sont constituées. Ensuite, pour l'étude générale des dictionnaires bilingues traditionnels, nous avons présenté les renseignements contenus dans l'entrée *grocery* de trois dictionnaires récents, soit le RCS, le OXHA et le DCB, et nous les avons comparés.

Afin de faciliter la comparaison entre les dictionnaires bilingues traditionnels et les dictionnaires bilingues automatiques, nous avons choisi l'entrée *grocery* du DCB, déjà examinée dans la comparaison des dictionnaires traditionnels et des dictionnaires automatiques, et l'avons intégrée aux dictionnaires automatiques des deux seuls systèmes de TA auxquels nous avons accès assez facilement, soit Globalink et Logos. Même si nous avons préparé plusieurs entrées, il nous a été impossible d'intégrer tous les renseignements qui se trouvaient dans le DCB. Cependant, notre comparaison des dictionnaires traditionnels et des dictionnaires automatiques est loin d'être

exhaustive dans le sens où les différents systèmes de TA ne fonctionnent pas nécessairement de la même façon et ne permettent pas à l'utilisateur d'inclure les mêmes informations dans leurs dictionnaires. Selon la documentation fournie pour certains systèmes (ex. Logos, METAL), leurs dictionnaires bilingues exigeraient plus d'information que ce qu'on trouve dans le DCB.

La conclusion peut-être la plus importante que nous puissions tirer de nos essais de modification et de création d'entrées de dictionnaires automatiques est que, même quand l'entrée est bien préparée, il n'est pas toujours possible d'obtenir les résultats voulus, car si les dictionnaires, et surtout le dictionnaire bilingue, sont une des pièces maîtresses des systèmes de TA, ils dépendent intimement des autres composantes. Ainsi, le lexicographe qui prépare un dictionnaire automatique ne peut travailler seul. En fait, le développement de bons systèmes de TA nécessite une étroite collaboration entre les développeurs de systèmes (informaticiens et linguistes) et les lexicographes qui préparent des dictionnaires pour ces systèmes.

Nous croyons également que le travail des lexicographes «traditionnels» a son importance en lexicographie automatique et qu'il serait possible de modifier les entrées des dictionnaires bilingues traditionnels de façon à ce qu'elles soient plus facilement utilisables par les systèmes de TA.

Depuis quelques années, les dictionnaires sont surtout rédigés en format SGML. Parmi ses nombreux avantages, le balisage SGML permet, notamment, de choisir les parties de l'entrée qui doivent être reproduites selon le genre de dictionnaire (imprimé, CD-ROM, etc.). Ainsi, le

lexicographe traditionnel peut dorénavant ajouter à des entrées traditionnelles des renseignements essentiels pour un dictionnaire automatique, sans pour autant compromettre la qualité ou la lisibilité du dictionnaire traditionnel. Il suffirait d'ajouter à la DTD qui sous-tend le format SGML³⁰ d'une ou plusieurs sections qui contiennent des renseignements destinés aux dictionnaires automatiques.

Au DCB, la DTD (annexe 1) est constituée de sections obligatoires (il faut, par exemple, au moins une division de sens) et de sections optionnelles (ex. la section des composés, car tous les mots ne forment pas nécessairement des composés), qui peuvent être répétées ou non. À l'intérieur de chaque section, certains éléments (ex. la catégorie grammaticale) peuvent aussi être obligatoires, alors que d'autres ne le sont pas (ex. les marques d'usage). La DTD conçue au DCB permet d'inclure tous les renseignements voulus dans les entrées dictionnaires; c'est en préparant plusieurs feuilles de style que les lexicographes pourront visualiser et/ou imprimer une même entrée de différentes façons selon les besoins. En somme, le document SGML contient beaucoup plus d'information que ce qui apparaîtra dans la version imprimée du DCB. Par conséquent, il est possible d'ajouter des renseignements qui ne sont utiles que pour un dictionnaire automatique dans une entrée traditionnelle, sans pour autant compliquer la présentation de l'entrée imprimée.

³⁰ «A Document Type Definition consists of rules that apply SGML to the markup of documents of a particular type» (Roberts, description du projet du DCB 1998, basé sur Turner *et al.* 1996 : 205).

Il semblerait logique de grouper ensemble, à l'intérieur d'une entrée SGML, toute l'information utile pour un système de TA, notamment les propriétés sémantiques et les restrictions de sélection. Une question s'impose toutefois : cette information doit-elle paraître immédiatement après l'élément du dictionnaire traditionnel auquel elle s'applique ou est-il préférable de l'inscrire en un seul bloc, à la toute fin de l'entrée?

Dans le premier cas, il faudrait créer des sous-sections, appelées 'MT' par exemple³¹, qu'on inclurait dans la DTD après chaque élément de l'entrée traditionnelle qui exige des renseignements supplémentaires pour la TA (ex. dans les divisions de sens — après les combinaisons libres — et après chaque collocation, chaque expression idiomatique et chaque composé).

La deuxième solution proposée implique la création d'une section séparée (qui pourrait aussi s'appeler 'MT') que nous pourrions placer à la fin de l'entrée traditionnelle complète ('entry.full' dans l'annexe 1, après le bloc 'cmp2'). L'inconvénient avec cette approche est qu'il serait difficile d'identifier quels éléments d'informations s'appliqueraient à un sens particulier et quels éléments ne seraient pertinents que pour une collocation ou un composé. Il faudrait donc, à l'intérieur de ce bloc 'MT', reproduire la structure sémantique de l'entrée afin de faire la part des choses entre les divers éléments d'information.

³¹ La DTD du DCB est faite en anglais, c'est pourquoi nous proposons 'MT' et non 'TA'.

Dans un cas comme dans l'autre, si nous créons une (ou des) sous-section(s) obligatoire(s), il faudra ajouter les renseignements typiquement 'MT' aux entrées déjà terminées pour qu'elles puissent parser, car elles ne seraient pas des 'instances SGML valables'. D'autre part, si nous ne créons qu'une (ou des) sous-section(s) optionnelle(s), seules les entrées contenant cette (ou ces) sous-section(s) pourront faire partie d'un dictionnaire automatique. Par conséquent, toutes les entrées faites avant la création de cette (ou de ces) sous-section(s) ne seraient d'aucune utilité pour la construction d'un dictionnaire automatique.

Dans l'ensemble, la première solution proposée semble la plus logique du point de vue du lexicographe, parce qu'en général, il travaille sens par sens et section par section. Afin d'illustrer clairement de quelle façon l'information ajoutée pourrait apparaître dans une entrée traditionnelle, nous avons ajouté des exemples de sous-sections 'MT' dans chaque section de l'entrée pour *grocery* du DCB.

grocery *n* 1 **grocery** *inv (store)* épicerie *f*, magasin d'alimentation *m*, marché d'alimentation *m* [(CD)]. * **he went to the grocery to get some milk** il est allé chercher du lait à l'épicerie. (MT) propriété sémantique, propriété sémantique, prop. sémantique (etc.). restriction de sélection, restriction de sélection (etc.)³².

2 **groceries** (*commodities, especially food*) provisions *fpl*; (*food*) épicerie *f*, nourriture *f*. * **to do the groceries = to buy groceries = to shop for groceries** faire son épicerie [(CD)] = faire l'épicerie [(CD)] = faire le marché [(CD)] = faire son marché [(CD)] = faire ses courses; **he arrived loaded down with bags of groceries** il est arrivé à la maison les bras chargés de sacs d'épicerie [(CD)] = il est arrivé à la maison les bras chargés de sacs de provision; **how much do you spend on groceries?** combien

³² L'information proposée dans la section 'MT' n'est pas exhaustive : il ne s'agit que d'un exemple.

dépensez-vous en épicerie?. (MT) propriété sémantique, propriété sémantique, prop. sémantique (etc.). restriction de sélection, restriction de sélection (etc.).

3 {modifier} [empty]. * **grocery bill** facture d'épicerie *f* [(CD)]; **grocery list** liste d'épicerie *f* [(CD)]; **grocery products** produits d'épicerie *mpl* = articles d'épicerie *mpl*; **grocery shopping** faire son épicerie [(CD)] = faire l'épicerie [(CD)] = faire le marché [(CD)] = faire son marché [(CD)] = faire ses courses. (MT) propriété sémantique, propriété sémantique, prop. sémantique (etc.). restriction de sélection, restriction de sélection (etc.).

(cmpl) **grocery bag** sac d'épicerie *m* [(CD) (au Québec, on ne semble pas faire la distinction entre le sac vide (sac à épicerie) et le sac plein (sac d'épicerie))] = sac à épicerie *m* [(CD)] = sac à provisions *m*. (MT) propriété sémantique, propriété sémantique, prop. sémantique (etc.). restriction de sélection, restriction de sélection (etc.).; **grocery cart** *a* (in store) chariot *m* = chariot d'épicerie *m* = chariot de supermarché *m*; *b* (small shopping cart) chariot *m*. (MT) propriété sémantique, propriété sémantique, prop. sémantique (etc.). restriction de sélection, restriction de sélection (etc.).; **grocery chain** chaîne de magasins d'alimentation *f* = chaîne d'alimentation *f*. (MT) propriété sémantique, propriété sémantique, prop. sémantique (etc.). restriction de sélection, restriction de sélection (etc.).; **grocery store** épicerie *f* = magasin d'alimentation *m* = marché d'alimentation *m* [(CD)]. (MT) propriété sémantique, propriété sémantique, prop. sémantique (etc.). restriction de sélection, restriction de sélection (etc.). qv corner *n*.

Malgré qu'il soit possible, d'une manière ou d'une autre, d'inclure l'information utile à la TA dans des entrées traditionnelles en format SGML, il reste cependant à déterminer comment ces renseignements pourraient être versés automatiquement dans les dictionnaires automatiques.

L'entreprise Alis Technologies de Montréal, par exemple, a prévu un projet qui vise précisément à atteindre ce but : créer une interface d'entrée commune dans laquelle on transférerait automatiquement les renseignements qui se trouvent déjà dans les dictionnaires électroniques et qui, par la suite, pourrait permettre d'ajouter, automatiquement encore, les données requises par

les dictionnaires automatiques de différents systèmes. Toutefois, on n'en est, là encore, qu'à l'étape des recherches. Entre-temps, les lexicographes traditionnels possédant une bonne connaissance du contenu des dictionnaires automatiques peuvent continuer, pour leur part, à préparer le terrain, en s'assurant que les éléments requis par ces dictionnaires sont présents dans les nouveaux dictionnaires traditionnels qu'ils préparent.

ANNEXE 1

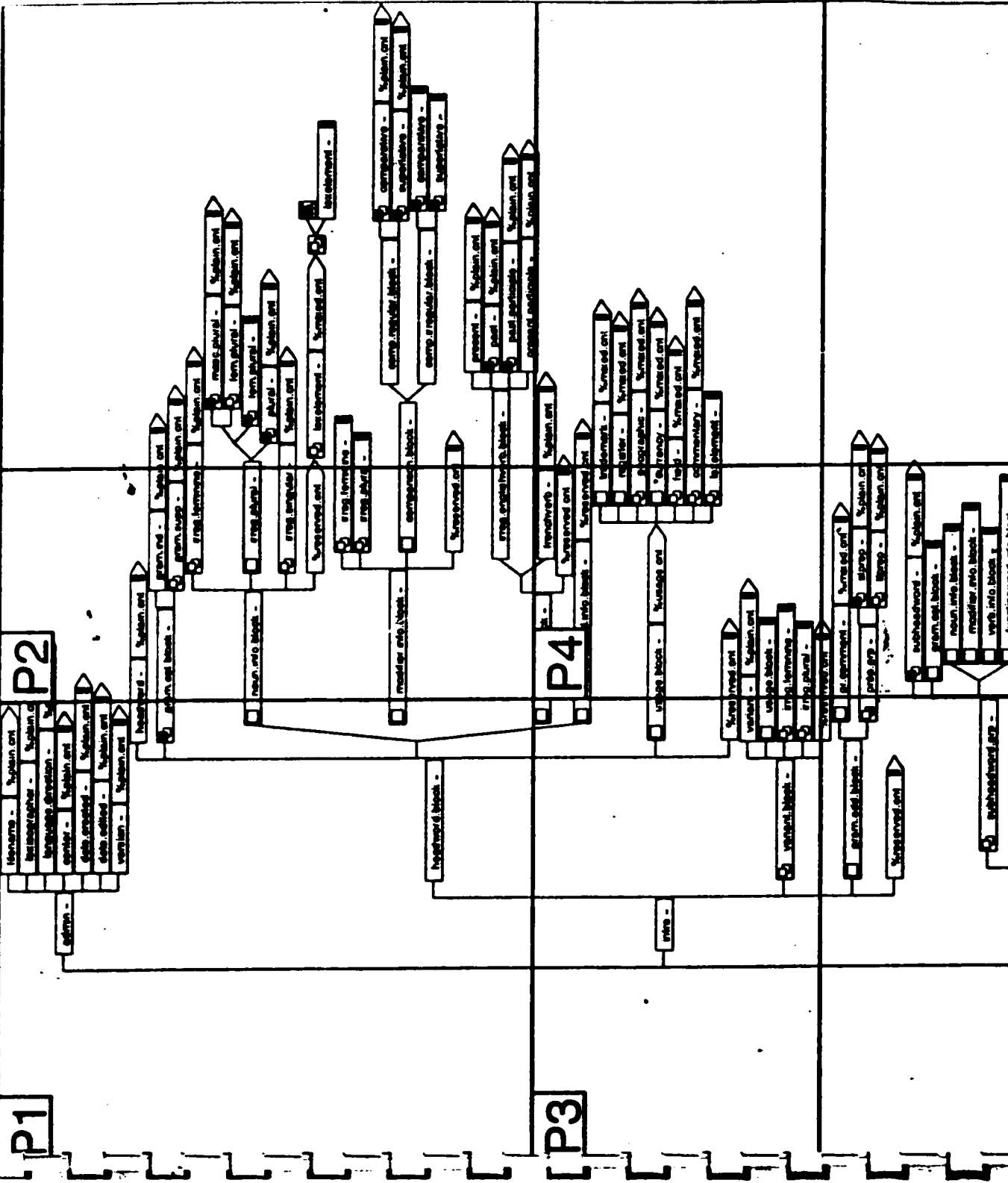
DTD Entry.Full

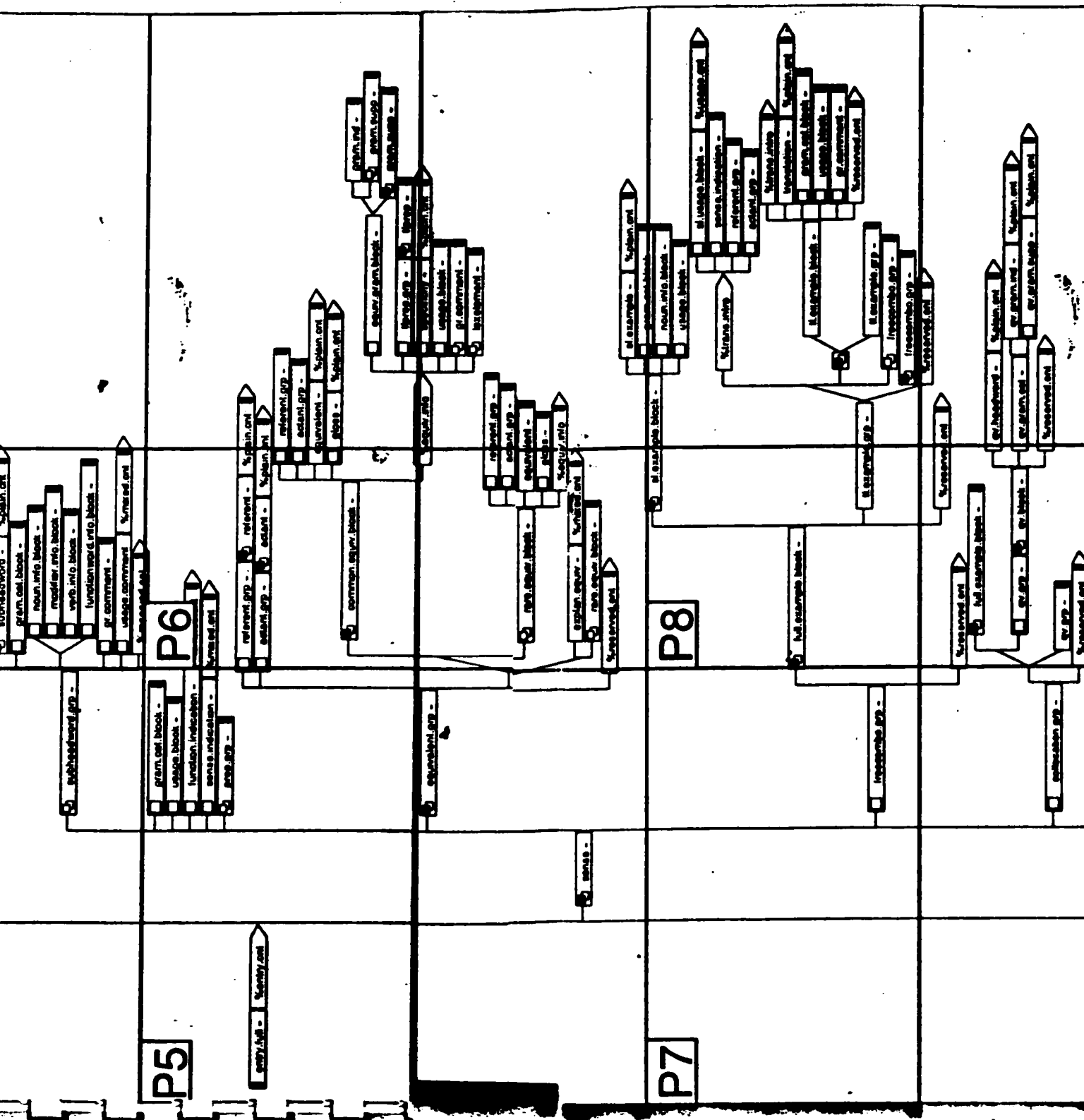
P1

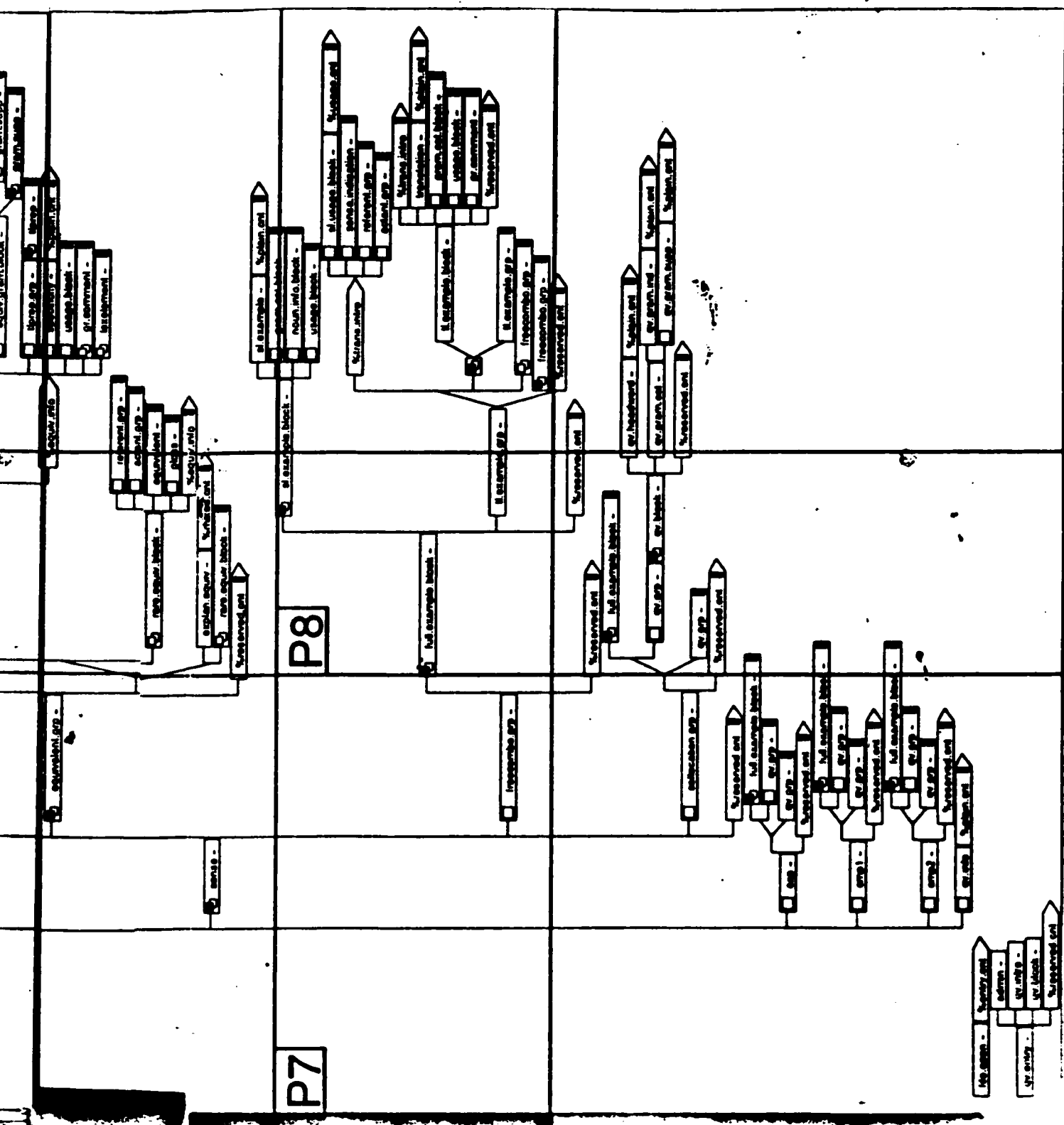
P2

P3

P4







BIBLIOGRAPHIE

Ouvrages et articles cités

- ARNOLD, Doug (1993) : «Sur la conception du transfert», Pierrette Bouillon et André Clas (dir.), *La traductique. Études et recherches de traduction par ordinateur*, Montréal, Les Presses de l'Université de Montréal et AUPELF/UREF, pp. 64-76.
- ARNOLD, Doug *et al.* (1994) : *Machine Translation. An Introductory Guide*, Oxford et Cambridge (Massachusetts), NCC Blackwell Ltd. et Blackwell Publishers, VIII + 240 p.
- ATKINS, B. T. S. *et al.* (1994) : «Computational Approaches to the lexicon: An Overview», B. T. S. Atkins et A. Zampolli (dir.), *Computational Approaches to the Lexicon*, Oxford, Oxford University Press, p. 17-45.
- ATKINS, Beryl T. et Beth LEVIN (1991) : «Admitting Impediments», Uri Zernik (dir.), *Lexical Acquisition. Exploiting On-Line Resources to Build a Lexicon*, Hillsdale (New Jersey), Lawrence Erlbaum Associates, Inc., Publishers, p. 233-262.
- CALZOLARI, Nicoletta et Eugenio PICCHI (1994) : «A Lexical Workstation: From Textual Data to Structured Database», B. T. S. Atkins et A. Zampolli (dir.), *Computational Approaches to the Lexicon*, Oxford, Oxford University Press, p. 439-467.
- CHEVALIER, Monique *et al.* (1978) : *TAUM-MÉTÉO. Description du système. Janvier 1978*, Montréal, Université de Montréal, IV + 47 p.
- CHEVALIER, Monique *et al.* (1981) : «La traductologie appliquée à la traduction automatique», *Meta*, numéro spécial sous la direction de Richard Kittredge, vol. XXVI, n° 1, pp. 35-47.
- GRISHMAN, Ralph *et al.* (1994) : «Complex Syntax: Building a Computational Lexicon», *COLING 94. The 15th International Conference on Computational Linguistics. Proceedings. August 5-9, 1994, Kyoto*, COLING 94 Organizing Committee, p. 268-272.
- HEID, Ulrich (1993) : «Le lexique : quelques problèmes de description et de représentation lexicale pour la traduction automatique», Pierrette Bouillon et André Clas (dir.), *La traductique. Études et recherches de traduction par ordinateur*, Montréal, Les Presses de l'Université de Montréal et AUPELF/UREF, pp. 169-196.

- HONGYAN, Jing et Kathleen MCKEOWN (1998) : «Combining Multiple, Large-Scale Resources in a Reusable Lexicon for Natural Language Generation», *COLING-ACL '98. 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics. Proceedings of the Conference. August 10-14, 1998*, Montréal, Université de Montréal, p. 607-613.
- HUTCHINS, W. John (1986) : *Machine Translation. Past, Present, Future*, Collection Computers and their Applications, Chichester et New York, Ellis Horwood Limited et John Wiley & Sons, 382 p.
- HUTCHINS, W. John et Harold L. SOMERS (1992) : *An Introduction to Machine Translation*, Londres, San Diego, Academic Press, XXI + 362 p.
- JACQMIN, Laurence (1993) : «Classification générale des systèmes de traduction automatique», Pierrette Bouillon et André Clas (dir.), *La traductique. Études et recherches de traduction par ordinateur*, Montréal, Les Presses de l'Université de Montréal et AUPELF/UREF, pp. 43-63.
- KLAVANS, Judith L. (1990) : «Building a Computational Lexicon using Machine Readable Dictionaries», Tamas Magay et J. Zigány (dir.), *BudaLEX '88 Proceedings. Papers from the EURALEX Third International Congress, Budapest, 4-9 september 1988*, Budapest, Akadémiai Kiadó, pp. 265-279.
- KLAVANS, Judith L. et Evelyne TZOUKERMANN (1996) : «Combining Corpus and Machine-Readable Dictionary Data for Building Bilingual Lexicons», *Machine Translation*, vol. 10, n° 3-4, p. 1-34.
- LAFFLING, John (1991) : *Towards High-Precision Machine Translation. Based on Contrastive Textology*, Collection Distributed Language Translation, Berlin, New York, Foris Publications, vol. 7, XIII + 178 p.
- LEHRBERGER, John (1982) : «Automatic Translation and the Concept of Sublanguage», Richard Kittredge et John Lehrberger (dir.), *Sublanguage. Studies of Language in Restricted Semantic Domains*, Berlin et New York, Walter de Gruyter, p. 81-106.
- LEHRBERGER, John et Laurent BOURBEAU (1988) : «Machine Translation. Linguistic characteristics of MT systems and general methodology of evaluation», *Linguisticae Investigationes Supplementa. Studies in French & General Linguistics / Études en Linguistique Française et Générale*, Amsterdam, Philadelphie, John Benjamins Publishing Company, vol. 15, XIII + 241 p.

- MATSUMOTO, Yuji *et al.* (1998) : «A Computational Lexicographer's Workbench», Thierry Fontenelle *et al.* (dir.) *Actes Euralex'98. Communications soumises à EURALEX'98 (Huitième Congrès International de Lexicographie) à Liège, Belgique, Liège, Université de Liège*, p. 195-202.
- MEIJIS, Willem (1992) : «Computers and Dictionaries», Christopher S. Butler (dir.), *Computers and Written Texts*, Collection Applied Language Studies, Oxford et Cambridge (É.-U.), Basil Blackwell Ltd, p. 141-165.
- NIRENBURG, Sergei (1994) : «Lexicon Acquisition for NLP: A Consumer Report», B. T. S. Atkins et A. Zampolli (dir.), *Computational Approaches to the Lexicon*, Oxford, Oxford University Press, p. 313-347.
- ODIJK, Jan *et al.* (1994) : «Dictionaries», M. T. Rosetta (dir.), *Compositional Translation*, Dordrecht, Boston, Londres, Kluwer Academic Publishers, pp. 119-132.
- REY-DEBOVE, Josette et Alain REY (dir.) (1993) : «Chaton nm», *Le Nouveau Petit Robert*, Paris, Dictionnaires Le Robert, p. 353.
- ROBERTS, Roda P. (1997) : *Bilingual Canadian Dictionary. Bilingual Dictionary Methodology for Research Assistants*, version 8.1, Université d'Ottawa, XIV + 113 p.
- SANFILIPPO, Antonio (1994) : «Word Knowledge Acquisition, Lexicon Construction and Dictionary Compilation», *COLING 94. The 15th International Conference on Computational Linguistics. Proceedings. August 5-9, 1994, Kyoto*, COLING 94 Organizing Committee, p. 273-277.
- SEGAL, Ami (1991) : «Machine translation of natural languages: the TOVNA MTS solution: a learning system», Catriona Picken (dir.), *Translating and the Computer 12. Applying Technology to the Translation Process*, Londres, The Association for Information Management, pp. 83-109.
- TOOPE, Michael (1996) : *Examples in the Bilingual Dictionary*, Thèse de maîtrise, Ottawa, Université d'Ottawa, VII + 243 p.
- TURNER, Ronald C. *et al.* (1996) : *README.1ST. SGML For Writers and Editors*, Upper Saddle River, New Jersey, Prentice Hall PTR, XXI + 241 p.

WEHRLI, Éric (1997) : *L'analyse syntaxique des langues naturelles. Problèmes et méthodes*, Paris, Milan et Barcelone, Masson, VIII + 249 p.

WHITELOCK, Peter et Kieran KILBY (1995) : *Linguistic and Computational Techniques in Machine Translation System Design*, 2^e édition, Londres, UCL Press Limited, XII + 208 p.

Ouvrages et articles consultés

ALLEGGRANZA, Valerio *et al.* (1991) : «Linguistics for Machine Translation: The Eurotra Linguistic Specifications», Charles Copeland *et al.* (dir.), *The Eurotra Linguistic Specifications*, Collection Studies in Machine Translation and Natural Language Processing, Bruxelles, Luxembourg, Office for Official Publications of the Commission of the European Community, vol. 1, pp. 15-123.

Aslib, The Association for Information Management (1996) : *Translating and the Computer 18. Papers from the Aslib conference held on 14 & 15 November 1996*, Londres, The Association for Information Management, s.p.

----- (1997) : *Translating and the Computer 19. Papers from the Aslib conference held on 13 & 14 November 1997*, Londres, The Association for Information Management, s.p.

BALKAN, Lorna (1992) : «Translation tools», *Meta*, vol. XXXVII, n° 3, pp. 408-420.

BARNBROOK, Geoff (1993) : «The Automatic Analysis of Dictionaries. Parsing Cobuild Explanations», Mona Baker, Gill Francis et Elena Tognini-Bonelli (dir.), *Text and Technology. In Honour of John Sinclair*, Philadelphie et Amsterdam, John Benjamins Publishing Company, p. 313-331.

BOUILLON, Pierrette et André CLAS (1993) : *La traductique. Études et recherches de traduction par ordinateur*, Montréal, Les Presses de l'Université de Montréal et AUPELF/UREF, 507 p.

BRADEN-HARDER, Lisa et Wlodek ZADROZNY (1991) : «Lexicons for Broad Coverage Semantics», Uri Zernik (dir.), *Lexical Acquisition. Exploiting On-Line Resources to Build a Lexicon*, Hillsdale (New Jersey), Lawrence Erlbaum Associates, Inc., Publishers, p. 369-388.

- CHANDIOUX, John et Marie-France GUÉRAUD (1981) : «Météo : un système à l'épreuve du temps», *Meta*, numéro spécial sous la direction de Richard Kittredge, vol. XXVI, n° 1, pp. 18-22.
- CLARK, Bob (1996) : «The Logos Living Dictionary — an unprecedented internet resource», Aslib, The Association for Information Management (dir.), *Translating and the Computer 18. Papers from the Aslib conference held on 14 & 15 November 1996*, Londres, The Association for Information Management, 8 p.
- COPELAND, Charles *et al.* (dir.) (1991) : *The Eurotra Linguistic Specifications*, Collection Studies in Machine Translation and Natural Language Processing, Bruxelles, Luxembourg, Office for Official Publications of the Commission of the European Community, vol. 1, 166 p.
- DIMITROVA, Ludmila *et al.* (1998) : «Multext-East: Parallel and Comparable Corpora and Lexicons for Six Central and Eastern European Languages», *COLING-ACL '98. 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics. Proceedings of the Conference. August 10-14, 1998*, Montréal, Université de Montréal, p. 315-319.
- DOAN-NGUYEN, Hai (1998) : «Accumulation of Lexical Sets: Acquisition of Dictionary Resources and Production of New Lexical Sets», *COLING-ACL '98. 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics. Proceedings of the Conference. August 10-14, 1998*, Montréal, Université de Montréal, p. 330-335.
- DORR, Bonnie Jean (1993) : *Machine Translation: A View from the Lexicon*, Cambridge (Massachusetts), Londres, The MIT Press, xi + 434 p.
- ECKLE-KOHLER, Judith (1998) : «Methods for quality assurance in semi-automatic lexicon acquisition from corpora», Thierry Fontenelle *et al.* (dir.) *Actes Euralex'98. Communications soumises à EURALEX'98 (Huitième Congrès International de Lexicographie) à Liège, Belgique*, Liège, Université de Liège, p. 119-127.
- Eurotra (1990) : *Eurotra Reference Manual 7.0*, pp. 25-26.
- FARWELL, David *et al.* (1993) : «Automatically Creating Lexical Entries for ULTRA, a Multilingual MT System», *Machine Translation*, n° 8, pp. 127-145.

- FARWELL, David et Yorick WILKS (1993) : «ULTRA: A Multilingual Machine Translator», Sergei Nirenburg (dir.), *Progress in Machine Translation*, Amsterdam, Oxford, Washington, DC, IOS Press, pp.45-51.
- FEDERICI, Stefano (1998) : «An efficient algorithm for the automatic building of a lexicon from textual corpora», Thierry Fontenelle *et al.* (dir.) *Actes Euralex'98. Communications soumises à EURALEX'98 (Huitième Congrès International de Lexicographie) à Liège, Belgique*, Liège, Université de Liège, p. 129-139.
- FONTENELLE, Thierry *et al.* (1993) : «L'unité lexicale dans le système de traduction assistée par ordinateur METAL^{MD}», Pierrette Bouillon et André Clas (dir.), *La traductique. Études et recherches de traduction par ordinateur*, Montréal, Les Presses de l'Université de Montréal et AUPELF/UREF, pp. 364-376.
- FUCHS, Catherine *et al.* (1993) : «Traduction automatique», Catherine Fuchs (dir.) *Linguistique et traitements automatiques des langues*, Paris, Hachette, p. 193-222.
- Globalink Inc. : *Guide de l'utilisateur de Globalink Power Translator Pro pour Windows 3.1*, pp. 79-129.
- HATLEY, John (1997) : «Logos as an Internet and Intranet Application», Aslib, The Association for Information Management (dir.), *Translating and the Computer 19. Papers from the Aslib conference held on 13 & 14 November 1997*, Londres, The Association for Information Management, 20 p.
- HEID, Ulrich (1990) : «Bilingual Lexicography and Transfer Dictionaries for Machine Translation — Treating Structural Differences Between French and German», Tamas Magay et J. Zigány (dir.), *BudaLEX '88 Proceedings. Papers from the EURALEX Third International Congress, Budapest, 4-9 septembre 1988*, Budapest, Akadémiai Kiadó, pp. 107-119.
- HELMREICH, Stephen *et al.* (1993) : «Questions de traduction automatique au Computing Research Laboratory (CRL)», Pierrette Bouillon et André Clas (dir.), *La traductique. Études et recherches de traduction par ordinateur*, Montréal, Les Presses de l'Université de Montréal et AUPELF/UREF, pp. 334-363.
- HOHNHOLD, Ingo et Thomas SCHNEIDER (1991) : «Terminological Records and Lexicon Entries. A Contrastive Analysis», *Meta*, vol. XXXVI, n° 1, pp. 161-173.
- ISABELLE, Pierre (1986) : «Le projet TAUM de traduction automatique», *Revue québécoise de linguistique*, vol. 5, n° 2, p. 13-46.

- KINGSCOTT, Geoffrey (1998) : «Lernout & Hauspie – determined to be the no.1», *Language Today*, n° 9, p. 5-9.
- KNOBLOCK, Craig A. (dir.) (1996) : «Statistical versus knowledge-based machine translation», *IEEE Expert*, avril 1996, pp. 12-17.
- KNOWLES, Francis E. (1983) : «Dictionaries and Computers», R. R. K. Hartmann (dir.), *LEXeter '83 Proceedings. Papers from the International Conference on Lexicography at Exeter, 9-12 September 1983*, Tübingen, Max Niemeyer Verlag, p. 301-314.
- KUMANO, Akira et Hideki HIRAKAWA (1994) : «Building an MT Dictionary from Parallel Texts based on Linguistic and Statistical Information», *COLING 94. The 15th International Conference on Computational Linguistics. Proceedings. August 5-9, 1994*, Kyoto, COLING 94 Organizing Committee, p. 76-81.
- LANDSBERGEN, Jan (1987) : «Montague Grammar and Machine Translation», Peter Whitelock *et al.* (dir.), *Linguistic Theory and Computer Applications*, Londres, San Diego, Academic Press, pp. 113-147.
- LOFFLER-LAURIAN, Anne-Marie (1996) : *La traduction automatique*, Paris, Presses Universitaires du Septentrion, 157 p.
- MAGAY, Tamas et J. Zigány (dir.) (1990) : *BudaLEX '88 Proceedings. Papers from the EURALEX Third International Congress, Budapest, 4-9 septembre 1988*, Budapest, Akadémiai Kiadó, xi + 580 p.
- MELBY, Alan (1987) : «On Human-Machine Interaction in Translation», Sergei Nirenburg (dir.), *Machine Translation. Theoretical and Methodological Issues*, Collection Studies in Natural Language Processing, Cambridge, Londres, Cambridge University Press, pp. 145-154.
- NIRENBURG, Sergei (dir.) (1987) : *Machine Translation. Theoretical and Methodological Issues*, Collection Studies in Natural Language Processing, Cambridge, Londres, Cambridge University Press, xv + 350 p.
- (1993) : *Progress in Machine Translation*, Amsterdam, Oxford et Washington, DC, IOS Press, xiii + 320 p.

- NIRENBURG, Sergei (1993) : «L'interlangue et le traitement du sens dans les systèmes de traduction automatique», Pierrette Bouillon et André Clas (dir.), *La traductique. Études et recherches de traduction par ordinateur*, Montréal, Les Presses de l'Université de Montréal et AUPELF/UREF, pp. 91-108.
- OLIVA, Karel (1990) : «The Role of the Lexicon in a Semi-Direct MT System», Tamas Magay et J. Zigány (dir.), *BudaLEX '88 Proceedings. Papers from the EURALEX Third International Congress, Budapest, 4-9 septembre 1988*, Budapest, Akadémiai Kiadó, pp. 313-316.
- PICCHI, Eugenio *et al.* (1990) : «Implementing a Bilingual Lexical Database System», Tamas Magay et J. Zigány (dir.), *BudaLEX '88 Proceedings. Papers from the EURALEX Third International Congress, Budapest, 4-9 septembre 1988*, Budapest, Akadémiai Kiadó, pp. 317-329.
- PICKEN, Catriona (dir.) (1991) : *Translating and the Computer 12. Applying Technology to the Translation Process*, Londres, The Association for Information Management, viii + 191 p.
- PLASSARD, François (1989) : *Panorama de la Traduction Assistée par Ordinateur*, Paris, Électricité de France, Direction des études et recherches, Service Information, Prospective et Normalisation, 70 p. + annexes.
- POLLARD, Carl et Ivan A. SAG (1987) : *Information-Based Syntax and Semantics*, Palo Alto, Center for the Study of Language and Information, vol. 1, x + 233 p.
- RICO PÉREZ, Celia et Aurora MARTÍN DE SANTA OLALLA SÁNCHEZ (1997) : «New Trends in Machine Translation», *Meta*, vol. XLII, n° 4, pp. 605-615.
- RITCHIE, Graeme (1987) : «The Lexicon», Peter Whitelock *et al.* (dir.), *Linguistic Theory and Computer Applications*, Londres, San Diego, Academic Press, pp. 225-256.
- ROSETTA, M. T. (dir.) (1994) : *Compositional Translation*, Dordrecht, Boston, Londres, Kluwer Academic Publishers, xviii + 478 p.
- SANTOS, Diana (1990) : «Lexical gaps and idioms in machine translation», *COLING-90. Papers presented to the 13th International Conference on Computational Linguistics on the occasion of the 25th Anniversary of COLING and the 350th Anniversary of Helsinki University*, Helsinki, Université d'Helsinki, p. 330-335.

- SCIARONE, Abondio (1983) : «The Organization of the Bilingual Dictionary», R. R. K. Hartmann (dir.), *LEXeter '83 Proceedings. Papers from the International Conference on Lexicography at Exeter, 9-12 September 1983*, Tübingen, Max Niemeyer Verlag, p. 413-419.
- TEN HACKEN, Pius *et al.* (1991) : «Dictionaries in Eurotra», Charles Copeland *et al.* (dir.), *The Eurotra Linguistic Specifications*, Collection Studies in Machine Translation and Natural Language Processing, Bruxelles, Luxembourg, Office for Official Publications of the Commission of the European Community, vol. 1, pp. 125-160.
- TUCKER, Allen B. (1987) : «Current Strategies in Machine Translation Research and Development», Sergei Nirenburg (dir.), *Machine Translation. Theoretical and Methodological Issues*, Collection Studies in Natural Language Processing, Cambridge, Londres, Cambridge University Press, pp. 22-41.
- TURCATO, David (1998) : «Automatically Creating Bilingual Lexicons for Machine Translation from Bilingual Text», *COLING-ACL '98. 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics. Proceedings of the Conference. August 10-14, 1998*, Montréal, Université de Montréal, p. 1299-1306.
- VASCONCELLOS, Muriel (dir.) (1988) : *Technology as Translation Strategy*, Collection American Translators Association Scholarly Monograph, Binghamton, State University of New York Press, vol. II, 248 p.
- WHEELER, Peter J. (1986) : «Logos: Single-Software and Multi-Target System», Karl Kummer (dir.), *American Translators Association Conference — 1986. Proceedings of the 27th Annual Conference of the American Translators Association, Cleveland, Ohio, October 16-19 1986*, Medford (New Jersey), Learned Information, Inc, pp. 135-144.
- (1988) : «The Translator and the Dictionary Experience», Muriel Vasconcellos (dir.), *Technology as Translation Strategy*, Collection American Translators Association Scholarly Monograph, Binghamton, State University of New York Press, vol. II, pp. 149-158.
- WHITELOCK, Peter *et al.* (ed.) (1987) : *Linguistic Theory and Computer Applications*, Londres, San Diego, Academic Press, X + 329 p.
- WILSS, Wolfram (1993) : «Basic Concepts of MT», *Meta*, vol. XXXVIII, n° 3, pp. 403-413.

Dictionnaires consultés

- ATKINS *et al.* (dir.) (1993) : *Le Robert & Collins. Dictionnaire français-anglais anglais-français Senior*, 3^e édition, Paris et Glasgow, Dictionnaires Le Robert et Harper Collins Publishers, XXXVIII + 1967 p.
- CORRÉARD, Marie-Hélène et Valerie GRUNDY (dir.) (1994) : *The Oxford-Hachette French Dictionary. Le Dictionnaire Hachette-Oxford français-anglais anglais-français*, Oxford, New York, Toronto et Paris, Oxford University Press et Hachette Livre, LX + 1949 p.
- CRYSTAL, David (1991) : *A Dictionary of Linguistics and Phonetics*, 3^e édition mise à jour et augmentée, Oxford, Cambridge (Massachusetts), Blackwell Publishers, XIV + 389 p.
- DUBOIS, Jean *et al.* (1973) : *Dictionnaire de linguistique*, Paris, Librairie Larousse, XL + 516 p.
- (1994) : *Dictionnaire de linguistique et des sciences du langage*, Paris, Librairie Larousse, LX + 514 p.
- GALISSON, Robert et Daniel COSTE (dir.) (1976) : *Dictionnaire de didactique des langues*, Paris, Librairie Hachette, 612 p.
- GINGUAY, Michel (1992) : *Dictionnaire anglais-français d'informatique. Bureautique — Télématique. Micro-informatique*, 11^e édition révisée et augmentée, Paris, Masson Éditeur, 286 p.
- LEROY, Maurice (1963) : *Les Grands Courants de la Linguistique Moderne*, 3^e édition, Bruxelles et Paris, Presses Universitaires de Bruxelles et Presses Universitaires de France, X + 198 p.
- MAROUZEAU, J. (1943) : *Lexique de la terminologie linguistique. Français, allemand, anglais*, Collection Georges Ort-Geuthner, 2^e édition augmentée et mise à jour, Paris, Librairie Orientaliste Paul Geuthner, 241 p.
- MOUNIN, Georges (dir.) (1974) : *Dictionnaire de la linguistique*, Paris, Presses Universitaires de France, XXXIX + 339 p.
- REY-DEBOVE, Josette et Alain REY (dir.) (1993) : *Le Nouveau Petit Robert*, Paris, Dictionnaires Le Robert, XXXV + 2490 p.

VEDEV, Yvan (1990) : *Elsevier's Dictionary of Mathematical and Computational Linguistics*, Amsterdam, Elsevier Science Publishers B.V., 682 p.

Consultants (communications personnelles)

ADRIAENS, Geert (25 juin et 13 juillet 1998), METAL.

BENNETT, Winfield Scott (7 et 14 juillet 1998), Director of Marketing, Logos Corporation.

FARWELL, David (21 juillet 1998), Computing Research Laboratory, New Mexico State University, Directeur du projet du système ULTRA.

ISABELLE, Pierre (octobre 1998), Université de Montréal, RALI.

LUCKHARDT, Heinz-Dirk (25 et 30 juin et 20 juillet 1998), Universität des Saarlandes, responsable du système SUSY (sur Internet).

MCLAUGHLIN, Steve (13 juillet 1998), GMS (Gesellschaft für multilinguale Systeme), METAL.

SOMERS, Harold (7 juillet 1998), Professor of Language Engineering, Centre for Computational Linguistics, Department of Language Engineering, UMIST (University of Manchester Institute of Science and Technology).

SUTHERLAND, Allison (22 et 27 juillet 1998), Computational Linguist, Globalink Inc.

VALENTINI, Erwin (7 juillet 1998), European Commission, Directorate General XIII-E6, Eurotra.

VAN EYNDE, Frank (1^{er} juillet 1998), CCL KUL (Centre for Computational Linguistics KU Leuven), Eurotra.

Liste de sites Web, d'adresses électroniques et de groupes de discussion

1. Globalink : www.globalink.com
tnadeau@globalink.com, techsupp@globalink.com
2. Systran : www.systransoft.com
info@systransoft.com, admin@systranet.com
3. Météo : padina.info.umoncton.ca/chandioux/nospr.html
104213.451@compuserve.com
4. METAL (T1) : www.cbs.dk/departments/dl/ska/ovd4.htm#top
www.lhs.com
5. Logos : www.logos.it/index
scottb@logos-usa.com
dictionary@logos.it
6. Tovna : software-guide.com/cdprod1/swhrec/018/553.shtml
7. SUSY : www.phil.uni-sb.de/FR/Infowiss/projects/susystest.html
d.luckhardt@is.uni-sb.de, dlu@rz.uni-sb.de
8. Rosetta : shop.logos.it/language_today/software/rosetta/html
106350.2265@compuserve.com, fairlangtk@aol.com
9. Traduction en général (ex. TA) : www.geocities.com/Athens/7110/lantra.htm
seagate.sunet.se/archives/LANTRA-L.html

INDEX

Alex	50, 71, 83-85, 89, 90, 101, 109, 113, 115
ambiguïté	24, 33, 51, 60
analyse	8, 16, 26, 27, 29-31, 35, 40, 45, 46, 51-54, 56, 59, 66-68, 71, 118
analyse grammaticale	16, 23
analyse linguistique	38, 52, 70, 119
analyse morphologique	21, 44
analyse sémantique	50, 52, 69
analyse structurelle	67
analyse syntagmatique	16
analyse syntaxique	16, 44
arbre	25, 30, 56, 59-63
architecture	3, 28
argument	17, 46, 47, 60, 96
Ariane	34, 45, 56, 57
ATEF	56
attribut	13, 25, 53, 54, 59, 60, 75, 80, 81, 89, 90, 97, 98, 115
attribut-valeur	59, 60
base de données	4, 51
base de données lexicales	6, 63
base de données linguistiques	53, 64
base de données relationnelle	49
clé sémantique	60
clé syntaxique	59, 60
collocation	45, 50, 72, 84, 92, 94, 99, 117
combinaison libre	92, 94, 107, 110, 117
composé	57, 74, 92, 94, 112, 117
consultation du (des) dictionnaire(s)	2, 23, 25, 27, 38, 48, 51, 54, 55, 57, 61-64
corpus	7, 118
décalage	14, 16, 99
dérivation	10, 31, 59, 60
désambiguïsation	22, 38, 40, 51, 55, 62, 67
désambiguïser	55, 62
diagnostic	51, 52
dictionnairel, 3-9, 11-13, 19, 21, 23, 27, 29, 31, 32, 34-45, 48-51, 53-71, 73-77, 79, 81-84, 88-90, 92, 94-96, 98-101, 103, 106-108, 110, 114, 116, 118-121, 126	

dictionnaire automatique	2-7, 12, 20, 91, 95, 97, 99, 100, 102, 110, 112, 115, 118-122, 125, 126
dictionnaire bilingue	1, 2, 5, 8-10, 12, 21, 22, 27, 29, 30, 33, 35, 36, 39, 42, 45, 47, 49, 53, 56, 62, 63, 66, 68-71, 91, 96, 98, 100, 114, 119-121
dictionnaire bilingue automatique	2, 8, 9, 12, 22, 71, 91, 95, 96, 99-101, 107, 117, 120
dictionnaire bilingue traditionnel	2, 8, 9, 12, 91, 94-96, 100, 101, 117, 120, 121
Dictionnaire canadien bilingue (DCB)	1, 92-94, 101-107, 110, 112, 114, 115, 120, 121
dictionnaire contextuel	39, 40
dictionnaire de formes	31, 43
dictionnaire de génération	25, 31, 32, 66-69
dictionnaire de transfert	25, 30, 46, 55, 68
dictionnaire d'analyse	25, 30, 62, 63, 66-69
dictionnaire d'utilisateur	50, 88, 89, 100, 104
dictionnaire électronique	4-7, 125
dictionnaire général	38, 42, 43, 79, 91, 120
dictionnaire imprimé	4
dictionnaire pour apprenants	5
dictionnaire spécialisé	37, 38, 50, 51, 79
dictionnaire traditionnel	2, 7, 12, 20, 95, 100, 101, 107, 120, 126
dictionnaire unilingue	2, 8, 9, 12, 25, 27, 29, 35, 45, 46, 48, 49, 53-55, 63, 66, 67, 69-71, 119
dictionnaire unilingue automatique	99, 100
dictionnaire unilingue traditionnel	12
divergence	16, 17
DLT	62, 63
enseignement	52
espéranto	26, 62, 63
expression	60, 61, 92, 109, 117
expression idiomatique	40, 43, 50, 53-55, 57, 60, 92, 120
flexion	10, 57, 80, 108, 109, 116
formule	62, 63
génération	26, 27, 29, 31, 32, 39, 45, 46, 51, 53-57, 59, 66-68
génération morphologique	32, 44
génération syntaxique	32, 44
Globalink	21, 34-38, 71-77, 84, 89, 90, 101, 103, 106-108, 110, 112, 114-116, 120
glossaire	19, 47
grammaire	3, 15, 16, 68
grammaire de cas	16, 50
grammaire formelle	15, 16

hiérarchie	15
hiérarchisation	99
hiérarchisé	30, 31, 46
homographie et homographe	11, 12, 22, 39, 40, 67, 99
information grammaticale	37, 39, 47, 69, 91, 92, 102, 114
information lexicale	4, 70
information lexicographique	9
information linguistique	44
information morphologique	3, 43, 45, 54, 66, 67
information phonologique	3
information pragmatique	3
information sémantique	3, 14, 41, 45, 50, 66, 67, 95
information syntaxique	3, 4, 39-41, 43, 45, 54, 55, 66, 67, 95, 100
information traductionnelle	98, 100
intelligence artificielle	51
Intercodeur	47, 48
interlangue	25-29, 58, 59, 62, 63, 69
langage contrôlé	18, 33
lemme	11, 12, 69
Lexical Knowledge Bank (LKB)	63
lexique	4, 11, 28, 59
Logos	33, 45, 49-51, 71, 72, 83, 84, 89, 90, 95, 101, 103, 104, 106-109, 111, 113-116, 120
Metal	25, 33, 45, 46, 99
Météo	33, 35, 36, 42, 44, 120
mise à jour du (des) dictionnaire(s)	32, 37, 63, 94
mode différé	33, 34, 51
mode interactif	34, 51
module	32, 44, 46, 47, 66-68, 99
module de génération	24, 67
module de transfert	24-26, 44
module d'analyse	24, 67
module lexical	46
module morphologique	56
module syntaxique	56
Montague (grammaire)	59
morphologie	10, 31, 53, 54, 63
opérationnel	29

parsage	53, 54
polysémie et polysème	12, 18, 40, 67, 72
postédition	18, 19, 32, 33, 52
préédition	19, 32, 33, 55
processus de traduction ..	9, 21, 23, 25, 27, 32, 33, 39, 41, 44, 48, 51, 53, 55, 57, 61, 62, 64, 119
programme de transfert	25
programme de validation automatique	48
programme d'application	6
programme lexical	48
propriété sémantique	12-14, 50, 60, 89, 95, 99, 101
règle	15, 16, 31, 32, 43, 51, 52, 63, 90, 98, 99
règle d'ajustement	22
règle grammaticale	22, 23
règle morphologique	53, 54
règle sémantique	50
règle syntaxique	40
règle typographique	53
représentation	24, 25, 30-32, 56, 62, 63, 66, 67
représentation abstraite	25, 62
représentation conceptuelle	26
représentation de transfert	55
représentation du monde	26
représentation hiérarchisée	31
représentation interlangue	29, 59, 63, 70, 119
représentation intermédiaire (RI)	24, 55, 66, 67, 69, 70
représentation structurale	29
représentation syntaxique	24, 29
restrictions de sélection	14, 53, 95, 98, 99, 101, 117
rôle sémantique	46, 47
Rosetta	58-60
Semantha	50
sémantique	11, 55
séquence	52, 56
sous-catégorisation	13, 16, 68, 98, 117
sous-langage	18
structure	6, 10, 18, 25, 31, 44, 47
structure argumentale	16, 17

structure d'interface (SI)	24, 25
structure syntaxique	17, 53, 60
SUSY	45, 54, 55
syntaxe	31, 48
système basé sur l'exemple	1, 21, 118
système bilingue	29, 119
système de transfert	7, 21, 24-29, 32, 35, 45, 49, 51, 56, 58, 66, 67, 69, 70, 89, 100, 119
système direct	21-23, 35, 36, 66, 68, 70, 89, 119
système expérimental	35, 45, 54, 56, 59, 62, 119
système hybride	29
système interlangue	21, 26-29, 35, 58, 59, 62, 66, 67, 69, 70, 100, 119
système multilingue	119
système opérationnel	35, 39, 42, 45, 49, 51, 119
système «apprenant»	51
Systran	33, 35, 36, 39-41, 120
Tovna	45, 51-53
traduction automatique assistée par l'homme (TAAH)	19
traduction en mode différé	32
traduction entièrement automatique (TEA)	20
traduction humaine assistée par ordinateur (THAO)	19
traduction interactive	32, 33
trait sémantique	43, 44, 46, 47, 55
traitement de textes	21, 33
transfert	25-31, 41, 45, 51-56, 58, 63, 66, 68
transfert intermédiaire	25
transfert lexical	56, 57
transfert structurel	56, 57
typographie	53
valeur	25, 45, 57, 60, 97, 98, 100
valeur sémantique	46
valeur syntaxique	48
WOBUSU	55

INDEX DES AUTEURS ET DES CONSULTANTS CITÉS

Arnold (1993)	25
Arnold <i>et al.</i> (1994)	1, 14, 97, 98, 118
Atkins <i>et al.</i> (1994)	3, 5
Atkins et Levin (1991)	6
Bennett, W. Scott (communication personnelle)	49-51
Calzolari et Picchi (1994)	4
Chevalier <i>et al.</i> (1978)	42, 43
Chevalier <i>et al.</i> (1981)	30, 31
Grishman <i>et al.</i> (1994)	4
Heid (1993)	14, 15, 17, 68, 69, 99
Hongyan et McKeown (1998)	3, 4
Hutchins (1986)	23, 25
Hutchins et Somers (1992)	12, 13, 18, 32, 40-43, 45, 48, 55-57, 61, 64
Isabelle, Pierre (communication personnelle)	24, 26, 27, 36, 67, 68
Jacqmin (1993)	28, 34
Klavans (1990)	3-5, 7, 118
Klavans (1996)	3, 5, 7, 118
Laffling (1991)	62, 64
Lehrberger (1982)	19
Lehrberger et Bourbeau (1988)	22, 96, 97
Matsumoto (1998)	7
Meijs (1992)	6, 118
Nirenburg (1994)	3, 5, 6, 118
Odijk (1994)	59
Roberts (1997)	92
Sanfilippo (1994)	7
Segal (1991)	53
Somers, Harold (communication personnelle)	28
Sutherland, Allison (communication personnelle)	38
Toope (1996)	117
Turner <i>et al.</i> (1996)	122
Wehrli (1997)	15
Whitlock et Kilby (1995)	35, 36, 46, 47