

THÈSE

Doctorat de l'université de Nancy 2
(spécialité sciences du langage)

par

Sébastien Haton

Analyse et modélisation de la polysémie
verbale dans une perspective multilingue : le
dictionnaire bilingue vu dans un miroir

**Thèse dirigée par Jean-Marie Pierrel (Nancy 1, UMR-CNRS ATILF) et Bernard
Combettes (Nancy 2, UMR-CNRS ATILF)**

Soutenue le 25 novembre 2006

Composition du jury

Directeurs :

M. PIERREL Jean-Marie, Professeur d'Informatique, Université Henri Poincaré, Nancy,
UMR-CNRS ATILF

M. COMBETTES Bernard, Professeur de Lettres et Sciences du Langage, Université Nancy
2, UMR-CNRS ATILF

Rapporteurs :

M. CHAROLLES Michel, Professeur de Linguistique, Université de Paris III, UMR-CNRS
LATTICE, ENS Ulm

M. FRANÇOIS Jacques, Professeur de Linguistique, Université de Caen, CRISCO, FRE 2805

Examineur :

M. APOTHÉLOZ Denis, Professeur de Sciences du Langage, Université Nancy 2, UMR-
CNRS ATILF

REMERCIEMENTS

Dans le cadre institutionnel, je voudrais tout d'abord remercier l'ensemble des personnes qui m'ont aidé et soutenu tout au long de mon cursus universitaire et qui m'ont permis d'en arriver là aujourd'hui.

Merci à l'ensemble des enseignants-chercheurs de l'UFR sciences du langage de Nancy 2 pour avoir appuyé ma demande de dérogation pour une inscription en première année de DEUG et pour avoir par la suite facilité mon intégration dans leur équipe pédagogique, comme tuteur, vacataire, moniteur et enfin ATER.

Merci à l'ensemble des enseignants-chercheurs de l'UFR PLISE de Strasbourg pour m'avoir si aimablement accueilli dans leur giron l'an dernier dans le cadre d'un demi-ATER et merci également d'avoir aménagé mes horaires pour me permettre de terminer ma thèse à Nancy dans les meilleures conditions. En passant, merci à mes amis Amar, Roxane et leurs deux adorables petites filles pour m'avoir fait partager leur doux foyer une nuit par semaine durant cette période.

Merci à l'équipe pédagogique de la section Informatique de l'IUT Charlemagne pour m'avoir accepté comme moniteur pendant deux ans et merci de m'avoir fait confiance pour l'enseignement de matières pointues desquelles je n'étais pas vraiment spécialiste à mon arrivée.

Merci à tous les membres de l'ATILF sans exception. Grâce à eux, j'ai passé cinq années très sereines en matière de relations humaines et cela n'est pas pour rien dans l'accomplissement de mes recherches de doctorat. Merci Dominique pour toutes ces heures passées devant la relieuse...

Merci à mes directeurs Jean-Marie et Bernard qui m'ont fort bien accompagné pendant ces cinq années et qui ont toujours eu l'intelligence de « faire avec » mon caractère plutôt que de m'imposer leurs vues et leur mode de fonctionnement.

Merci également aux professeurs et chercheurs rencontrés au hasard des congrès et des projets et qui m'ont manifesté de la sympathie.

Enfin, merci au Ministère de la Recherche qui m'a octroyé une allocation de recherche en 2001 sur la base de mon projet de recherche de l'époque dans le cadre du programme COGNITIQUE. Que les membres de ce projet soient également remerciés.

Bien sûr, mes plus profonds remerciements et toute mon affection iront à ma famille et à mes amis qui ont tous peu ou prou contribué à me faciliter la vie pour que mes années de doctorat soient des années heureuses et non des années pesantes.

En disant « tous » mes amis et « toute » ma famille, je veux que toutes les personnes ainsi englobées se reconnaissent dans cet hommage. Toutefois, j'ai une pensée très appuyée pour mon artiste préférée Véronique Lafont dont l'aimable présence m'a donné toute l'énergie dont je manquais avant son arrivée ; une pensée également pour mes parents sans qui « rien n'aurait été possible » au double sens de l'expression (ils comprendront). Et un petit coucou hilare aux deux clowns à quatre pattes qui tous les soirs dédramatisent par leur accueil délirant les petits tracas quotidiens d'un doctorant en fin de thèse....

À Elisabeth, Gérard et Françoise qui tenaient à être présents.

TABLE DES MATIERES

REMERCIEMENTS	3
TABLE DES MATIERES	4
TABLE DES FIGURES.....	7
INTRODUCTION	8
1. Présentation du sujet : origine, problématique générale et objectifs	9
2. Organisation du document.....	22
2.1 Des verbes différents	22
2.2 Plan général	23
3. Annexes	29
CHAPITRE I.....	31
Problématique et fondements théoriques	31
1. Un sujet vaste	32
2. Sur les notions de base	33
2.1 Qu'est-ce que la polysémie ?.....	33
2.2 Synonymie.....	53
2.3 La place des lexies dans le lexique : définitions et notions	55
3. Le verbe au centre de l'étude	60
3.1 Définition	60
3.2 Pourquoi les verbes ?.....	63
3.3 La classification des verbes en linguistique.....	64
3.4 La polysémie verbale.....	65
3.5 Transition	66
4. Modèles théoriques.....	67
4.1 Modèles de traduction selon J. François.....	67
4.2 Graphes lexicaux, graphes conceptuels.....	68
4.3 Quelques théories et applications en sémantique lexicale	83
4.4 La tentation de toucher à tout.....	93
5. Transition.....	93
CHAPITRE II	95
Outils et analyse lexicographiques : Lexiques Inversés et Construction de Champs Lexicaux et Sémantiques ...	95
1. Introduction de partie	96
1.1 Présentation	96
1.2 Quelques difficultés à surmonter.....	97
2. Le lexique inversé.....	98
2.1 Genèse	98
2.2 Construction des lexiques inversés (LI)	98
3. Exploitation des données des dictionnaires	99
3.1 Retour sur les principes d'origine.....	99
3.2 Détail des entrées dans nos dictionnaires de référence.....	100
4. Un exemple flagrant de dissymétrie : les traductions du verbe <i>aborder</i> en italien	107
4.1 Préparation française du verbe, pour <i>aborder</i> <i>aborder</i>	107
4.2 Traductions.....	112
5. De la polysémie faible à l'explosion combinatoire : le verbe abandonner	120
5.1 Polysémie faible	120
5.2 Un champ lexical large.....	121
5.3 Les traductions d'abandonner.....	123
6. L'intégration des locutions et périphrases verbales dans le champ de l'étude	127
6.1 Intégration des périphrases verbales dans les champs lexicaux : exemple à partir de <i>se mettre</i> à en français.	127
6.2 Travail préliminaire sur <i>se mettre</i> à.....	129

6.3	Intégration par technique de « balayage »	132
6.4	Bilan de l'intégration des périphrases à l'étude, et des lexies aux champs lexicaux de façon générale	133
7.	Prolongement des LI.....	134
7.1	La traduction intermédiaire : Paraphrase ou pas paraphrase ?.....	134
7.2	Représentation des lexiques multilingues issus du LI. Projet de superposition de champs sémantiques en plusieurs langues.	135
8.	Transition avant étude sur corpus.....	140
8.1	Le statut des traductions dans le miroir	140
8.2	Les limites de l'observation des dictionnaires.....	141
9.	Les verbes « de parole » en contexte.....	143
9.1	Quelques mots sur la linguistique de corpus	143
9.2	Traitement des verbes de parole	148
CHAPITRE III.....		161
Étude autour du verbe <i>traverser</i> . Compléments et complémentarité des données contextuelles et lexicographiques.....		161
1.	La version lexicographique de <i>traverser</i>	162
1.1	Des acceptions de <i>traverser</i>	162
1.2	Des traductions de <i>traverser</i>	165
1.3	Comparaison LI / LD en anglais.....	166
2.	Vers une analyse plus approfondie de <i>traverser</i> en contexte	169
3.	Pourquoi compléter les données par le corpus	172
3.1	Une question qui divise ou qui anime la communauté	172
3.2	Le plus beau corpus du monde ne peut donner que ce qu'il a	173
4.	Traverser dans le corpus.....	174
4.1	Introduction à la traversée	174
4.2	Diagrammes ontologiques des traversées à partir de la base littéraire	176
4.3	Mouvement et bornage dans les traversées	184
4.4	Modélisation.....	199
4.5	Analyse contrastive	211
4.6	Conclusion de la traversée.....	223
CHAPITRE IV.....		228
Modélisation et algorithme de construction du graphe des Champs Sémantiques Multilingues Unifiés ou CSMU		228
1.	Objectifs de la modélisation et Exploitation des données	229
1.1	Problématique originelle	229
1.2	Problématique enrichie.....	230
1.3	Méthodologie et Travaux Préliminaires	231
1.4	Construction des bases de données.....	238
2.	Sur la notion de synonymie	240
2.1	Préliminaires.....	240
2.2	Les niveaux de synonymie	241
3.	Recensement des variables et lecture du tableau.....	246
3.1	Étiquettes de mot.....	246
3.2	Liens de synonymie.....	246
3.3	Synonymes et quasi-synonymes.....	247
3.4	Collocations sujet.....	247
3.5	Collocations objet.....	248
3.6	Constructions syntaxiques indirectes.....	249
3.7	Indices non langagiers.....	250
4.	Unification des contextes	250
4.1	Pourquoi unifier.....	250
4.2	Comment unifier.....	251
5.	Écriture de l'algorithme	252

5.1	Choix de la procédure de mise en œuvre.....	252
5.2	L'algorithme.....	272
5.3	Atouts et perspectives.....	290
	Conclusion.....	292
	Bilan final, Perspectives et Obstacles.....	292
1.	Rappel des objectifs et des résultats.....	293
2.	Contribution pour le dictionnaire du XXIème siècle.....	294
3.	Des difficultés en grand nombre.....	297
4.	Conclusion finale.....	299
	Glossaire.....	300
	Bibliographie.....	301
	Articles et Ouvrages.....	301
	Dictionnaires.....	315
	Quelques Pages Internet.....	316

TABLE DES FIGURES

Figure 1 : Graphe de « Prix Goncourt » par John Sowa (1992).....	70
Figure 2 : Chemin dit « d'activation »	80
Figure 3 : Graphique récapitulatif, seconde version. Mise en exergue de la dissymétrie.....	115
Figure 4 : Tableau récapitulatif du verbe aborder	116
Figure 5 : Tableau des correspondants anglais du verbe ABANDONNER : première représentation du lexique inversé français-anglais.....	125
Figure 6 : Fragment du tableau de données autour de <i>aborder</i>	137
Figure 7 : Fragment du tableau de données autour de <i>abandonner</i>	138
Figure 8 : Fragment du tableau de données autour de <i>se mettre (à)</i>	139
Figure 9 : Fragment de CSMU pour les verbes de parole à partir de données attestées.....	159
Figure 10 : tableau de correspondance des traductions anglaises de <i>traverser</i>	167
Figure 11 : Fragment du tableau de données autour de <i>traverser</i>	168
Figure 12 : traversée dans la largeur	177
Figure 13 : traversée dans la longueur	178
Figure 14 : traversée angulaire.....	179
Figure 15 : transpercée dynamique	179
Figure 16 : traversée de l'extérieur vers l'intérieur	180
Figure 17 : traversée de l'intérieur vers l'extérieur.....	180
Figure 18 : traversée de l'intérieur vers l'intérieur avec objectif	181
Figure 19 : traversée en intérieur sans objectif.....	181
Figure 20 : transpercée statique.....	182
Figure 21 : traversée statique en largeur	183
Figure 22 : traversée semi-dynamique	183
Figure 23 : typologie aspectuelle	192
Figure 24 : Modèle 1 pour traverser.....	202
Figure 25 : Modèle 2 pour traverser.....	203
Figure 26 : Modèle 3 pour traverser.....	205
Figure 27 : Modèle 4 pour traverser.....	207
Figure 28 : Modèle 5 pour traverser.....	208
Figure 29 : Modèle 5bis pour traverser	209
Figure 30 : Fragment de CSMU pour <i>traverser</i> en contexte.....	225
Figure 31 : mise en forme du fragment d'entrée "to leave"	236
Figure 32 : seconde mise en forme d'un fragment d'entrée de to leave	253
Figure 33 : fragment sélectionné de <i>parler</i>	253
Figure 34 : Ligne sélectionnée de <i>to leave</i>	259
Figure 35 : Fragment sélectionné de <i>to leave</i>	266
Figure 36 : Typologie des constructions syntaxiques associées au verbe	268

INTRODUCTION

1. Présentation du sujet : origine, problématique générale et objectifs

L'objectif de recherche que nous nous sommes fixé est le suivant : à partir de l'analyse de la polysémie des verbes dans un cadre lexicographique et sur la base des lexiques miroirs dont nous expliquerons précisément le principe, nous avons cherché à établir une méthodologie de récupération de l'intégralité des données de dictionnaires bilingues pour les inclure dans des modèles à des fins d'exploitation informatique. En constatant certaines lacunes naturelles des dictionnaires, et ainsi en admettant que ces derniers ne représentent pas la langue dans son intégralité, le cadre de recherche s'est naturellement élargi vers des analyses linguistiques sur corpus, en l'occurrence sur la base littéraire FRANTEXT, dans l'unique but de montrer que les faits de langue observés dans la littérature sont complémentaires des données des dictionnaires et qu'ils peuvent être modélisés dans le même espace que ces dernières. Nous avons travaillé sur trois dictionnaires bilingues, français-anglais, français-espagnol et français-italien et la catégorie verbale a été le matériau central de notre travail. Il faut également noter que toutes nos études sont centrées sur le français et sont destinées à l'informatisation, conformément à la politique scientifique de l'ATILF qui est un laboratoire d'analyse et de traitement informatique de la langue française.

Pour mener à bien cet objectif assez vaste, nous avons procédé en trois étapes, lesquelles correspondent au plan général de la thèse :

Tout d'abord, nous discutons de la polysémie dans un cadre lexicographique et nous essaierons de montrer quelles sont les ressources réelles des dictionnaires bilingues ainsi que leurs principales lacunes, pour la plupart inévitables. Dans un deuxième temps, nous présenterons des travaux sur données littéraires destinés à pallier quelques-unes de ces lacunes inévitables des dictionnaires pour montrer que les dictionnaires et les corpus peuvent être complémentaires dans certaines circonstances. Et enfin, nous présenterons les modèles que nous avons conçus à partir de nos précédentes analyses ainsi que l'algorithme de programmation qui permet de transformer ces modèles en graphes lexicaux exploitables informatiquement.

Le premier travail que nous avons effectué a été de faire une analyse critique et descriptive des dictionnaires bilingues, de leur contenu, de leurs structures et surtout de ce qu'ils recèlent réellement en termes de liens et d'informations cachés. À titre d'exemple, dans un seul extrait de l'entrée *to leave* tiré du Robert & Collins français-anglais dernière édition, il y a 28 types d'informations différents que nous avons relevés et décrits dans cette thèse, ce qui est beaucoup :

leave /li:v/ (vb : prêt, ptp **left**)

VI a (= *go away from*) [+ *town*] quitter, partir de ; (*permanently*) quitter ; [+ *room, building*] sortir de, quitter ; [+ *person, job, one's husband, wife*] quitter ; [+ *one's children*] aban-

donner ♦ **he left Paris in 2001** il a quitté
Paris en 2001 ♦ **we left Paris at 6 o'clock**
nous sommes partis de Paris *or* nous
avons quitté Paris à 6 heures...

(...)

e (*Math*) **three from six ~s three** six moins
trois égalent trois...

f (*in will*) [+ *money*] laisser {*to à*} ; [+ *object*,
property] laisser, léguer {*to à*}...

VI (= *go away*) [*person, train, ship etc*] partir,
s'en aller ; (= *resign*) partir, démissionner

♦ **to ~ for Paris** [*person, train*] partir pour
Paris ; [*ship*] partir *or* appareiller pour
Paris...

▶ **leave off** VI (*= *stop*) s'arrêter ♦ **where did
we ~ off?** (*in work, reading*) où nous sommes-
nous arrêtés? ♦ **~ off !** arrête !, ça suffit ! *

VT SEP a (*= *stop*) arrêter (*doing sth* de faire
qch)...

▶ **leave out** VT SEP a (= *omit*) (*accidentally*)
oublier, omettre ; (*deliberately*) exclure ;
[+ *line in text*] (also *Mus*) [+ *note*] sauter...

La lecture d'un tel dictionnaire ressemble presque à un exercice de spécialistes d'autant qu'il faut se familiariser avec les typographies différentes choisies pour caractériser chacun des types. Ce qui nous intéresse prioritairement dans cet extrait et dans le reste du dictionnaire c'est : quelles sont les informations nécessaires pour caractériser un lien de traduction et comment les modéliser. L'autre grand intérêt de cet extrait est qu'il nous apprend beaucoup de choses en plus de nous donner les traductions de *to leave*. Si nous observons à nouveau le début de l'extrait, après les informations phonétiques et grammaticales, on observe que *quitter et partir de* sont assimilables à des synonymes l'un de l'autre lorsqu'ils traduisent *to leave* avec comme objet direct *ville*. De même *to go away from* est présenté comme un synonyme de *to leave* dans les mêmes conditions d'emploi et par voie de conséquence comme une traduction potentielle de *quitter et partir de* et pouvant être traduit par eux. Ainsi, cette simple relation de traduction crée deux liens de synonymie et deux liens de traduction qui sont des informations que nous avons appelées « cachées ou indirectes » dans notre document, mais ce ne sont pas les seules. À la lecture de ce fragment, il est permis de penser que la présence de ces données cachées est un phénomène généralisé et qu'un simple dictionnaire bilingue contient également et dans une certaine mesure deux dictionnaires de synonymes monolingues. En tout état de cause, il faut que ces liens apparaissent clairement dans notre modélisation.

Mais la richesse cachée des dictionnaires ne s'arrête pas là et nous allons maintenant exposer brièvement le principe des lexiques miroirs qui est une manière de

réconcilier les deux parties d'un même dictionnaire bilingue car leur confrontation montre dans les faits que ce sont vraiment deux dictionnaires très différents.

Pour bien illustrer le fonctionnement des lexiques miroirs ou inversés, nous utiliserons à nouveau l'entrée du verbe *to leave*. Ce que nous appelons le lexique droit (le LD) de *to leave* est l'ensemble des traductions françaises de *to leave* à l'entrée de celui-ci dans la partie anglais-français du dictionnaire et qui contient donc entre autres *quitter* et *partir de*. Maintenant, ce que nous appelons le lexique miroir ou inversé (le LI) de *to leave* correspond à l'ensemble des verbes français pour lesquels *to leave* est une des traductions potentielles dans la partie français-anglais du dictionnaire. En tant qu'utilisateur moyen, nous nous attendions lors de la confrontation des lexiques droit et inversé à observer une certaine symétrie entre les deux. Or, nous avons constaté que les cas de symétrie pour les traductions entre lexèmes ne représentent que de 20 à 50 % du total. En d'autres termes, lorsqu'un verbe est proposé comme traduction d'un autre dans une partie d'un dictionnaire bilingue, il y a moins d'une chance sur deux que la réciproque soit vraie dans l'autre partie du dictionnaire. Pour effectuer cette comparaison, nous avons construit trois lexiques miroirs complets entre le français et l'espagnol, le français et l'anglais et le français et l'italien, ce qui représente 9000 formes verbales du français et leurs correspondants inversés dans les trois autres langues. Cette dissymétrie est encore plus flagrante si l'on essaie de rattacher les justes acceptions du verbe étudié à ses traductions. Nous avons fait ce travail à partir de quelques verbes français, dont *traverser* qui bénéficiera d'un traitement tout particulier. Tout d'abord nous avons extrait du dictionnaire Le Petit Robert toutes les acceptions de *traverser* et nous les avons numérotées. Ensuite, nous avons fait le relevé des verbes anglais qui entretiennent un lien de traduction avec lui et nous avons associé à chaque lien l'indice correspondant. La deuxième colonne du tableau que l'on peut observer à la page 167 correspond au lexique droit de *traverser* et il contient le relevé des indices se rapportant aux acceptions de *traverser* si le verbe anglais de la première colonne est proposé comme traduction de *traverser*. À l'inverse, la troisième colonne qui correspond au lexique miroir de *traverser* contient ces indices lorsque c'est *traverser* qui est la traduction du verbe anglais. On peut observer dans ces conditions d'analyse qu'il y a un seul cas de symétrie dans ce relevé, celui de la correspondance entre *traverser* et *to occur to* ce qui représente ici un cas sur quarante-deux. L'étude statistique globale n'a que peu d'intérêt, ce que nous retenons est surtout cette très grande asymétrie à l'intérieur d'un même dictionnaire. Même si nous admettons que la symétrie ne peut être la règle, il s'agit tout de même d'un phénomène d'une ampleur très inattendue. Une fois encore il faudra en rendre compte dans les modèles.

Ces deux constats ont eu deux conséquences directes sur notre recherche, l'une entraînant l'autre : tout d'abord, la multiplicité des liens réellement présents dans le dictionnaire ainsi que la difficulté de décrire l'asymétrie au cas par cas par les acceptions font qu'il est quasiment impossible de continuer à avoir une approche sémasiologique du problème, c'est-à-dire de continuer à travailler à partir des lexèmes

polysémiques et de leurs dégroupements en acceptions. Il a donc fallu rechercher une démarche plus onomasiologique, c'est-à-dire qui traite d'abord de la signification pour remonter vers le mot. C'est pourquoi nous avons introduit dans le sujet la notion de lexie au sens que Pottier puis Mel'çuk et son équipe lui ont donné, à savoir un mot pris dans une acception bien spécifique. Dans le cadre de notre propre modélisation, une lexie est un mot muni d'un faisceau de traits qui la distinguent de toutes les autres lexies du même mot. Voyons toujours notre exemple de *to leave*, ces traits peuvent être une traduction, un ou plusieurs synonymes liés à la traduction, éventuellement des restrictions sur le choix des collocations sujet et objet, sur les constructions syntaxiques voire sur le registre, le domaine d'emploi et sur certaines informations extra-langagières. Bref une lexie dans nos modèles est caractérisée par tout ce qui est jugé pertinent pour caractériser une relation de traduction dans un dictionnaire bilingue.

Néanmoins, un certain nombre de contraintes pour la plupart liées au format de l'édition empêchent les ouvrages lexicographiques d'offrir une visibilité entièrement satisfaisante des phénomènes linguistiques qui préexistent à leur création. Il est également évident qu'un dictionnaire grand public se doit de refléter la langue académique, et accessoirement les faits de langue qui sont présentés dans les dictionnaires monolingues du même éditeur. Il est peu probable qu'un dictionnaire sérieux se permette de répertorier minutieusement les créations des auteurs littéraires en les présentant comme des faits de langue établis, notamment pour des ouvrages qui sont destinés à des locuteurs non natifs. C'est pourquoi nous avons été amené à poursuivre nos réflexions par des analyses linguistiques débarrassées des contraintes liées au cadre lexicographique, en nous fondant exclusivement sur des données attestées extraites de la base littéraire FRANTEXT. En particulier, nous avons relevé deux phénomènes linguistiques que les dictionnaires bilingues ne peuvent pas traiter de par leur nature et c'est ce que nous essaierons d'explicitier dans la partie de ce document qui suit l'analyse critique des dictionnaires.

Tout d'abord, il est très difficile de sélectionner des traductions concurrentes pour une même acception en fonction de la nature sémantique des arguments du verbe, voire en fonction du co-texte un peu plus éloigné, ce dernier étant moins dépendant du verbe mais participant également à sa caractérisation sémantique et aspectuelle. Même dans le dictionnaire grand public le plus élaboré (comme le Robert & Collins), le problème demeure car l'ouvrage ne propose que le choix d'un objet ou d'une classe d'objets sans entrer plus en détails dans sa nature sémantique. De plus, le dictionnaire ne propose jamais plus d'une information collocationnelle à la fois : si l'objet est précisé, le sujet ne l'est pas, et inversement. C'est pourquoi nous avons décidé de nous pencher sur le verbe *traverser* et son co-texte, dans le cadre d'un travail collaboratif au sein de l'ATILF.

Le verbe *traverser* est fortement dépendant de son environnement co-textuel quant à sa caractérisation aspectuelle, ce qui le rend difficile à traduire si ce co-texte

n'est pas précisément étudié. Dans les études générales sur ce verbe comme celles de Muller & Sarda, Victorri ou celle de Vendler pour son équivalent anglais *to cross*, traverser est systématiquement traité comme un verbe dit d'accomplissement, c'est-à-dire dont l'emploi suppose que le sujet qui traverse passe d'un côté à l'autre de l'objet traversé, les bornes spatiales de celui-ci étant supposées dépassées à l'issue de la traversée. On retrouve cette vision prototypique dans les dictionnaires puisque le Petit Robert donne les définitions du verbe en respectant à la lettre ce trait aspectuel d'accomplissement, de même que les dictionnaires bilingues. Or, un premier passage en revue du corpus littéraire montre que *traverser* peut prendre d'autres valeurs aspectuelles selon ses emplois et que ces valeurs aspectuelles sont calculables à partir de plusieurs facteurs : le sens du verbe lui-même, le co-texte extra-prédicatif qui fixe les bornes de l'action (hors ou à l'intérieur de l'objet traversé), la nature spatiale des arguments du verbe et en dernier lieu la vision subjective du locuteur, ou plus précisément la manière dont le locuteur perçoit l'objet traversé. En combinant tous ces facteurs nous sommes parvenus à une modélisation des occurrences de *traverser* qui les fait tous apparaître dans des graphiques mathématiques pourvus d'un axe spatial et d'un axe temporel. En transposant les occurrences de *traverser* dans ces modèles, nous avons pu prouver que le verbe rentre potentiellement dans les quatre catégories aspectuelles majeures de Vendler (accomplissement, achèvement, activité et état). La prise en compte de ces facteurs peut avoir une incidence sur le choix de la meilleure traduction pour une occurrence donnée. Sans entrer dans le détail de notre analyse contrastive, nous citerons sans le détailler un exemple, construit celui-là, qui illustre ce que nous venons de dire :

Jean a traversé la forêt dont la traduction générique peut être *John crossed the forest*.

Traverser peut ici être traduit en anglais par des prédicats contenant chacune des trois prépositions habituelles pour le traduire (*across*, *through* et *over*). Trois interprétations sont possibles, chacune de ses interprétations étant soumise aux facteurs dont nous faisons la liste plus haut. Le choix de la bonne traduction pour cette phrase dépend directement de leur prise en compte minutieuse. Lorsque le co-texte et certains éléments liés à l'aspect et au contenu sémantique des arguments sont pertinents et aisément repérables, nous pouvons en effectuer la saisie dans la base de données. Nous reviendrons sur celle-ci au moment de parler de la modélisation.

La deuxième chose que les dictionnaires ne peuvent pas faire, et pour cause, c'est montrer ce qu'ils ne montrent pas. En effet, nous l'avons déjà dit, les dictionnaires destinés au grand public sont avant tout des ouvrages académiques qui se doivent de représenter la langue telle qu'elle est normativement établie à un moment du temps. Ceci exclut de pouvoir y trouver des sens de mots originaux ou des constructions syntaxiques qui n'auraient pas été validées par les lexicographes comme faisant partie intégrante de la langue. Or, l'observation des œuvres littéraires fait clairement apparaître que les sens et constructions « inhabituels » ou « inattendus »

sont loin d'être des phénomènes résiduels et qu'ils ne sont pas non plus toujours récents ni isolés. À ce sujet, les auteurs littéraires font la preuve que la langue n'est pas seulement un système figé mais surtout un socle normatif à partir duquel de nombreuses fantaisies d'expression sont possibles. L'émergence de sens nouveaux pour les verbes eux-mêmes est parfois, il est vrai, difficile à décrire. En revanche, il est plus aisé de répertorier les constructions syntaxiques nouvelles (surtout si elles se répètent dans des œuvres littéraires différentes) de même que quelque originalité dans la nature sémantique des arguments du verbe. Les verbes qui font partie du vaste paradigme des verbes de parole¹ sont particulièrement malléables de ce point de vue et c'est ce qui nous a conduit à en proposer l'étude à partir de FRANTEXT dans le but avoué d'enrichir les bases de données d'origine lexicographique sans en modifier la structure. Notre travail sur ces verbes a été effectué dans le cadre du PDR ILF « transitivité ». À partir d'observations sur l'opposition entre verbes de parole transitifs et intransitifs, nous avons formulé l'hypothèse que la plupart des verbes classés intransitifs devaient pouvoir être employés dans des formes transitives avec une certaine fréquence. Pour effectuer une analyse complète des occurrences des verbes de parole, nous avons élaboré une grille d'analyse syntaxique et sémantique qui puisse les décrire toutes. Le codage de 2300 occurrences littéraires dans ces grilles a montré que notre hypothèse de départ était justifiée pour un grand nombre de verbes alors même que les dictionnaires ne le montrent pas, ou très peu et très lentement. À titre d'exemple, nous citerons le cas du verbe *aboyer* qui n'est répertorié comme verbe de parole dans le *Robert & Collins* que sous la forme verbe intransitif équivalent de *crier*. Dans les textes de Frantext postérieurs à 1960, les formes transitives *d'aboyer* en tant que verbe de parole sont très nettement majoritaires et se rencontrent avec les constructions syntaxiques les plus variées. Bien sûr, on peut toujours objecter que ces formes sont accidentelles et ne reflètent que la propension des locuteurs d'une langue à en façonner le matériau pour des raisons stylistiques ou de méconnaissance de la langue. Cependant, ce serait nier l'ampleur du phénomène et c'est ce qui nous a conduit à intégrer ces faits de langue dans nos modèles. Le fait qu'aucune traduction ne soit proposée initialement pour ces occurrences *d'aboyer* n'est en soi pas problématique pour notre travail d'ensemble. En effet, nos bases de données sont conçues pour contenir des lexies, c'est-à-dire des emplois bien particuliers d'un mot mais elles n'impliquent pas que ces lexies sont nécessairement associées préalablement à une traduction. En l'occurrence, les lexies qui seront saisies pour *aboyer* auront comme traits renseignés les collocations sujet et objet, la construction syntaxique du mot source et éventuellement une information complémentaire permettant de voir que l'on a affaire à l'emploi métaphorique du verbe d'origine.

La finalité de ces travaux sur corpus est d'observer la réalité du langage sans contraintes *a priori* afin de ramener la description de celui-ci aux intentions « réelles »

¹ En prenant toutes les précautions qui s'imposent pour désigner cet ensemble de verbes atypique et très hétérogène. Voir à la fin du chapitre II.

des auteurs ou à tout le moins à s'en rapprocher le plus possible. Nous espérons montrer à partir de ces analyses comment l'observation de données attestées peut compléter ou enrichir nos modèles sans en bouleverser la structure dans le but de réunir des informations lexicales de sources différentes dans une même base d'informations qui est finalement une banque de données.

Le dernier volet de notre travail concerne la modélisation proprement dite des lexies repérées dans les dictionnaires et dans les corpus littéraires. La première étape de la modélisation est la transformation en bases de données des dictionnaires bilingues. Nous laisserons de côté pour l'instant les occurrences littéraires car la saisie des lexies qu'elles induisent n'est probablement pas automatisable. Revenons à notre fragment d'extrait de *to leave*. Comme nous l'avons dit, ce fragment contient outre un lien de traduction plusieurs liens de synonymie et de traduction présumée qu'il va falloir ordonner. Pour en rendre compte, nous avons effectué une nomenclature de liens de synonymie, lesquels se retrouvent dans l'algorithme final. Trois de ces liens sont directs (traduction directe, synonymie monolingue et traduction réciproque), trois autres sont indirects (synonymie partielle en langue source, partielle en langue cible et traduction indirecte).

Tout le travail de modélisation d'une lexie réside dans le codage de toutes les informations linguistiques présentes dans la ligne. En l'occurrence, notre fragment induit la définition de plusieurs liens et structures pour les lexies présentes :

- Deux liens de synonymie de traduction directe (ST) entre *to leave* et *laisser* et entre *to leave* et *partir de*
 - Un lien de synonymie partielle en langue source (SPS) entre *to leave* et *to go away from*
 - Un lien de synonymie partielle en langue cible (SPTC) entre *laisser* et *partir de*
 - Deux liens de synonymie indirecte (SI) entre *to go away from* et *laisser* et entre *to go away from* et *partir de*.
 - Une structure de traits pour chacune des quatre lexies *to leave*, *to go away from*, *quitter* et *partir de* dont la plus complète est celle de la lexie source *to leave* qui contient :
 - Son lemme : *to leave*
 - Sa langue: Anglais
 - Sa catégorie grammaticale: Verbe
- Ces trois critères sont communs à toutes les lexies du même mot *to leave*.
- Elle contient également la collocation sujet (pas renseignée donc valeur par défaut qui est théoriquement « sans importance d'un point de vue sémantique »), la collocation objet (ici *house*), la construction syntaxique (pas renseignée donc transitif direct), le registre (pas renseigné donc « neutre ») et les informations complémentaires (absentes).

- Trois autres structures sont créées pour les trois autres lexies selon un procédé un peu différent car les informations qui les concernent ne sont pas forcément au même endroit dans la base.

Et bien entendu, pour permettre l'exploitation informatique de tous les objets qui viennent d'être décrits, les liens et les structures sont munis automatiquement d'indices numériques permettant de les réunir ultérieurement par l'intermédiaire du programme de création de graphes.

Voilà par l'exemple ce que l'algorithme que nous avons conçu fait sur l'ensemble de notre base de données lexicales en la traitant ligne par ligne. Pour justifier l'existence de plusieurs sources de données distinctes, et notamment la source des corpus, l'algorithme prévoit que lorsque les cases mot-source ou synonymes ne sont pas renseignées, il ne crée qu'une structure pour la lexie sans lien de synonymie.

L'algorithme de création de graphes « multilexical », car nous n'osons dire pleinement multilingue, a été programmé tel quel quelques semaines avant la fin de la thèse. Nous n'avons évidemment pas eu le temps de tester ses potentialités mais nous savons déjà que la transformation des bases de données en graphes est acquise. D'autre part, la perspective de disposer de dictionnaires électroniques correctement balisés nous autorise à espérer l'automatisation complète de toute la chaîne de transformation du dictionnaire en base de données puis de la base de données en graphe. La conservation de l'étape intermédiaire « Base de Données » nous semble très importante car c'est le seul espace commun à la fois aux données qui sont automatisables et à celles qui ne le sont pas (notamment les données littéraires originales). La réunion des deux types de données constitue une bibliothèque ou banque lexicale de grande ampleur, ouvrant des perspectives particulièrement riches.

Pour en finir avec la problématique générale, nous souhaitons dire quelques mots sur les potentialités du travail d'informatisation réalisé à l'issue de notre recherche. Tout d'abord, si notre algorithme est capable de transformer un dictionnaire en graphe, l'inverse est également vrai moyennant quelques aménagements mineurs. De ce point de vue, on peut dire que notre méthode permettra la génération automatique de dictionnaires selon un paramétrage très souple. Par exemple, admettons que nous avons saisi l'intégralité de notre dictionnaire bilingue et que nous voulons le restituer sous une forme parfaitement symétrique. Deux lignes de code informatique suffisent pour demander à ce que tous les liens de synonymie de traduction soient symétrisés et le dictionnaire bilingue symétrique est prêt à être généré.

Les applications les plus ambitieuses que nous envisageons sont de deux types : nous espérons d'abord pouvoir intégrer dans les bases de données des informations issues de dictionnaires monolingues. Enfin, nous espérons pouvoir générer, au moins partiellement, des dictionnaires inédits à partir de ceux qui auront déjà été transformés. Par exemple, admettons encore que nous avons saisi un dictionnaire français-anglais et un dictionnaire français-syldave. À partir de cela, nous souhaitons réfléchir aux

moyens informatiques à mettre en œuvre pour générer un dictionnaire anglais-syldave complètement inédit parce que jamais personne n'a entrepris de rédiger un tel dictionnaire.

Enfin, nous terminerons cette présentation en citant deux obstacles majeurs :

Le premier est que nous n'avons travaillé que sur des langues relativement proches, facilitant leur mise en rapport dans des champs lexicaux communs. En revanche, nous n'avons pas du tout réfléchi à la comparaison entre le français et des langues au fonctionnement très différent, notamment au niveau des verbes.

Le deuxième problème est que nous n'avons travaillé que sur la catégorie verbale et ses satellites. À ce stade nous ignorons totalement ce que cela induirait en termes de profondes modifications dans la structure de la base de données que d'y inclure des lexies d'autres catégories grammaticales.

En premier lieu et avant les descriptions sus-mentionnées, l'étude est censée respecter :

- Le principe d'exhaustivité

Ce principe est double, il touche tout autant le traitement lexicographique du lexème dans sa langue que son paradigme d'équivalents en langue cible.

Par ce principe, nous imposons d'abord la prise en compte de toutes les entrées lexicales d'une unité verbale qu'elles soient répertoriées ou non dans les dictionnaires monolingues. D'autre part, une traduction présente dans le dictionnaire est une traduction à prendre en compte, quelle que soit la manière dont elle apparaît. La récupération d'une nouvelle entrée lexicale est une opération délicate car l'arbitraire du lecteur est ici mis en jeu. Par ce principe, nous proposons de ne pas borner le dégroupement de sens à celui que présenterait un dictionnaire monolingue, mais au contraire de l'ouvrir si besoin est en fonction de sens nouveaux qui émergent du fait de l'évolution de la langue et/ou de ce que proposent les dictionnaires bilingues et les corpus littéraires.

- Le principe de limpidité

Nous entendons par limpidité le principe qui stipule que la lecture et la justification de chaque entrée lexicale doivent être décodables sans effort particulier. En d'autres termes, le locuteur non natif d'une des deux langues confrontées n'est pas censé hésiter sur l'acception à laquelle renvoie le mot qu'il ne connaît pas. Pour ce faire, un lexique limpide doit faire apparaître autant d'informations linguistiques qu'il est nécessaire, en mettant de côté les données non pertinentes ou surnuméraires, lesquelles brouillent la perception de l'utilisateur.

- Le principe de réécriture

Ce principe est à la base de tout travail lexicographique selon Dubois & Dubois (1971:84) : il s'agit tout simplement de la règle de la paraphrase.

En lexicographie bilingue, il conditionne le transfert d'une unité lexicale polysémique vers une vedette donnée par une paraphrase en langue-source. Le travail de paraphrase n'est pas nécessaire dans la perspective d'une implantation informatique de nos lexiques puisque les différences au sein de ceux-ci se font par les informations linguistiques extra-lexicales. La réécriture – voire la reformulation mentale – demeure malgré tout un exercice obligatoire pour décoder une unité polysémique.

- Les principes de sous-catégorisation syntaxique et sémantique

Une entrée lexicale pourra être divisée en sous-entrées lexicales en fonction des contraintes de sélection des arguments et des constructions particulières. Les disparités potentielles dans les collocations sujet et complément ainsi que des différences de la structure syntaxique du syntagme verbal (SV) entraînent parfois un dégroupement très fin des verbes qui en sont pourvus. Nous appellerons ce principe le principe de sous-catégorisation (cf. les collocations du verbe *abandonner* comme discriminateurs forts au chapitre II) ;

Chaque traduction d'un verbe est associée à un ou plusieurs types de construction syntaxique qui peut être la transitivité directe par défaut. En outre, des collocations génériques sont souvent proposées à l'appui de la traduction. Ce sont généralement ces critères qui permettent au lecteur de déterminer à quelle acception du mot vedette correspond(ent) le ou les candidat(s) proposé(s).

Ces critères accompagnent le principe d'exhaustivité en cela qu'une entrée lexicale pourra être divisée en sous-entrées lexicales en fonction des contraintes de sélection des arguments et des constructions particulières.

- Le principe de valeur de transfert

Lors d'une traduction, les facteurs qui entrent en jeu pour le passage d'une langue à une autre (le « transfert ») sont de différents types. Outre les contraintes de sélection des arguments, cela concerne la hiérarchie des traits définitoires (registre de langue, idiolecte, genre, paradigme de conjugaison, etc.) et les traits nécessaires pour la traduction (polysémie, sous-catégorisation, traits sémantiques intrinsèques...).

Selon le niveau de langue considéré, la traduction acquiert ce que nous appelons une certaine « valeur de transfert », en complément à la paraphrase et aux contraintes de sous-catégorisation dont nous venons de parler.

- Le principe du miroir

en complément à ces cinq principes qui sont immédiatement reconnaissables, nous en proposons un nouveau que nous appelons le « principe du miroir ». Le principe du miroir ramène les informations ignorées d'une des parties d'un dictionnaire bilingue puis les fusionne avec les informations de l'autre partie. En

procédant de cette façon, nous obtenons un objet lexicographique « parfaitement » symétrique entre ses deux parties. Dans la plus grande partie du document présent, c'est la transgression systématique - et involontaire - de ce principe par la lexicographie qui nous concernera.

Le principe du miroir est tout simplement le processus de fabrication des LI. Dans tous les cas de polysémie verbale, il s'agit d'un miroir déformant en raison de la dissymétrie observée.

- Autre(s) principe(s) à suivre

Au-delà des principes énoncés ci-dessus, il ne faut pas oublier le respect de l'existant avant tout : il est important pour nous de ne pas dénaturer ni augmenter ni altérer le matériau lexicographique. En cas d'erreur manifeste, nous nous autorisons cependant à rejeter des traductions mais il s'agit de cas très exceptionnels et concernent essentiellement des coquilles orthographiques ou typographiques.

Parmi les sources d'inspiration bibliographiques et scientifiques, plusieurs méritent d'être citées dès à présent. Dans une perspective bilingue, le sujet s'inspire en partie des travaux de Jacques François sur *l'analyse comparative des champs sémantiques et lexicographie bilingue* (in Changement, causation, action, 1989) qui expose le même type de recherche sur quelques verbes français et allemands dont il confronte les structures sémantiques et aspectuelles sous-jacentes. Quelques pistes de réflexion sont empruntées à Igor Mel'çuk, mais avec prudence tant ses modèles, particulièrement détaillés, paraissent inapplicables à des champs entiers à forte polysémie. Pour les questions d'organisation du sens et des contextes thématiques, référence est faite, entre autres, à Georges Kleiber (1990, 1999), Bernard Victorri et Catherine Fuchs (1996), Gilles Fauconnier (1984), Anna Wierzbicka (1996), Eveline Martin ou encore Hüllen & Schulze (1988). Et pour appuyer le travail sur la polysémie verbale, nous utilisons en particulier les travaux de Kleiber (1987) et Maurice Gross (1975) pour le français, et quelques aspects de la classification verbale de Levin (1993) pour l'anglais.

On trouvera de nombreuses autres pistes de réflexion dans les travaux récents autour de l'univers des dictionnaires électroniques : Ceux de l'équipe de Jean Véronis sur le recensement des collocations à partir d'un corpus journalistique, ceux de l'équipe de Gaston Gross sur les classes d'objets, ceux de Jacques François sur la classification verbale à partir de critères aspectuels et participatifs. L'étude de ces travaux récents est toutefois limitée par le caractère figé des différents outils proposés. Nous voulons dire par là que les modèles construits sont peu évolutifs, ou trop opaques selon le cas. Leur exploitation dépasse rarement le cadre de recherche dans lequel ils ont été conçus. Dans l'ensemble, il y a rarement possibilité pour l'utilisateur d'exploiter les dictionnaires ou lexiques à des fins personnelles d'implémentation. En revanche, cet écueil est compensé par la relative richesse des exemples proposés ainsi

que par une certaine « complétude », étant entendu que celle-ci est uniquement le fruit d'un dépouillement de corpus restreint (au sens thématique).

En ce qui concerne nos propres données, l'étude préalable doit permettre de nous diriger vers des tentatives de modélisation de la polysémie verbale. Cette question a été évoquée entre autres par James Pustejovsky, mais ses modèles conduisent selon nous beaucoup plus au traitement de la polymorphie syntaxique qu'à celui de la polysémie proprement dite (cf. chapitre I sur l'auteur et ses travaux). Nous appuyant sur diverses perspectives du domaine (Bouillon & Clas 1993, Hutchins 1986, Dancette 1995), un de nos objectifs majeurs est d'approcher un modèle susceptible d'intégrer aussi bien la polymorphie² que les traits sémantiques intrinsèques ou exprimés ainsi que tous les traits pertinents pour la préparation à la traduction d'une unité verbale. Pour essayer d'atteindre ces représentations multiples, nous puisons largement dans les travaux de Dominique Dutoit sur les « réseaux sémantiques multidimensionnels universalistes aprioriques » (Dutoit, 2000) ainsi que dans l'algorithme de construction des graphes de proximité sémantique établi par Sabine Ploux (1997) et Hyungsuk Ji (Ploux & Ji, 2003) en français et en anglais, sans oublier les travaux du CRISCO portant également sur le dictionnaire électronique des synonymes (Manguin & François 2004 en particulier).

Dans la partie qui traite des analyses sur FRANTEXT, nous nous appuyerons spécialement sur les visions de la linguistique de corpus proposées par Kleiber (1997), Ooi (1998) ou encore Sinclair (1996, cité par Ooi).

Dans cet ordre d'idées, le travail de recherche que nous proposons comporte trois volets qui seront reliés par l'exploitation de traits et de principes communs :

En premier lieu, la prise en compte d'un vaste ensemble d'entrées lexicales et sous-entrées lexicales conduit à construire une nouvelle façon d'aborder le lexique multilingue en y intégrant sous forme codée toutes les informations pertinentes pour la détection et la discrimination des sens particuliers d'une même unité verbale. Ce codage devra conduire au regroupement sous forme de champs sémantiques des unités verbales apparentées et conduira à réfléchir sur les moyens à mettre en œuvre pour représenter conceptuellement les relations lexicales et sémantiques au sein de ces champs dans une perspective de traitement multilingue et non plus seulement bilingue.

Le deuxième volet s'ouvre sur une perspective globaliste de la langue, c'est-à-dire qui entend décrire la langue dans sa réalité au-delà de la forme figée que les dictionnaires lui imposent, et inversement. Nos deux travaux de recherche effectués sur le verbe *traverser* d'une part et sur les verbes de parole d'autre part sont censés illustrer la complexité du langage réellement produit ainsi que la nécessité d'ouvrir les portes du lexique aux occurrences scripturales originales. Si nous y parvenons, nous

² Nous voyons la polymorphie comme une faculté pour une même unité de sens de prendre des formes différentes.

tenterons de réaliser la fusion de toutes les données, lexicographiques et littéraires, au sein d'un espace unique et automatisable.

À l'issue de ce deuxième volet, nous proposons d'en ouvrir un troisième, si tout va bien : il sera possible d'exploiter les réseaux sémantiques et conceptuels (cf. Sowa 1983 ; Dutoit 2000 ; Ploux 1997 ; Gaume 2003 ; Manguin & François 2004, entre autres) réalisés dans l'optique d'une mise en relation entre les différentes langues visées de l'organisation mentale du sens des unités verbales. L'objectif ultime de ce traitement des données est d'essayer d'apporter une solution nouvelle aux systèmes de traduction automatique en prenant en compte à la fois les exigences dues à la richesse de la langue et les impératifs liés à l'implémentation de ces données. Cet apport peut être double, par les réseaux sémantiques et conceptuels d'une part et par l'émergence de nouvelles règles bigrammes ou trigrammes³ (voire davantage) d'autre part. L'intensité du problème à traiter est variable en fonction des langues étudiées : ainsi, la relative symétrie entre le français et l'italien semble moins problématique en coût d'informations que le passage du français à l'anglais. L'intérêt de l'étude réside ainsi autant dans la correspondance lexicale entre les langues que dans les méthodes permettant de réduire l'écart cognitif qui les sépare.

Dans le cas le plus favorable, nous souhaiterions élaborer des « cases conceptuelles » où un même sens sera étiqueté dans toutes les langues étudiées au même endroit dans un seul réseau sémantique. Cela permettrait en particulier de supprimer certaines étapes intermédiaires de transfert habituellement indispensables en traduction automatique. Par extension, l'objectif ultime de ce travail est d'aboutir à une représentation de toutes les langues étudiées au sein d'un graphe unique de correspondance lexicale, sans intermédiaire entre ces langues.

³ Nous n'avons pas effectué l'écriture de ces règles, pas plus que nous n'avons essayé de les intégrer à notre algorithme final. Toutefois, certaines de nos analyses, notamment celle du verbe *traverser* montrent qu'une large prise en compte du co-texte peut être très utile pour la désambiguïsation de certains énoncés. C'est une des pistes que nous devons explorer s'il y a lieu pour la suite de ce projet.

2. Organisation du document

2.1 Des verbes différents

Cette thèse présente l'étude de plusieurs verbes selon ce que nous cherchons à faire émerger. La justification d'une telle hétérogénéité est double. Tout d'abord, chaque verbe ou ensemble de verbes analysé(s) met en exergue un ou plusieurs phénomène(s) bien précis qui ne sont pas nécessairement partagés par les autres. En second lieu, les analyses que nous allons détailler ont toutes un lien avec des recherches ciblées (collaborations, congrès, travaux personnels) dans lesquelles nous avons été partie prenante :

- Le verbe *aborder* a été le premier étudié, dans deux travaux de maîtrise, le premier sur la métaphore des verbes de mouvement, le second pour les travaux fondateurs des Lexiques Miroirs ou Inversés, malgré l'impossibilité physique de soutenir ce dernier.
- *Abandonner* a été étudié pour son étonnant champ lexical en français et dans les autres langues en regard de sa polysémie peu marquée. Son comportement en traduction a été détaillé pour le congrès international « Représentation du Sens » à Montréal en juin 2003.
- *Traverser* a fait l'objet d'une longue et fructueuse collaboration avec Laurent Mascherin (doctorant à l'ATILF travaillant sur l'aspect et la morphologie du verbe), laquelle a conduit à une présentation au colloque international « Space and Time in language and literature » à Rijeka, Croatie, en avril 2005. Nous nous sommes intéressés aux contenus aspectuels des prédications contenant *traverser* aussi bien qu'à l'influence du co-texte sur ledit aspect (références tirées de FRANTEXT).
- *Se mettre à* est la seule périphrase verbale dont nous avons poussé l'étude sur le même modèle que les verbes monolexicaux. Initialement motivé par une participation au congrès international sur « les périphrases verbales » à Caen (2004), notre travail a permis de définir un cadre applicatif et théorique pour faciliter l'intégration des périphrases et des locutions verbales à nos graphes sémantiques.
- *Dire, parler* et les autres verbes de parole sont le matériau central des travaux menés par le Projet de Recherche ILF « transitivité ». Notre implication dans ce PDR a été marquée par une redéfinition de la notion de verbe de parole, une réécriture de la nomenclature de travail

et l'étude systématique du comportement en contexte langagier de tous les verbes de parole ambivalents du français (à partir de FRANTEXT une fois encore), c'est-à-dire permettant aussi bien des formes transitives qu'intransitives.

Même si leurs études peuvent paraître disjointes, chacun de ces verbes a permis la mise en relief d'un ou de plusieurs phénomène(s) spécifique(s). D'autre part, la méthodologie dégagée pour chacun de ces prédicats verbaux est commune à l'ensemble.

En définitive, la diversité revendiquée de nos matériaux d'analyse nous a amené à penser que seule une analyse verbe par verbe, ou au mieux champ lexical par champ lexical, est valide dans l'optique de montrer précisément le fonctionnement de chacun.

2.2 Plan général

2.2.1 Première partie : problématique et choix théoriques.

En première partie, nous présentons la problématique, l'état de l'art et les choix théoriques associés à notre sujet.

Le premier thème abordé est la polysémie, en tant que thème linguistique très discuté. Qu'est-ce que la polysémie ? À quoi sert la notion ? Nous débiterons avec une distinction entre polysémie et homonymie avec l'idée de continuum indéterminable (cf. Victorri) et tenterons de nous orienter vers une fusion de la distinction dans le rapport homographique intra-catégoriel⁴. Diverses questions seront posées à l'appui de notre démonstration : La polysémie comme outil d'analyse linguistique est-elle sans réalité dans la langue ? La polysémie existe-t-elle en tant que telle en dehors d'une vision métalinguistique ? Loin d'être des remises en cause des notions bien établies dans le domaine de la sémantique, nos réponses à ces questions correspondent aux objectifs que nous nous sommes fixé, à savoir la représentation des relations sémantiques entre les mots de différentes langues, relations débarrassées de toute contrainte éditoriale.

Nous aborderons également, plusieurs formes d'organisation des verbes selon des critères syntaxiques et/ou sémantiques (de Maurice Gross pour le français à Jacques François (comparaison entre le français et l'allemand et classement des verbes du français), Levin pour les verbes anglais, les travaux du PDR ILF sur la transitivité des verbes de parole, etc.).

Pour compléter et étendre notre vision de la polysémie, nous présenterons la synonymie comme lien sémantique majeur dans chaque langue et entre les langues.

⁴ Une unité lexicale peut appartenir à des catégories grammaticales différentes, on parlera alors de rapport inter-catégoriel. Dans le cas d'une homonymie portant sur une unité de même catégorie, avec une graphie identique, nous parlons de rapport homographique intra-catégoriel.

À partir de nos considérations sur la polysémie et la synonymie, nous exploiterons le dictionnaire bilingue comme un facteur de correspondance lexico-conceptuelle « complète » entre deux langues, en partant du principe qu'il n'y a pas de vide sémantique intralangue ni vide sémantique interlangue. De cela découle l'idée d'unifier les langues de l'étude dans un graphe unique (cf. Dutoit pour un début de méthodologie).

Bien entendu, nous proposons une définition la plus précise possible du verbe, en nous appuyant sur les considérations de Touratier (1996) et des auteurs qu'il cite pour étayer son point de vue sur la notion. À partir de la définition choisie, nous décrirons au passage quelques travaux typologiques liés aux verbes (M. Gross en syntaxe, G. Gross pour le traitement des unités polylexicales, Levin pour une classification sémantique, Vendler dans une perspective aspectuelle, etc.).

En conclusion à ce premier chapitre, nous présenterons la genèse et le mode d'utilisation des « Lexiques Inversés », intégrant les locutions verbales comme unités lexico-sémantiques homogènes car un « verbe » n'est pas un « mot » au sens strict. Nous utilisons globalement le terme lexie en tant qu'unité lexicale homogène sur le plan sémantique.

2.2.2 Outils et analyse lexicographiques : Lexiques Inversés et Construction de Champs Lexicaux et Sémantiques

L'observation critique des dictionnaires bilingues fait apparaître :

- une forte dissymétrie des deux parties de chaque dictionnaire ;
- l'importance du paradigme de traductions pour chaque lexème courant, ce qui semble induire une forte polysémie sous-jacente (hypothèse partiellement corroborée par la consultation de dictionnaires monolingues) ;
- l'absence d'organisation claire des critères de sélection du mot-vedette approprié (même avec un choix de typographies différentes pour les différencier).

Pour pallier le premier problème, nous avons construit des lexiques multilingues à partir de tous les verbes du français (expressions composées comprises) en utilisant la méthode de la fusion des données, d'une façon similaire à Sabine Ploux pour les dictionnaires de synonymes (PLOUX, 1997) : en plus des traductions du verbe recueillies dans la partie français-langue cible des dictionnaires bilingues, nous avons également récupéré les verbes en langue cible pour lesquels chaque verbe est une traduction proposée. C'est le principe des « Lexiques Inversés » (LI), selon l'expression que nous avons choisie. La création des LI permet dans un premier temps

de rétablir la cohésion de surface entre les deux parties du dictionnaire pris comme référence. C'est également un pas supplémentaire vers l'exhaustivité en traduction potentielle (i.e. dans le choix proposé à l'utilisateur).

Afin d'assister ce choix et pour réduire les effets de la polysémie absolue, les traductions recensées et fusionnées sont organisées dans une base de données en fonction de critères discriminatoires pertinents (des constructions syntaxiques jusqu'aux variations contextuelles) mais initialement exclusivement issues du dictionnaire. Nous obtenons ainsi des micro-paradigmes de traduction pour chaque « sens » de la lexie ; le regroupement des données donnant lieu à un dégroupement nouveau des acceptions.

Pour répondre à la troisième critique, les critères discriminatoires sont mis en exergue dans la base. Ainsi, ce n'est plus le mot-vedette qui est prééminent, mais ce qui en motive le choix. Pour effectuer ce dernier, il peut s'avérer nécessaire de proposer une paraphrase de l'emploi du mot-source à traduire (ou de sa construction). D'où la nécessité d'étudier le mot-source conjointement à ses synonymes contextuels en langue source.

Le travail sur les LI induit quelques questions non subsidiaires qu'il s'agira pour nous de traiter en profondeur :

- la question de la dissymétrie ou de l'asymétrie en lexicographie bilingue : introduction de ce problème récurrent avec la présentation détaillée de la dissymétrie lexicographique des traductions du verbe *aborder* entre le français et l'italien ;
- la question de la distributionnalité des entrées lexicales en fonction des traductions potentielles : avec l'analyse de la distribution interlangue du verbe *abandonner*. Ses équivalents en anglais, son paradigme de synonymes et quasi-synonymes en français et leurs traductions. Problèmes et solutions ;
- le traitement des périphrases et des locutions verbales : discussion et travail autour de la périphrase aspectuelle *se mettre à*, méthodologie pour l'intégrer à nos lexiques en tant qu'unité verbale à part entière. Discussion de son statut à la lumière de la définition du verbe de Touratier et consorts.

Au centre de ce chapitre, nous proposons l'exploitation des « lexiques inversés » pour aboutir aux « champs sémantiques multilingues unifiés » qui seront largement décrits au chapitre IV. Pour une première expérimentation de ces derniers, nous avons cherché un lexème « efficace » et facile d'utilisation (nombreuses traductions, petit nombre d'acceptions). Le verbe *abandonner* a été retenu pour cette

étude en raison de sa faible polysémie et de son noyau de sens « stable ». Par récupération des traductions directes, inversées et indirectes, le nombre total de traductions d'*abandonner* est considérable et justifie les moyens mis en œuvre pour son analyse et sa représentation.

Enfin, nous concluerons ce chapitre par un large aperçu de nos travaux sur les verbes de parole en contexte langagier et sur l'influence qu'ils ont eu sur notre méthodologie d'ensemble. En effet, le cadre par trop restrictif des dictionnaires n'offre pas une vision globale satisfaisante des faits de langue observés dans les corpus. De ce fait et pour élargir le champ de notre démonstration linguistique, nous avons puisé dans les travaux effectués en collaboration avec les membres du PDR ILF sur « la transitivité » la matière nécessaire pour faire une analyse linguistique sur un champ lexical entier. Les verbes de parole ont servi de matière aussi bien au PDR qu'à nos recherches personnelles ; leur nomenclature est l'une des plus riches qui soit en français puisqu'on recense près de neuf-cent verbes ou locutions verbales pour lesquelles au moins une acception respecte explicitement les critères que nous définissons précisément dans ce sous-chapitre pour autoriser leur intégration à la nomenclature des verbes de parole.

L'étude dans son ensemble pose la question du champ sémantique pré-construit et de ses caractéristiques afférentes, conduisant souvent à une forte remise en cause des données langagières issues des dictionnaires et de la grammaire normative pour une classe de verbes reliés par un sème commun. Le paradigme des verbes de parole que nous avons réécrit, leur analyse sur corpus et la mise en correspondance interlangue illustrent bien qu'un dictionnaire ne peut pas donner de vision exhaustive de la langue, et ce quelle que soit sa dévolution.

2.2.3 *Étude autour du verbe traverser. Compléments et complémentarité des données contextuelles et lexicographiques.*

Le travail effectué à partir des seules sources lexicographiques n'est pas entièrement satisfaisant car il ne rend pas compte du potentiel réel de création sémantique pour chaque unité polysémique. Et par ailleurs le dictionnaire même fusionné ne fournit toujours pas de manière explicite les informations permettant de distinguer les acceptions en fonction de leurs critères respectifs. C'est pourquoi nous avons effectué des études purement linguistiques sur un verbe seul (*traverser*) afin de confronter les résultats de l'analyse des données issues de corpus aux données strictement lexicographiques.

Ce chapitre pose une nouvelle fois la question de la compositionnalité sémantique pour les données lexicales et donne lieu à une analyse approfondie des occurrences du verbe *traverser* en contexte, dans des textes littéraires francophones. *Traverser* est tout d'abord considéré à partir du matériau lexicographique afin de faire ressortir l'asymétrie dont il est l'objet dans un seul ouvrage bilingue. Ensuite, à partir des occurrences attestées et des conclusions que nous en tirons, nous essayons

d'établir des correspondances aspectuelles et lexico-conceptuelles entre *traverser* et les verbes associés des autres langues que nous confrontons. Pour cela, nous prenons en compte les éléments co-textuels qui participent au contenu sémantico-aspectuel du verbe, en particulier la configuration spatio-temporelle des arguments de *traverser* ainsi que les éléments extra-prédicatifs dont la portée va jusqu'au contenu prédicatif.

Cette étude d'abord très générale conduit à considérer le verbe *traverser* comme un verbe inclassable sur le plan aspectuel si tous ces éléments ne sont pas caractérisés. Il en résulte pour nous l'écriture de modèles quadridimensionnels qui rendent compte dudit aspect, et nous terminons le chapitre sur des considérations contrastives en fonction de tous les critères spatiaux et temporels que nous aurons retenus.

Le cas des sens métaphoriques du verbe sera également traité, mais surtout pour montrer la difficulté qu'il y a à essayer de les classer et de les représenter dans des modèles concrets.

Suite à ces deux analyses distinctes et complémentaires sur *traverser* et les verbes de parole, deux hypothèses seront à vérifier :

Première HYPOTHÈSE : Ces analyses produisent des relations lexicales équivalentes et/ou complémentaires aux dictionnaires.

Seconde HYPOTHÈSE : Ce qu'on retrouve comme constantes syntaxico-sémantiques dans une langue se retrouve dans les autres langues de manière équivalente.

2.2.4 Cinquième Partie : Modélisation et algorithmique des CSMU

Notre objectif final est la constitution de graphes lexicaux multilingues illustrant la non biunivocité des correspondances lexicales entre les langues, et permettant de la dépasser. En effet, nous avons déjà insisté sur le fait qu'il est rare qu'à un terme en langue source corresponde une seule traduction en langue cible. Par surcroît, le décalage interlangue grandit de manière exponentielle à chaque ajout d'une langue dans l'étude.

Le chapitre IV présente très en détail le traitement des critères qui serviront de catégories discriminantes au sein des CSMU. Les Champs Sémantiques Multilingues Unifiés (CSMU) sont des graphes de relations lexicales à lecture multiple, les CSMU pouvant être visualisés dans leur totalité ou bien partiellement par l'appel d'une lexie qui y est présente. Il contient également l'essentiel de notre travail préparatoire à la programmation d'un graphe de relations lexicales qui soit conforme à la méthodologie que nous avons développée auparavant.

Selon nos souhaits, la forme de ces champs permettrait de construire des dictionnaires bilingues ou multilingues en utilisant n'importe quelle langue présente dans le graphe comme langue source, voire en abolissant totalement le statut de langue

source. De même, toute unité lexicale du graphe pourra être sélectionnée comme lexie centrale de référence pour la visualisation du graphe.

L'ambition de la mise en algorithmes des CSMU peut être perçue comme les moyens à mettre en œuvre pour opérer la transformation d'un ou plusieurs dictionnaire(s) (notamment bilingues) en graphe multilingue unique, lui-même transformable en dictionnaire(s) à son tour. L'ajout d'une nouvelle langue et de nouvelles données ne doit pas en principe démultiplier le graphe. L'un des objectifs finaux demeure de restituer un dictionnaire de langues coopératif pour l'aide à la traduction à partir des CSMU.

Pour décrire précisément l'algorithme d'implantation du graphe, nous aurons recours aux notions et outils informatiques d'objets structurés, d'ensembles et de fonctions. Les premiers contiennent l'ensemble des traits distinctifs pour une lexie donnée (globalement ou en contexte), les seconds sont utilisés en guise de listes de traits ou de valeurs associées à un attribut d'objet donné et les troisièmes permettront l'écriture de sous-programmes nécessaires à la bonne implantation des liens de synonymie que nous définissons également dans ce chapitre.

La partie la plus approfondie de ce chapitre présente l'intégralité de l'algorithme final que nous commenterons très largement en langage naturel. En guise de conclusion intermédiaire, nous tenterons de mettre en avant les atouts d'un tel outil sans en négliger les faiblesses.

2.2.5 Partie conclusive : Bilan et perspectives

En guise de conclusion ouverte à nos travaux, cette dernière partie reprend certains points de la problématique d'ensemble, dont les suivants :

- la volonté de briser le décalage habituel qui fait représenter les systèmes lexicaux de chaque langue indépendamment les uns des autres ;
- la visibilité des « champs sémantiques multilingues unifiés ». En quoi ces CSMU peuvent apporter quelque chose dans le domaine du TAL ? Sont-ils seulement utiles d'un point de vue théorique ? ;
- l'exploitation des outils informatiques pour la représentation et la manipulation de ces champs : opportunité de coupler au langage XML, à des logiciels de gestion des bases de données ou à des graphes visuels (que nous devons justifier hors d'un point de vue purement « artistique ») ;
- « une contribution au dictionnaire du 21^{ème} siècle ». L'idée est que si un dictionnaire peut générer un graphe, alors le même graphe peut

générer un dictionnaire. Cette réalité est la meilleure justification de l'utilité des Lexiques Miroirs, et s'avère la mise en relief souhaitable de leur malléabilité ;

- une réflexion à propos de l'intérêt et nécessité d'avoir une double-vision du lexique : en aval, il y a la lexicographie, et en amont la reconstitution du lexique par contenu sémantique et étude de corpus. Les deux aspects ne sont pas incompatibles. Tout notre travail peut s'apparenter à un plaidoyer en filigrane pour une linguistique mixte qui postule que les lexies ont une réalité en contexte et une réalité hors contexte et qu'aucune n'est plus valide que l'autre pour des raisons que nous justifierons tout au long de ce document ;
- nous effectuerons également un petit détour par les prédicats non verbaux : tout ce qui n'est pas verbe n'est pas forcément différent du verbe d'un point de vue sémantique, comme nous l'aurons déjà évoqué dans la définition du verbe dans le premier chapitre.

En guise de conclusion finale à la conclusion générale, nous tenterons une dernière fois de défendre l'intérêt réel ou supposé du cheminement de nos travaux jusqu'à l'écriture d'un algorithme dont l'apport en tant qu'outil de recherche doit être précisé. Le mot de la fin sera pour le « plaidoyer pour une linguistique mixte » que nous annonçons plus haut car la linguistique au sens large ne saurait ignorer aucun pan de la réalité de la langue.

3. Annexes

En raison de leur caractère très volumineux, nous avons décidé de ne pas faire figurer nos travaux annexes dans le présent document. L'intégralité des travaux de saisie et de mise en forme des données annoncés ci-dessous aura été portée à la connaissance des membres du jury et se trouve sur un CD-Rom joint à la thèse.

Lexiques Inversés

Lexique Inversé Français-Espagnol en l'état.

Grille d'analyse de quelques verbes de parole en contexte

Cette grille comporte les analyses des occurrences extraites de FRANTEXT d'un nombre important de verbes de parole selon le protocole que nous décrivons succinctement dans le présent document.

Base de données de *abandonner*

Échantillon des CSMU autour du lexème-noyau *abandonner*. Saisie intégrale de synonymes et des traductions potentielles de ce verbe pour expérimentation par le programme de création de graphes. Le graphe qui en découle constitue l'annexe suivante.

Graphe du verbe *abandonner*

Faute de disposer du logiciel adéquat, le graphe en question ne se présente pas sous la forme des graphes que l'on a l'habitude de visualiser, avec des étiquettes et des flèches, mais simplement sous la forme d'un ensemble de liens entre lexies générés par un programme de mise en relation des lexies. Malgré les apparences, cela revient exactement au même.

La base de données test que nous avons saisie contient 4500 lignes, c'est-à-dire 4500 liens de traduction entre lexies, centrées sur le verbe *abandonner*, ses synonymes français et ses traductions en anglais dans le Robert & Collins. Si nous observons juste la première ligne de cette base, nous voyons qu'elle contient trois unités lexicales, le programme va donc générer pour cette ligne trois structures de lexies et trois liens de synonymie. Le graphe créé automatiquement à partir de cette base de données est ici rendu sous une forme lisible par tous. Chaque triplet correspond à la création d'un lien de synonymie reliant deux lexies. Par exemple, la première ligne signifie que la lexie de *abandon* qui porte l'indice 1 possède un lien de synonymie de type traduction directe (ou ST) avec la lexie d'*abandonner* qui possède l'indice 2. En remontant au mot, on peut en déduire qu'*abandon* et *abandonner* sont synonymes de type ST via ces lexies. Pour les mêmes lexies 1 et 2, on observe que *forsake*, ou plus exactement la lexie de *forsake* d'indice 3, est codé comme synonyme SPS d'*abandon* et synonyme SI d'*abandonner* comme nous l'avons annoncé. La base de données de 4500 lignes a permis de créer un graphe contenant plus de 12700 lexies et à peu près autant de liens de synonymie. Ce graphe simple est nous l'espérons le point de départ d'un travail de recherche très conséquent dans le domaine des dictionnaires et lexiques électroniques.

CHAPITRE I
Problématique et fondements théoriques

1. Un sujet vaste

Bien qu'étant à l'origine un travail orienté vers la description et l'exploitation de données lexicographiques, la thèse que nous défendons a largement dépassé ces bornes pour aller explorer de nombreux domaines de la linguistique.

En premier lieu, l'usage systématique de dictionnaires comme instruments linguistiques de référence a conduit à une réflexion sur le comportement des unités lexicales décrites dès qu'elles apparaissent en contexte. À ce titre, notre recherche s'interroge longuement sur l'enrichissement des données par le corpus, et par-dessus tout sur le thème connexe : quelle linguistique mener en sémantique lexicale ?

Derrière cette vaste question, nous entrebâillons la porte vers un débat central des sciences du langage contemporaines : quelle linguistique « sans » corpus et quelle linguistique « de » corpus ? Nous montrerons le plus largement possible qu'il n'y a pas incompatibilité entre la linguistique de corpus et les approches traditionnelles, majoritaires avant ce mouvement d'envergure. D'autre part, nous entendons également démontrer que le ou les corpus ne peuvent pas réellement nourrir des théories et des applications qui vont au-delà de ce qu'ils représentent pour eux-mêmes, ce qui pour notre chapelle conduit à rejeter la « *corpus-driven linguistics* » de Sinclair en tant que panacée (voir au chapitre II, sous-chapitre 9).

Le travail de lexicographie ne s'effectue pas au hasard, en fonction du « caprice » momentané d'un rédacteur face à un lexème particulier ; il s'exerce au contraire dans le strict respect d'une méthodologie fondée sur des principes tangibles. Toutefois, l'exemple du TLF nous montre que si tout rédacteur isolé a des principes généralement stables, ses collaborateurs ont tendance à développer leur propre méthodologie, au risque de mettre en péril la cohésion de l'ensemble. A partir de nos lectures et de nos réflexions sur le sujet, nous avons décidé de proposer notre propre méthodologie de construction de lexiques. En premier lieu et avant la description des « lexiques inversés » sus-mentionnée, nous rappelons que l'étude est censée respecter :

- le principe d'exhaustivité ;
- le principe de limpidité ;
- le principe de réécriture ;
- les principes de sous-catégorisation syntaxique et sémantique ;
- le principe de valeur de transfert ;
- le « principe du miroir ».

Ces principes ne seront guère redéfinis ni réévalués mais serviront de toile de fond permanente à l'exposition de nos travaux de recherche.

2. Sur les notions de base

2.1 Qu'est-ce que la polysémie ?

2.1.1 Définition générale

La première définition connue de la polysémie a été donnée par Bréal (1897 :154/155) :

« Le sens nouveau, quel qu'il soit, ne met pas fin à l'ancien. Ils existent tous les deux l'un à côté de l'autre. Le même terme peut s'employer tour à tour au sens propre ou au sens métaphorique, au sens restreint ou au sens étendu, au sens abstrait ou au sens concret... A mesure qu'une signification nouvelle est donnée au mot, il a l'air de se multiplier et de produire des exemplaires nouveaux, semblables de forme, mais différents de valeur. Nous appellerons ce phénomène de multiplication la *polysémie*. »

Bréal (1897 :154/155)

Nous retenons de cette notion alors nouvelle le fait que Bréal l'associe à un « phénomène de multiplication », induisant l'interprétation de la polysémie comme phénomène de production autant que comme état de la langue. Autrement dit, la polysémie peut être perçue comme « la capacité des mots de *prendre un sens nouveau* », ainsi que le notent Victorri et Fuchs (1996 :11) à propos de la définition du même auteur.

De ce point de vue, il est possible de percevoir la polysémie comme un potentiel de la langue, celui de faire apparaître des significations nouvelles pour des unités qui ne les offraient pas précédemment. De fait, même si ce potentiel n'est pas forcément exploité, tous les mots de la langue seraient donc des unités polysémiques *en puissance*.

Ce dernier point est très important dans la mesure où il sous-tend l'extrême instabilité du lexique. Bien que nous n'en ayons pas conscience, ce dernier évolue et se modifie constamment, rendant superflue toute tentative d'en figer le contenu. Ce potentiel aura une influence sur les modèles que nous allons créer dans la mesure où nous devons prévoir l'opportunité de les faire évoluer à tout moment.

La linguistique et le linguiste tentent de distinguer la polysémie (capacité d'un mot à prendre plusieurs sens) de l'homonymie (fait pour une unité lexicale de constituer plusieurs mots) en se fondant sur des critères diachroniques : deux unités homographes, homophones et de même catégorie grammaticale sont polysémiques si elles partagent le même étymon, et homonymiques dans le cas contraire. Selon

Victorri & Fuchs (op. cit.), accepter cette définition revient à ignorer que l'évolution de la langue a favorisé l'éloignement des différentes acceptions de nombreuses unités lexicales à étymon unique.

« Un exemple bien connu est celui de *grève*, qui, à partir du sens de rivage, a pris progressivement l'acception d'arrêt de travail par l'intermédiaire du nom de *la Place de Grève*, lieu où l'on transbordait les marchandises circulant sur la Seine à Paris, et qui était devenu le lieu de rassemblement des travailleurs : *être en grève* a signifié au 18^e siècle chercher du travail, par extension du sens 'attendre l'embauche en place de Grève' (...). »

Victorri & Fuchs (1996)

Bien qu'édifiant, cet exemple ne satisfait pas totalement notre vision de la polysémie pour une raison très simple : L'apparition de l'expression *être en grève* ne résulte pas d'un processus de polysémie tel que Bréal le décrit, à savoir qu'il ne s'agit pas ici du potentiel du mot *grève* à désigner autre chose que le « bord de mer » par analogie à celui-ci. Au contraire, l'expression en question a été construite directement à partir d'un nom de lieu éponyme⁵ de *grève* qui avait déjà lui-même perdu tout lien sémantique évident avec son terme d'origine. De ce point de vue, il n'y a jamais eu de processus polysémique pour la création de l'expression *être en grève*, mais simplement un processus métonymique⁶ de corrélation entre une action et le lieu où elle se produisait habituellement.

De même, nous pensons que l'exemple de *grève* n'est pas comparable avec la polysémie de *bureau*, également cité par les auteurs, laquelle fut également générée par métonymies successives (*bureau* = sous-main → meuble de travail → pièce où l'on travaille → lieu de travail → ensemble des personnes travaillant ensemble (dans un bureau ou pas)) où chaque nouveau sens est directement motivé par le précédent.

Si le critère diachronique est insuffisant ou plus assez pertinent pour discriminer les deux notions, il faut par conséquent essayer de les caractériser par une opposition sémantique en synchronie, ce qui rend d'autant plus difficile la distinction entre polysémie et homonymie.

Pour d'autres auteurs, il y a polysémie lorsqu'il y a « existence de *sens* différents pour un même mot, mais de sens perçus comme reliés d'une manière ou d'une autre. » (Cadiot & Habert, 1997:3)

Or, il est relativement difficile de déterminer une gradation des degrés de proximité de sens entre deux acceptions d'un même homographe, et bien davantage d'en systématiser la méthode pour toutes les unités d'une langue. Si l'on admet la

⁵ C'est-à-dire portant le même nom qu'un élément qui lui est relié par métonymie, en l'occurrence la Place de Grève.

⁶ La métonymie décrit un processus de correspondance lexicale entre deux éléments reliés par une relation de proximité ou de contiguïté concrète ou abstraite de type « personne-lieu où se trouve cette personne », « œuvre-créateur de l'œuvre », etc.

définition précédente, il est manifeste que celle-ci s'oppose à une perception diachronique de la langue puisque le hasard de l'évolution d'une langue peut aussi bien éloigner des acceptions étymologiquement corrélées⁷ qu'en rapprocher d'autres entre deux homographes d'origines distinctes. Ce dernier phénomène est un processus d'influence ou d'imprégnation : les sens éloignés d'un même terme peuvent se rapprocher par l'intermédiaire de leur rattachement à une unité de forme équivalente. Ce phénomène est comparable à la paronymie, laquelle permet également l'apparition de nouveaux sens pour un terme donné à partir d'un sens d'un mot qui lui est proche phonétiquement ou orthographiquement.

Plus que tout autre, le phénomène de la polysémie est sensible à l'instabilité du continuum lexico-sémantique dans la langue. Il en est même le corollaire notionnel.

2.1.2 Des difficultés de proposer une définition stable

La polysémie est instable en tant que telle. Ainsi que le montrent les paragraphes précédents, définir précisément ce qu'est la polysémie n'est pas chose facile.

« La polysémie, comme beaucoup d'autres termes linguistiques, est une notion qui se laisse très facilement appréhender de manière intuitive, mais qui se révèle beaucoup plus rétive à une définition rigoureuse. On peut aisément se convaincre de la réalité et de l'importance de ce phénomène, en donner de nombreux exemples, et en décrire les principaux traits. Mais ces descriptions restent informelles, et dès que l'on cherche à dépasser ce niveau superficiel, pour tenter de caractériser exactement cette notion, d'en déterminer les limites précises, d'en donner des critères définitoires, on se heurte à la nécessité de construire un cadre théorique global, avec son lot de définitions axiomatiques et de postulats toujours discutables, qui nous éloignent de la relation immédiate que nous entretenons avec la langue, et qui peuvent sembler d'une complexité démesurée par rapport à l'objet étudié. »

(Victorri & Fuchs, 1996 :11)

Pour omniprésente qu'elle soit, la réalité de la polysémie se heurte à un écueil de taille dans la perspective d'une systématisation : la langue est en perpétuelle évolution et les unités qui la composent en sont les éléments les plus perméables. Il faut bien sûr des décennies avant qu'une règle grammaticale nouvelle s'impose ou pour qu'une autre se modifie durablement. Toutefois, l'apparition de nouvelles acceptions à l'échelon lexical est constante, généralement imprévisible et

⁷ Nous citerons ici l'exemple traditionnel du verbe *voler*, lequel possède deux acceptions majeures à étymologie commune (le sens de *dérober* étant issu d'un des sens mineurs de *se déplacer dans les airs*, à savoir *poursuivre et chasser en volant* (pour *un faucon*)). Il n'est guère possible de rétablir un quelconque lien sémantique entre ces deux acceptions aujourd'hui.

passablement soudaine. Ainsi, la définition de la polysémie passe inévitablement par la reconnaissance de cette mouvance incessante et par l'acceptation du fait qu'aucune étude systématique au temps présent ne demeurera pérenne à moyen voire à court terme. Nous reviendrons à ce sujet sur quelques travaux lexicographiques (Trésor de la Langue Française, TLF) et lexicologiques (Dictionnaire Explicatif et Combinatoire de Mel'çuk et Polguère, le DEC, Dictionnaire Intégral de Dominique Dutoit, Eurowordnet, etc.) soumis à l'érosion du temps et à la nécessité de la guérir à défaut de la prévenir, nécessité qui n'est d'ailleurs pas toujours suivie d'effet.

Aucune étude linguistique convaincante n'a à ce jour essayé d'aborder ostensiblement la question de la prospective en sémantique lexicale. Dans une perspective plus ludique que scientifique, tout juste pouvons-nous citer nos propres travaux (Haton, 2004) menés en collaboration avec quelques collègues de l'ATILF (P. Bernard, W. Del Mancino, H. Gerner et E. Jacquy) dans le cadre du projet « Polyglotron » initié par la Cité de la Science à Toulouse et pour lequel il nous avait été demandé d'imaginer quelques phrases en français du vingt-septième siècle. Ce fut une tâche certes ludique mais au contenu scientifique fort discutable. D'autres travaux, plus récents et plus « sérieux », ont été lancés à l'ATILF par Mathieu Valette (2004), lequel dans un premier temps a tenté d'explorer par la morphémique les mécanismes de néologie liés aux textes racistes sur Internet. Dans un avenir proche, Mathieu Valette espère expliquer à partir du mécanisme de néologie la manière dont les nouveaux sens vont émerger à partir des sens existants⁸.

L'autre difficulté liée à la définition de la polysémie réside dans les termes qui la composent. Dans sa forme la plus élémentaire, la polysémie est le propre pour un *mot* de pouvoir prendre plusieurs *sens*. Cela impose de savoir et d'explicitier ce qu'est un mot, et également ce qu'est le sens. Or, ces deux notions sont tout particulièrement délicates et font partie des « autres termes linguistiques » qui ne se laissent pas facilement définir, selon Victorri et Fuchs.

Concernant le premier, il s'agit de faire une différence initiale entre le mot en tant qu'artefact que l'on peut directement appréhender par l'observation et le méta-objet linguistique. Cette distinction sert paradoxalement de clin d'œil à la polysémie en opposant le matériau à l'artefact, comme c'est couramment le cas dans la langue.

Il est encore plus difficile d'appréhender le sens de *sens*. En réalité, il n'est pas possible d'en donner une définition claire sans l'associer à d'autres notions ou à des situations particulières où le sens est figé en une substance 'inférieure'. En lexicographie, par exemple, le sens est restreint au cadre définitoire, de même que le mot y est représenté par une entrée lemmatique. Dans un dictionnaire, un sens d'un mot est une définition. Il est évident que cette restriction est rendue nécessaire par l'usage que l'on fait des ouvrages lexicographiques, mais elle ne permet en rien de

⁸ Perspective annoncée et présentée lors d'un atelier de rentrée CNRS-ATILF-Nancy2. Septembre 2005.

résoudre l'énigme du sens car toute définition est circulaire, elle renvoie à d'autres unités de la langue, lesquelles sont à leur tour définies et ainsi de suite. Pour percevoir l'essence du sens, il est nécessaire de retourner à la racine, à sa genèse même. D'un point de vue cognitif, le sens est ce qui précède l'expression langagière et ne peut donc être décrit par le langage puisqu'il procède d'un mécanisme qui est antérieur à celui-ci. Cependant, peut-on caractériser « l'essence du sens » au moyen du langage ? Et si oui, peut-on le faire sans risquer de retomber dans la circularité que nous venons de décrire ? Un choix doit être fait ici : soit nous considérons que le sens linguistique est toujours descriptible par la langue, auquel cas nous le décrirons et en resterons là, soit nous admettons que le sens relève d'une substance d'un autre ordre, égaré dans les dédales mystérieux des connexions neuronales. Et sur ce plan, nous nous déclarerons incompetent.

Pour rester cohérent avec notre démarche, et surtout pour nous autoriser à continuer de parler de sens lexical, nous suggérons de ne voir dans cette notion de sens que le résultat de l'observation des dictionnaires et des données, à savoir qu'un sens d'un mot renvoie à une acception de celui-ci, ou plus prosaïquement à une description globale de cette acception (liens contrastifs à l'appui).

2.1.3 La polysémie a-t-elle plusieurs sens ?

La polysémie a-t-elle plusieurs sens...et vice-versa ? « Les sens uniques mènent à des impasses » disaient Coulon et Kayser (1982). La polysémie dans son ensemble mène plutôt au delta, serions-nous tenté de répondre en écho.

La *polysémie* servant d'abord à désigner la faculté d'un terme à prendre plusieurs sens, nous pouvons nous demander si le terme est lui-même polysémique, et par extension s'il sert à désigner des concepts différents.

A l'origine, la polysémie semble être une faculté car, selon Bréal, « un terme peut prendre plusieurs sens ». C'est également un état (de langue) : « un terme polysémique a plusieurs acceptions ». C'est encore un processus, ou un mode de production, en cela que « le langage autorise un terme à prendre de nouveaux sens qui s'ajoutent aux précédents ». C'est aussi probablement un outil d'analyse métalinguistique puisque « l'on observe qu'un terme peut prendre des acceptions différentes selon le contexte où il apparaît, ce qui conduit à une description dégroupée de ses sens ».

Pour clore cette liste non exhaustive, nous proposons une vision de la polysémie en tant que phénomène systématique (polysémie absolue) dont la définition porte les germes d'une neutralisation de la notion de polysémie en tant que réalité du langage en dehors de la seule description linguistique. De ce point de vue, il s'agit d'un prolongement de la notion de polysémie en tant qu'outil d'analyse métalinguistique.

2.1.3.1 De l'homophonie à la polysémie absolue (le sens entre tout et partie dans le discours)

Après avoir essayé de synthétiser les positions traditionnelles au sujet de la polysémie, nous allons maintenant entreprendre de dessiner les contours de la notion telle que nous la voyons ou plus précisément telle que nous allons la modeler pour qu'elle corresponde à nos besoins. De notre point de vue, sens et référence forment effectivement « un couple à rabibochoer » selon l'expression de Georges KLEIBER (1999:51). En restreignant la description du sens à un simple dégroupement d'acceptions, la linguistique se trompe (ou se tromperait) d'objectif : le but n'est plus de délimiter la langue ou les faits de la langue, comme les grammaires et les dictionnaires s'en chargent, mais plutôt de proposer une étude de la langue telle qu'elle est ; c'est-à-dire en remplaçant les énoncés dans le réel (cf. Siblot, 1990) sans ignorer le fait que :

« il y a une partie du sens qui est donnée ou préconstruite, c'est-à-dire conventionnelle ; »
(Kleiber, 1999).

Avant d'aller explorer les terres encore assez vierges de la polysémie absolue, il nous faut également réconcilier deux approches de l'analyse du sens qui ne nous paraissent pas réellement incompatibles : la sémantique compositionnelle et l'approche holistique (confrontées dans Gosselin, 1996). Selon la première, le sens d'un énoncé est la résultante de la somme de ses parties, i.e. des sens portés par les mots ou locutions qui le composent. Pour la seconde, en revanche, le sens d'un énoncé forme un tout indivisible distinct de la somme de ses parties. En observant les deux points de vue, nous remarquons qu'ils sont au moins d'accord sur un point, essentiel à nos yeux : le tout, l'énoncé, a **un** ou **du** sens. Comme cela arrive souvent dans ce genre d'oppositions, les deux approches nous semblent avoir raison toutes les deux, avec les nuances qui s'imposent.

L'approche compositionnelle considère que le tout est la somme de ses parties, mais ignore ainsi l'inévitable interinfluence des unités discrètes⁹ mises au contact les unes des autres. Nous pouvons à ce sujet référer à la phonétique et aux altérations que les phonèmes se font subir mutuellement lorsqu'ils sont accolés les uns aux autres, rendant ainsi difficile une reconnaissance de la parole qui ne peut pas se contenter de prendre comme points de référence absolue les spectres individuels de chaque son pour analyser et reconnaître le discours dans son intégralité. Le processus n'est bien entendu pas strictement équivalent dans la concaténation lexicale, mais le phénomène d'influence mutuelle demeure : dans de nombreux cas, un énoncé complet ne peut pas être décomposé en unités discrètes à sens autonome ni paraphrasé par des éléments discrets *stricto sensu*. Par exemple, l'énoncé suivant n'est pas interprétable par la somme de ses parties :

1) *Si vous continuez à piétiner mes plates-bandes, vous allez voir de quel bois je me chauffe*

D'un autre côté, imposer au discours la vision holistique revient à nier la persistance d'une autonomie sémantique des unités discrètes qui le composent. Autrement dit, il est difficile de nier que l'énoncé dans son ensemble est constitué d'unités plus petites que l'énoncé lui-même, et que ces unités ont des sens propres, des acceptions particulières dont le choix a été effectué en fonction des besoins du sens global. D'ailleurs, s'il n'y avait pas préexistence des sens autonomes des unités discrètes, il n'y aurait pas de dictionnaires ni de sens lexical.

Pour en conclure avec l'opposition compositionnalité/holisme, nous accepterons une vision plus consensuelle et, nous l'espérons, plus proche de la réalité du langage et de son rapport au sens : dans un premier temps, le sens pré-langagier existe sous la forme d'une entité insaisissable qui nécessite la verbalisation pour être rendu intelligible. En ce sens, il est effectivement holistique tant qu'il échappe à la langue. En second lieu, il lui est associé un ensemble fini et fermé d'unités langagières (nous pouvons penser aux mots et à leur caractère discret en l'occurrence) destiné à rendre au plus près la réalité du sens primitif à l'ensemble des interlocuteurs potentiels. Les énoncés ainsi construits sont constitués de mots dont aucun n'est choisi au hasard. Tous les mots sont retenus en fonction de leurs acceptions propres ou à tout le moins du potentiel qu'ils recouvrent. De ce point de vue, l'énoncé serait purement compositionnel et son analyse peut en apparence être restreint à l'étude de ses constituants.

En définitive, un énoncé est à la fois holistique et compositionnel. Tout dépend donc de la manière dont le linguiste entend régler les problèmes d'analyse et d'application. La solution personnelle à trouver pour chacun est d'abord d'ordre

⁹ Par unité discrète, il faut comprendre une unité qui possède une forme de base commune à toutes ses variantes sémantiques.

théorique. En ce qui nous concerne, notre position est de considérer que le discours est compositionnel dans la perspective de le modéliser via les graphes lexicaux ; d'autre part, nous admettons que tout énoncé est holistique afin de faciliter l'explicitation de notre position sur la polysémie absolue : un sens d'un énoncé est tout autant la résultante de facteurs linguistiques qu'extralinguistiques, ce qui rend originale et unique chaque production de parole.

La vision du langage dépend également du sens que l'on donne au mot *mot*. À partir de notre exemple 1), peut-on dire que *piétiner les plates-bandes de* et *voir de quel bois je me chauffe* sont des mots ou bien sont-ce des locutions polylexicales découpables en mots ? En observant cette phrase-exemple, on peut d'ailleurs légitimement se demander si les deux locutions qui la composent ne sont pas des unités discrètes à part entière.

Comme nous venons de le suggérer, admettre que tout énoncé en contexte est distinct de tous les autres énoncés produits conduit à un élargissement de la notion de polysémie à la « polysémie absolue » (un emploi = une entrée lexicale) pour se rapprocher 'par le haut' de la représentation langagière, et conséquemment du rapport multilingue. Avant de parler de positionnement théorique, l'objectif déclaré est d'essayer de modéliser la langue « telle qu'elle peut être » et non pas « telle qu'elle est » ou « doit être ». Outre que cette vision du potentiel de la langue est défendue par Dominique Dutoit dans ses recherches, nous faisons directement référence à la réflexion de Récanati ici. En effet, l'auteur annonce en 1997 l'apparition d'une nouvelle vision de la polysémie : la faculté qu'ont les contextes de ne jamais être semblables à aucun autre. Partant du principe qu'il n'y a pas deux situations d'énonciation équivalentes, il n'y a donc pas deux énoncés à sens identiques. Ce phénomène se répercute au niveau du mot en vertu d'une vision holistique du sens. Cette notion est reprise par Gosselin en 2001 sous le terme de « polysémie contextuelle », et par un phénomène combinatoire « polysémie contextuelle généralisée » :

Le calcul sémantique rencontre avec la polysémie contextuelle un obstacle décisif. Par "polysémie contextuelle", on désigne le fait qu'un marqueur (morphème lexical, grammatical, ou construction syntaxique) puisse prendre des significations au moins partiellement différentes en fonction des contextes linguistiques dans lesquels il se trouve. Comme ces contextes sont composés de marqueurs qui sont eux-mêmes, le plus souvent, polysémiques, on parlera de "polysémie contextuelle généralisée" (PCG) pour désigner le fait que la signification d'un marqueur puisse varier en fonction non seulement des formes, mais aussi des significations des autres marqueurs qui l'entourent (lesquelles varient aussi de semblable façon). Autrement dit, adopter le point de vue de la PCG, c'est, essayer de calculer l'interaction globale des marqueurs polysémiques dans l'énoncé, et au-delà, dans le texte.

La solution retenue pour rendre compte de cette PCG consiste à admettre que les différentes instructions codées par les marqueurs constitutifs d'un énoncé peuvent entrer en conflit, et que ces conflits sont résolus par des procédures régulières (et donc prédictibles) de déformation des structures sémantiques. Ce sont ces procédures de résolution de conflit que nous allons exposer de façon détaillée. On montrera aussi que ces mêmes procédures permettent de résoudre les conflits entre contraintes linguistiques (liées aux instructions sémantiques) et contraintes pragmatico-référentielles, dans le cadre de la construction globale de la cohérence.

Gosselin, 2001.

Dans une réflexion d'ensemble qui peut être assimilée à une pure conjecture, nous irons encore plus loin que Récanati et Gosselin en reconnaissant tout d'abord la divergence des contextes au sens large, jusqu'à la « PCG » mais en y adjoignant d'autres divergences qui concernent les locuteurs de la langue. En l'occurrence, chaque émetteur a sa propre sensibilité linguistique, sa propre vision du sens des mots, inconsciente ou non. Les récepteurs présentent les mêmes décalages, avec en plus une déformation du sens encodé qui est reçu de manière non prévue par l'émetteur. Autrement dit, le sens qui précède l'énonciation est différent du sens qui suit la réception.

Émetteurs différents, récepteurs différents et déformation du sens contribuent à systématiser la polysémie absolue, au risque de la rendre inutilisable en tant qu'outil d'analyse.

Néanmoins, cette « ambiguïté » induite par la polysémie n'entraîne aucune hésitation ni véritable erreur d'interprétation de la part du récepteur. De notre point de vue, il n'y a pas incompatibilité dans les termes, il s'agit surtout de l'illustration d'un inévitable décalage entre le sens encodé par le locuteur et celui décodé par le récepteur via un énoncé. La polysémie absolue se rapproche en cela de la polysémie logique, laquelle correspond à des sens contextuels apparentés sans être identiques :

Polysémie logique:

(9)a. *Ouvre la porte. Il fait chaud ici.*

b. *Passe la porte et tourne à gauche.*

(...)

• Polysémie logique:

– se produit quand un mot unique donne accès à plusieurs sens différents mais liés.

(Reboul, 1999)

En acceptant ce principe, nous suggérons une généralisation de la réalité de la polysémie absolue : non seulement chaque occurrence d'un terme constitue une entrée lexicale à part entière au sens de Récanati, mais elle est elle-même génératrice de polysémie selon le locuteur qui l'interprète. Ainsi caractérisée, elle va bien au-delà de

la description proposée par Fuchs au sujet de la polysémie en suggérant que le choix de la bonne acception ne garantit pas l'uniformité de l'interprétation d'un récepteur à l'autre. Une application pratique et indirecte de la polysémie absolue peut être perçue dans la production littéraire, en cela que la relative liberté de créativité des auteurs les conduit à employer certains mots dans des sens très inattendus, sinon soumis à interprétation de la part du lecteur. Les écrivains n'éprouvent jamais la nécessité d'explicitier certaines de leurs métaphores, laissant toute liberté à leurs lecteurs de construire un nouveau sens lexical original, ou pas.

Nous raccordons cette notion au phénomène de « sphère d'influence » d'une unité lexicale : dans notre esprit, une relation de traduction n'est pas biunivoque mais s'intègre dans un espace dont chaque point peut être assimilé à une nuance subtile du sens habituel de la relation entre deux unités¹⁰ (cf. Ploux, 1997, Ploux & Ji, 2003). D'un point de vue sociolinguistique, la polysémie absolue marque les décalages inévitables dans les compétences linguistiques des locuteurs d'une langue ainsi que dans leur sensibilité tout court.

2.1.4 Des polysémies « parallèles »

2.1.4.1 La polysémie verticale

La polysémie verticale a été caractérisée par GEVAUDAN (1997). Selon lui, elle désigne la possibilité pour un terme de s'englober lui-même, d'être son propre hyperonyme. L'exemple mis en avant est celui de *homme* qui sert aussi bien à désigner l'espèce humaine dans son ensemble (concept englobant) y compris les femmes pour des raisons linguistiques, historiques et culturelles que l'individu de sexe masculin qui fait partie de l'espèce (concept englobé).

¹⁰ Outre chez Ploux, on retrouve la définition du potentiel sémantique d'un mot « comme un espace conceptuel assimilable à un nuage de points » (in Manguin & François 2004) chez Croft et Cruse (2004) : « We can portray the total meaning potential of a word as a region in conceptual space, and each individual interpretation as a point therein. Understood in this way, the meaning potential of a word is typically not a uniform continuum : the interpretations tend to cluster in groups showing different degrees of salience and cohesiveness, and between the groups there are relatively sparsely inhabited regions » (Croft & Cruse:109-110). Cette vision est parfaitement conforme à la notre en matière de conceptualisation du sens lexical au sein d'un espace clos.

En soi, ce phénomène n'est pas très éloigné de la polysémie simple, laquelle consiste en la cohabitation de différentes acceptions apparentées pour un terme unique. Nous noterons d'ailleurs que les deux acceptions ne peuvent pas se rencontrer dans un même énoncé, en l'absence de qualification particulière sur au moins l'un des deux, et encore cela ne fait-il pas vraiment partie de la langue telle qu'elle est réellement produite :

- 2) *L'homme est un loup pour l'homme.*
- 3) ?? *Un homme est une sorte d'homme.*
- 4) ? *Dans les hommes, il y a les hommes et les femmes.*
- 5) ? *L'homme en tant qu'individu de sexe masculin est un homme en tant que membre d'une espèce.*

En cas d'emploi simultané des deux notions, l'une d'entre elles sera probablement remplacée par son synonyme habituel :

- 6) *L'espèce humaine comprend les femmes et les hommes.*

De ce point de vue, il est difficile de parler de polysémie verticale sans étendre le champ de la notion. En premier lieu, en effet, la plupart des noms concrets autorisent deux lectures, une générique et une spécifique, l'une englobant l'autre sur le plan notionnel. Cela est particulièrement vrai pour les noms d'espèces animales ou pour tous les ensembles où l'on a une dualité de genre ou de sexe. Ensuite, ce phénomène peut être étendu aux verbes car ceux-ci peuvent être considérés comme hyperonyme ou hyponyme d'eux-mêmes, à l'observation de leurs acceptions dans un dictionnaire courant. Le verbe *abandonner* a par exemple un nombre d'acceptions assez faible qui toutes peuvent être englobées sémantiquement par un sens primitif peu détaillé qui peut correspondre littéralement à une définition valable d'*abandonner*. Il y aurait par conséquent un proto-verbe *abandonner* qui exerce sa prédominance « verticale » sur toutes ses acceptions. Nous nous servons de ce point de vue pour écrire une primitive sémantique de ce verbe afin de faire ressortir l'homogénéité de ses emplois.

2.1.4.2 La polysémie logique

Caractérisée par James Pustejovsky (1996), la polysémie logique désigne le phénomène de cohabitation entre deux acceptions en apparence contradictoires mais liées pour le même objet linguistique (cf. Jacquy 2001). L'exemple le plus cité est celui de *livre*, dont chaque référent est aussi bien un artefact, objet manipulable, qu'un contenu informationnel. L'artefact possède une forme, une apparence, des couleurs et toutes autres caractéristiques de son état. Le contenu informationnel peut être qualifié de « plus ou moins intéressant », « de plus ou moins bonne qualité », etc. Il peut y

avoir concurrence entre les deux acceptions car, certes elles désignent le même objet, mais elles ne peuvent pas toujours être employées conjointement dans la même occurrence, sauf contraintes particulières portant sur la qualification :

- 7) *Le livre est rouge* (artefact)
- 8) *Le livre est intéressant* (contenu)
- 9) **Le livre est rouge et intéressant* (conflit artefact/contenu)

Si le mode attributif ne peut qualifier les deux sens en même temps, ceci n'empêche pas de les confronter dès lors que la qualification porte déjà sur *livre* en position d'épithète :

- 10) *Le livre rouge est intéressant*
- 11) *Ce passionnant livre est rouge*

Dans un souci de caractérisation du sens, il est difficile de dire que l'on a affaire à deux sens différents d'un même terme puisque les deux définitions peuvent renvoyer au même objet, en l'occurrence l'artefact livre et son contenu.

En essayant d'étendre cette notion aux verbes, nous nous apercevons que ces derniers résistent à la polysémie logique car leur rayonnement sémantique en tant que prédicats touche directement les arguments qui l'accompagnent. Par exemple, on acceptera difficilement :

- 12) ??*J'ai abandonné ma lecture (ou mon livre) et mes enfants,*

non pas parce que le sens du verbe est profondément différent mais parce que ses arguments objets induisent une incompatibilité subtile qui bloque leur co-occurrence.

Nous verrons vers la fin de ce chapitre comment Pustejovsky applique ce principe pour expliquer pourquoi certains verbes autorisent la sous-catégorisation d'objets qui ne sont pas prévus dans sa structure télique habituelle (cas de *to bake*).

2.1.5 Polysémie et traduction¹¹

2.1.5.1 Opposition de mondes lexico-conceptuels

2.1.5.1.1 Instabilité intégrale

Dans l'ouvrage collectif *La Traductique* (1993), l'instabilité des mondes lexico-conceptuels de langues différentes est présentée de la façon suivante par Pierrette Bouillon :

« L'analyse des langues naturelles se heurte en effet au problème de l'*ambiguïté* des langues naturelles, ambiguïté à la fois *lexicale* et *structurale*. On parle d'ambiguïté lexicale, quand un mot peut recevoir plusieurs catégories syntaxiques ou significations. Ce terme englobe aussi les ambiguïtés bilingues (ou *contrastives*) qui désignent deux mots qui, tout en étant la traduction l'un de l'autre, ne dénotent pas exactement le même concept ou le même ensemble de concepts. Le terme d'ambiguïté structurale fait référence à une phrase qui peut être dotée de plusieurs représentations. »
Bouillon (1993 :19)

L'accent est mis sur les écueils successifs de « l'ambiguïté des langues naturelles » et des « ambiguïtés bilingues ». La première est mise en évidence par l'actualisation des dictionnaires, comme nous le décrivons par ailleurs, ainsi que le caractérise Paul IMBS lorsqu'il parle de « continuum instable » pour désigner le lexique d'une langue (Introduction au premier volume du TLF, 1963). On la retrouve également dans le phénomène de polysémie et de polymorphie, dans la mesure où peu d'unités de la langue présentent une réelle stabilité sémantique et/ou syntaxique.

Les ambiguïtés dites bilingues sont tout aussi nombreuses et complexes. Tout d'abord, les mondes lexico-conceptuels de deux langues différentes ne sont pas parallèles, i.e. à un mot d'une langue donnée ne correspond presque jamais un mot et un seul dans l'autre langue. Les décalages entre les concepts ou les ensembles de concepts sont donc la règle. D'autre part, certains pans de la langue ne peuvent pas être rendus par la traduction, comme c'est généralement le cas pour les jeux de mots¹².

¹¹ Il sera davantage question ici des problèmes liés à la traduction et aux ambiguïtés diverses qui peuvent exister entre les langues que du rapport *stricto sensu* entre polysémie et traduction.

¹² Nous citerons un exemple tiré du roman de John Irving, « l'épopée du buveur d'eau », où le traducteur se voit obligé de signaler en note qu'il ne peut pas rendre en français le comique généré par la version anglaise. Dans cet exemple, une dame demande à son mari ce qui arrive à leur fils : « *Qu'est-ce qui se passe ? – Vérole, lui crie-t-il !* » et la dame tape des mains en silence. Ce jeu de mots est intraduisible puisque *clap* en anglais signifiait aussi bien *applaudir* (comme traduction de *to clap*) que *vérole* en langage familier. On trouve régulièrement ce genre de note de bas de page dans les romans traduits.

2.1.5.1.2 L'ambiguïté selon l'Humain...

À contexte suffisant, on ne rencontre pas vraiment d'ambiguïtés réelles dans la langue. Tout le monde connaît le processus de polysémie et « la polysémie n'est pas l'ambiguïté » nous rappelle Catherine Fuchs en 1996.

Quand une autre langue est en jeu, le danger du décalage augmente bien entendu. Pour autant, nous ne traiterons pas de la traduction littérale car, bien qu'elle semble proche de la traduction mécanique, elle relève avant tout d'une forte incompétence linguistique qui fait établir une biunivocité syntaxique et lexicale absolue à défaut de maîtriser les décalages.

Pour illustrer le faible impact des ambiguïtés sur l'humain, Pierrette Bouillon propose les exemples suivants :

13) *La fermière a tué la vache parce qu'elle ne donnait pas de lait*

14) *Je vois l'homme avec le télescope*

Selon elle, la première phrase n'est pas ambiguë pour l'être humain, c'est une ambiguïté *accidentelle* que la machine aura à résoudre. Quant à la seconde, elle serait ambiguë pour l'humain comme pour la machine, soit *réelle*. Notre lecture innocente de ces deux phrases montre à quel point la notion d'ambiguïté appliquée à des exemples construits n'est guère convaincante quand elle est péremptoire. En effet, la phrase (14) nous semble plus improbable qu'ambiguë et ne relève pas vraiment de la sphère de la langue, à moins de ne correspondre qu'à la seconde interprétation (c'est « l'homme » qui tient le télescope). Quant à la (13), nous avons tout de suite imaginé, avec un rien de mauvaise foi, une pauvre fermière ne donnant pas de lait et tuant sa vache de rage.

Ce type d'ambiguïté est intrinsèquement lié à la non-saturation apparente du référent du pronom qui suit la proposition principale (*elle* dans (13)). Observons maintenant l'extrait suivant tiré de la fable de la Fontaine *le cheval et l'âne* :

« (...)
Un âne accompagnait un cheval peu courtois,
Celui-ci ne portait que son simple harnois,
(...) »
Jean de La Fontaine

À partir de cet exemple, Maingueneau (2003 : 221, 236-237) propose une interprétation du pronom anaphorique « celui-ci » en ces termes :

« La reprise par *celui-ci* permet d'éviter l'ambiguïté de l'anaphore. »
Maingueneau (2003 : 237)

Si le pronom démonstratif est bien désambiguïsant, cela suppose que l'exemple (13) proposé par Pierrette Bouillon doit être considéré d'emblée comme ambigu.

L'utilisation de *celle-ci* aurait permis de rattacher son référent à celui de *la vache*, ce qui n'est pas le cas ici. En définitive, une telle ambiguïté apparente est en principe résolue par notre connaissance du monde, l'interprétation de la vache comme étant l'être « qui ne donnait pas de lait » prime par conséquent sur l'autre. Toutefois, rien n'interdit de penser que la fermière ne puisse pas donner de lait, il s'agit d'une situation envisageable dans le monde où nous sommes, même si cela n'apparaît pas spontanément comme étant le meilleur choix pour cette occurrence.

En définitive, générer de l'ambiguïté en discours est un exercice difficile et périlleux dans la mesure où les analyses sont elles-mêmes potentiellement ambiguës, comme nous avons essayé de le montrer, surtout si les exemples sont construits et analysés en dehors d'un co-texte désambiguïsant.

Il n'en est pas moins vrai qu'on s'entend souvent dire par des relecteurs ou des auditeurs : « à quoi as-tu voulu relier ce complément ? », « le sujet de ta phrase, c'est lequel ? ». Dès qu'on dépasse la structure simple Sujet-Verbe-Objet, le risque augmente d'aboutir à des ambiguïtés structurales, beaucoup plus fréquentes que les lexicales qui profitent plus du co-texte pour être résolues d'emblée.

2.1.5.1.3 ...et selon la machine

Pour la machine, tout est ambigu par définition... et rien n'est ambigu par absence de raisonnement conscient.

L'humain comme la machine biaisent aisément l'ambiguïté structurale dans le cas où les langues mises en rapport sont proches :

Si la résolution de toutes les ambiguïtés accidentelles n'est pas toujours nécessaire pour traiter des langues proches, elle est malheureusement indispensable pour traduire correctement des langues de familles plus éloignées. Ainsi, dans un système qui traite du français et de l'anglais, l'ambiguïté de la phrase 1 (NDLR : *L'homme a vu le cheval avec le télescope*) ne doit pas être levée et il suffit d'ordonner les modificateurs de la même façon dans les deux langues. Mais dans un système qui traduirait vers le japonais, cette distinction deviendrait obligatoire pour produire une phrase syntaxiquement correcte.

Il est donc souvent nécessaire de choisir la bonne interprétation, ce qui nécessite des connaissances linguistiques, contextuelles, générales, ou statistiques et même des capacités de raisonnement. Si les recherches ont beaucoup progressé dans ce domaine, la représentation et l'intégration de ces sources de connaissances restent encore souvent expérimentales et on se pose souvent la question de savoir jusqu'à quel point ce type d'informations est suffisant et adéquat pour résoudre les problèmes d'analyse.

Bouillon (1993 :19)

Un peu plus loin, Pierrette Bouillon pointe sur un problème d'autant plus important qu'il est difficile à traiter : la prise en compte des variantes contextuelles

dans l'interprétation des énoncés. C'est la connaissance de ces variantes qui permet à l'humain de n'être presque jamais décontenancé par la polysémie des termes mis en rapport, alors que la machine n'est que rarement capable de la dépasser :

« la plupart des systèmes traduisent phrase par phrase, sans en tenir compte »
(op.cit.).

Il reste que la machine ne peut résoudre que ce qu'on lui a appris à résoudre. Dès lors que les « bonnes » règles lui sont apprises, un programme informatique cessera de générer de représentations identiques pour des formes identiques, et s'adaptera aux nouveaux modèles implantés à la hauteur de leur qualité. Pour cela, les systèmes de traduction automatique qui entendent tenir compte du co-texte doivent passer par l'écriture de règles n-grammes et par la prise en compte des éléments contextuels appropriés (dits « déclencheurs » ou « triggers » en anglais).

2.1.5.2 Les vraies relations d'un terme avec ses traductions

Proposer une traduction d'un terme source revient à proposer un équivalent contextuel de celui-ci. Dans la plupart des cas, les traductions ne sont pas univoques, c'est-à-dire que la plupart des lexies ont plusieurs traductions (parfois énormément) et que chacune de ces traductions peut être considérée comme « la plus appropriée » ou la meilleure pour un contexte donné.

L'exemple du verbe *abandonner*, que nous présentons par ailleurs au chapitre II, est une excellente illustration du potentiel d'essaimage sémantique d'une unité à polysémie peu marquée. En regard de la lexicographie monolingue, *abandonner* et les verbes qui lui ressemblent sont soumis à une polysémie assez faible dans le sens où il existe un noyau de sens commun à toutes les acceptions du verbe. Ce qui est notable, c'est que les traductions de ces verbes en contexte se choisissent en fonction des arguments satellitaires du verbe bien davantage qu'en fonction de l'acception qu'ils induisent. De ce fait, les critères distinctifs dont on aura besoin pour sélectionner les traductions justes seront plus axés sur les structures argumentales (collocations) et sur les normes de registre et de contexte qui ne sont pas nécessairement liées au dégroupement initial des verbes en acceptions.

Comme dit précédemment, chaque traduction est supposée incarner « la meilleure » de sa catégorie. Toutefois, il s'agit là d'un point de vue difficile à systématiser dans la mesure où certains termes cibles restent des seconds choix ou des avatars de contextes luxueux, ce qui explique en partie leur absence des dictionnaires en tant que candidats visibles. Outre les exemples déjà mentionnés pour les traductions rares du verbe *abandonner*, nous pouvons citer la longue liste des verbes italiens traduisant *aborder* ou étant traduits par ce verbe : plusieurs de ces derniers sont des termes littéraires ou désuets qui n'ont pas lieu d'apparaître comme candidats directs selon le rédacteur (*avvicinare*, *appropinquare*) ; d'autres, comme *sbarcare*, possèdent

des acceptions qui évoquent la spécialisation ou l'extension d'une des acceptions de *aborder* en français sans toutefois en être suffisamment proches pour le traduire directement. Nous étudierons en détail cette relation au chapitre II.

Dans de nombreux cas, les dictionnaires bilingues proposent deux, parfois trois, très exceptionnellement quatre traductions pour un même terme dans des conditions d'emploi équivalentes. Dans une telle situation, il n'y a que rarement mention d'une hiérarchie préférentielle entre ces traductions proposées, ni davantage de marqueurs d'emploi spécifiques. Si l'on en reste à la lecture brute de l'entrée du dictionnaire, il est permis de considérer ces différents candidats comme des équivalents contextuels en langue cible. Or, on admet qu'il n'y a stricte équivalence sémantico-lexicale dans une langue donnée que par accident, ce qui complique notre vision du lexique.

La synonymie accidentelle est tout de même l'événement le plus saillant dans une telle situation. Conséquemment, les traductions entretiennent entre elles des relations de type synonymique, dites contextuelles car liées à un emploi particulier précisé par le dictionnaire.

Pour comprendre comment fonctionne le paradigme de traductions des traductions, il faut aller observer le processus de balayage¹³ et ce qui en découle. Par le « jeu du miroir qui ne réfléchit pas¹⁴ », la traduction d'une traduction n'est pas statistiquement le mot à traduire initialement, d'où d'ailleurs la raison d'être fondamentale de notre thèse. C'est également le moyen d'évaluation pour le système de traduction SYSTRAN. En recourant systématiquement à ce mode de double lecture, la concurrence entre traductions peut partiellement être résolue. Toutefois, il s'agit d'une gymnastique contraignante qui n'offre pas de ferme garantie quant au résultat.

Après avoir exploré la notion de polysémie puis établi que les traductions entretenaient des rapports d'équivalence subtils et codifiés, nous nous retrouvons face à quelques questionnements de poids : Une traduction peut-elle être considérée comme le synonyme en langue cible de la lexie¹⁵ source ? Et dans ce cas, la meilleure traduction n'est-elle point le meilleur synonyme d'une lexie, au moins sur le plan théorique ? Plus globalement, deux mots de langues différentes peuvent-ils être perçus comme synonymes l'un de l'autre ?

En ôtant de la définition de la synonymie le trait d'équivalence idiomatique, il reste que le signifié est potentiellement le même dans toutes les langues, simplement réalisé par des signifiants différents. D'autre part, les traductions des traductions ajoutent des traits de synonymie jusqu'alors invisibles, lesquels viennent s'insérer dans un réseau global de proximité sémantique.

¹³ Le balayage est le processus de récupération des données que nous utilisons pour construire nos bases. Son mode de fonctionnement a été évoqué dans l'Introduction et sera plus largement décrit aux chapitres II et V.

¹⁴ C'est-à-dire la confrontation du LI et du LD sans faire leur fusion.

¹⁵ La lexie est à prendre dans le sens de Mel'çuk et alii, eux-mêmes s'étant inspirés de la définition de Pottier (présent chapitre, 2.3.1).

Dans leur présentation du programme de traduction ULTRA (in Bouillon & Clas, 1993), Helmreich, Jin, Wilks et Guillen définissent la notion d'équivalence de traduction après avoir discuté des différents niveaux d'ambiguïté qui perturbent l'exercice de traduction automatique¹⁶. Les différentes ambiguïtés qu'ils relèvent sont au nombre de quatre :

- l'ambiguïté lexicale. Elle correspond au croisement entre la polysémie inhérente au lexique de chaque langue et les traductions possibles dans une autre langue. Un mot d'une langue a potentiellement plusieurs sens et donc potentiellement plusieurs traductions en langue cible. À l'inverse, plusieurs mots d'une même langue peuvent être traduits par un et un seul mot dans une autre langue (cas de *to hope*, *to wait for* et *to expect* traduisibles par le même verbe espagnol *esperar*) ;
- l'ambiguïté syntaxique. Cette dernière complique la situation car une phrase structurellement ambiguë dans une langue ne l'est pas forcément dans une autre (les auteurs reprennent la phrase de Bouillon en anglais *I saw the man with the telescope* dont la structure « doit être désambiguïsée car la traduction chinoise varie (...) » (Helmreich et alii, 1993 :336)) ;
- l'ambiguïté sémantique. Elle concerne notamment les cas de sens figurés, les métaphores et les métonymies pour lesquels une traduction littérale n'est pas souvent possible. Sans ambiguïté syntaxique, un énoncé peut souvent avoir un sens littéral et un sens figuré ce qui complique un peu plus la tâche de désambiguïsation et de traduction. Là aussi, ce problème induit qu'on sache « pourquoi » un énoncé est produit afin de savoir « ce qu'il veut dire » ;
- l'ambiguïté discursive. Résoudre la question de la recherche des référents des anaphores est une opération délicate et nécessaire car elle

¹⁶ « L'une des caractéristiques communes aux langages naturels est son ambiguïté. Le nombre d'interprétations possibles d'une phrase donnée dans une langue est en effet très important et est causé par le fait que chaque phrase est appropriée à un grand nombre de situations. La sémantique situationnelle parle à ce propos de l'« efficacité » du langage (Barwise et Perry 1983). La phrase « il n'est pas là », par exemple, peut être prononcée dans des circonstances très variées avec des interprétations différentes de « il » et de « là ». (...) » (Helmreich, Jin, Wilks et Guillen, 1993:335-336).

On reconnaît dans ces lignes les tenants de la polysémie absolue et, à demi-mot, l'influence du contexte pragmatique sur l'interprétation des énoncés : désambiguïser les anaphores et déictiques ne suffit pas, il faut également savoir « pourquoi » un énoncé est produit, comme les auteurs le précisent d'ailleurs pour expliciter le fonctionnement de leur programme.

conditionne beaucoup de critères syntaxiques et lexicaux dans une langue cible. « Pour traduire en français le texte anglais *John took the fruit from the table. Then he cleaned it*, il faut décider si *John* a nettoyé “le fruit” ou “la table”, puisque cette information est nécessaire pour déterminer le genre du pronom (...) et parfois aussi le choix du verbe. » (ibid. : 337). Dans cet exemple, heureusement, il n’y a pas d’ambiguïté sur l’autre pronom, à savoir « he », qui ne peut se rapporter qu’à « John ».

Une fois les différents niveaux d’ambiguïté définis, et éventuellement résolus, on s’aperçoit que la question de l’équivalence interlangue est toujours en suspens car la désambiguïsation n’implique pas la résolution de tous les problèmes liés à la traduction. En effet, les obstacles liés à la non biunivocité entre les lexiques, les grammaires et même les cultures pour chaque langue demeurent. Pour autant, nous avons choisi de ne pas remettre en question notre principe d’équivalence qui consiste en l’acceptation qu’une « meilleure » traduction d’un terme source dans un contexte donné constitue le « meilleur » synonyme de ce terme parmi d’autres « meilleurs » synonymes dans d’autres contextes.

En mêlant cette vision au phénomène de la polysémie absolue, nous parvenons à l’interrogation suivante : Y a-t-il toujours un « meilleur candidat » pour tous les contextes d’apparition d’une vedette ou bien est-ce une proposition pour pallier les effets de la polysémie contextuelle ? Autrement dit, il n’y aurait meilleur candidat qu’après passage dans le filtre *sens d’origine/encodage/interprétation*, et ce seulement si le problème lié à l’équivalence a été surmonté.

2.1.6 La polysémie remise en question

En acceptant une fois pour toutes la pertinence du parti pris de la polysémie « absolue » ou systématique, selon laquelle chaque occurrence d’un mot est dissemblable de toutes les autres, on aboutit à une forme de neutralisation de la polysémie décrite depuis Bréal.

Avant d’en arriver à ce point, notons que la polysémie en tant que concept est irréductible dans le travail métalinguistique pour expliquer le potentiel et l’évolution sémantique d’une forme donnée, et indispensable en métalexigraphie pour préparer et détailler le dégroupement des termes en acceptions aux caractéristiques différentes. Autrement dit, prétendre à notre niveau et pour ouvrir l’accès à nos modes de représentation personnels que « la polysémie n’existe pas » n’empêche pas la notion d’exister en tant que concept essentiel pour l’analyse linguistique.

Dès que l’on quitte le niveau de méta-analyse et que l’on cherche à rétablir l’intégralité des facteurs qui participent à l’interprétation des énoncés en situation, la polysémie se dilue peu à peu au contact des discriminateurs forts que sont le contexte

(au sens large) et les perceptions conjointes et volontiers incompatibles de l'encodeur et de l'interpréteur, lesquels n'ont évidemment pas la même vision de la situation d'énonciation ni de la finesse du sens, et conséquemment font que les locuteurs d'une même langue voient chaque occurrence différemment de leurs co-locuteurs.

En considérant dès lors que la polysémie est, au niveau de l'énoncé, un phénomène systématique qui permet de caractériser la différence entre deux occurrences identiques en forme mais dans des contextes différents, nous aboutissons à une première neutralisation de la notion en ce qu'elle représente fondamentalement. Par ailleurs, un même couple énoncé-contexte sera interprété différemment par des locuteurs distincts mais sans en dénaturer le sens originel, ce qui conduit à une seconde neutralisation au niveau le plus fin, celui de l'interprétation exacte : le niveau du récepteur indivisible. Il serait fastidieux d'aller rechercher un troisième niveau de neutralisation visant à démontrer qu'un seul interpréteur est susceptible de projeter plusieurs significations sur un même couple énoncé-contexte, dans la mesure où ce processus n'a pas de fin.

En guise de conclusion à notre réflexion sur la polysémie absolue, il est manifeste que cette dernière n'a plus une utilité majeure dans le cadre de nos travaux, puisqu'elle est absolue justement, et que ce n'est pas en son nom que nous allons définir les liens qui unissent les lexies dans nos représentations. En revanche, elle demeure indispensable en tant qu'outil d'analyse en métalangue pour les sémanticiens et les lexicographes soucieux de réduire le sens à des formalisations ou des rédactions à taille humaine. Pour une poursuite de la réflexion, nous renvoyons à Rastier et Victorri au sujet de l'opposition entre type (élément ultra-polysémique puisque utilisable dans un nombre infini de contextes) et occurrence (élément monosémique).

Pour éviter les effets néfastes de ce que nous venons de décrire, il nous a fallu effectuer un détour par la synonymie afin d'explicitier non plus le fonctionnement interne d'un lexème en fonction de ses entrées lexicales mais plutôt le type de liens qu'entretient une lexie avec les autres lexies. Pour ce faire, nous allons explorer le mécanisme de la synonymie orienté vers nos besoins, en n'oubliant pas de définir plus précisément les unités lexicales minimales sur lesquelles nous voulons travailler.

En définitive, toute notre réflexion mène à la conclusion qu'il existe deux polysémies bien distinctes qui ne sauraient finalement être opposées sur un même plan.

2.2 Synonymie

2.2.1 Définition

Dans notre esprit, la synonymie est en quelque sorte la polysémie du sens primitif, celui qui n'est pas directement verbalisé. Voici quelques définitions de la synonymie :

“Synonymy

Semantic relation of sameness or (strong) similarity in meaning of two or more linguistic expressions. In lexicology, grammar, or stylistics it is a term whose interpretations are as varied as the semantic theories in which it is found. (...) »

Bussmann (1996:470)

Ce qui revient à dire que la synonymie, comme la polysémie du reste, est une notion qui ne se laisse pas facilement “apprivoiser” dans les théories linguistiques, comme nous le savons depuis bien longtemps.

« The following distinctions are generally made. (a) Complete (absolute, strict or pure) (...): by definition, complete synonymy presupposes the unconditional substitutability of the given expressions in all contexts and refers both to denotative and to connotative semantic elements. (...) it appears that the concept of linguistic economy eliminates, in almost all cases, the possibility of complete synonymy at least in lexemes. (b) Partial synonymy, which refers either to lexemes which can be substituted in some but not all contexts (...) »

Bussmann (op.cit.:470-471)

D'après cette définition, la seule synonymie réellement active dans la langue est la synonymie partielle, car la « synonymie complète » suppose d'outrepasser le principe d'économie linguistique, laquelle prévaudrait sur la possibilité de substituer une unité à une autre dans tous les contextes. Le même décalage existe dans un rapport interlangue pour des raisons extérieures à l'économie du langage : il n'y a aucune raison *a priori* pour que deux langues adoptent le même système lexico-conceptuel dans une parfaite symétrie (ce qui au demeurant rendrait inutiles les recherches en traductologie).

2.2.2 Application aux lexiques multilingues

Les traductèmes¹⁷ sont assimilables à des synonymes contextuels « parfaits » du mot qu'ils traduisent. Ils respectent la définition de la synonymie de Bussmann en ce qui concerne les synonymes partiels. D'autre part, cette définition ne mentionne pas de restriction au sujet de la langue des unités mises en rapport. Nous en déduisons que deux mots de langues différentes peuvent être synonymes l'un de l'autre. C'est à partir de ce postulat que nous allons mener notre constitution d'un algorithme de construction de graphes lexicaux.

Entre la synonymie et la polysémie, pour le choix de la meilleure matière notionnelle pour la suite de nos travaux, la question ne se pose pas réellement en terme d'efficacité mais plutôt d'un point de vue de la réalité linguistique. Toutes les relations définies au moyen d'un dictionnaire bilingue se ramènent à des liens de synonymie ou ressemblant à de l'équivalence sémantique.

D'une certaine façon, nous soutenons que la synonymie est probablement moins un artefact linguistique que ne l'est la polysémie, dans la mesure où la première est fréquemment induite par un mécanisme conscient, soit pour trouver des équivalents dans le discours, soit pour rechercher des équivalents en traduction. La synonymie induit une stabilité du sens ou à tout le moins une forte ressemblance entre sens.

Les liens de synonymie que nous allons défendre au chapitre IV concernent non plus des unités lexicales au sens large, i.e. des formes, mais des emplois particuliers de chacun des mots mis en rapport. La polysémie absolue transforme un mot en une multitude d'entrées lexicales caractérisables (en théorie). La synonymie contextuelle va à l'inverse juguler l'expansion induite par cette « ultra-polysémie » en restreignant les unités lexicales aux couples acceptions-traductèmes présents dans les dictionnaires. En des termes plus simples, nous dirons qu'il n'y a pas sémantiquement plus proche d'une entrée lexicale que ses synonymes, d'un point de vue lexical ; en tout état de cause un synonyme d'une entrée lexicale d'une unité polysémique est toujours plus proche de la première que de toutes les autres entrées lexicales de l'unité polysémique.

Pour exploiter ces notions dans l'optique de notre modélisation, nous allons maintenant expliciter quels éléments lexicaux nous allons observer, et quelle terminologie adopter, à commencer par la « lexie » telle que Pottier (1991) et Mel'çuk & alii (1995) l'entendent.

¹⁷ Un traductème est un mot proposé en traduction d'un autre mot selon des critères de sélection précis. On peut l'assimiler à « la plus petite unité de traduction » pour une expression donnée.

2.3 La place des lexies dans le lexique : définitions et notions

2.3.1 Qu'est-ce qu'une lexie

A l'instar de Mel'çuk, Clas et Polguère (1995), nous utilisons la lexie comme unité de base pour nos représentations. La définition d'une lexie pose certains problèmes, en particulier celui d'être tributaire du concept de *mot*, lequel n'est pas précisément défini en linguistique :

« La lexie est l'unité de base de la lexicologie –en fait, son objet central et même, en schématisant un peu, son seul et unique objet. Par conséquent, le concept de lexie devrait être défini dès à présent de façon rigoureuse et muni de toutes les explications indispensables. (...)»

Le concept de lexie est une formalisation et, simultanément, une généralisation de la notion de MOT. Il n'est pas dans notre intention d'entreprendre ici une analyse poussée de cette notion, même si l'on sait bien que le mot est une unité centrale de la langue : on connaît tout aussi bien le caractère rétif du mot *mot*, qui, jusqu'à présent, a échappé aux tentatives de le circonscrire avec précision et a fait couler beaucoup d'encre pendant des décennies. (...)»

Pour le moment, il nous suffit de dire qu'une *lexie*¹⁸ ou *unité lexicale*, est soit un mot pris dans une acception bien spécifique (= *lexème*), soit encore une locution, elle aussi prise dans une acception bien spécifique (= *phrasème*). Ainsi, si nous considérons le mot PONT, nous pouvons en dégager, de façon intuitive et en nous fiant à des dictionnaires, à peu près une quinzaine de lexies (...) :

Lexie 1 : PONT au sens de 'construction reliant les deux rives d'une étendue d'eau...';

Lexie 2 : PONT au sens de 'jours chômés entre deux jours fériés';

Lexie 3 : (...)

...

De plus, on connaît toute une série de locutions comprenant le mot PONT, qui constituent d'autres lexies :

Lexie 6 : PONTS ET CHAUSSEES au sens de 'service administratif des voies publiques...';

Lexie 7 : PONT AUX ÂNES au sens de 'banalité connue de tous';

Lexie 8 : PONT AERIEN au sens de 'liaison régulière par avion entre deux points...';

Lexie 9 : COUPER LES PONTS [*avec N*] au sens de 'cesser délibérément les relations [*avec N*]...';

Telle que présentée dans ces exemples, une lexie est une entité trilatérale : elle a

un sens (le *signifié* saussurien),

¹⁸ Les auteurs précisent qu'ils emploient ce terme « tel que proposé et défini par B. Pottier (1991) »

une forme phonique/graphique (le *signifiant* saussurien),
et un ensemble de traits de combinatoire (le *syntactique* de la théorie Sens-
Texte : Mel'çuk 1993 : 117). »
Mel'çuk, Clas & Polguère (1995 : 15-16)

Ainsi définie l'unité lexicale, dite *lexie*, s'oppose au lexème, distincte du *mot* en cela qu'elle est caractérisée par un couple forme-acceptation. Le lexème est d'abord du domaine du *signifiant*, modérée par des critères distinctifs fondés sur les différences catégorielles (homographie avec catégories grammaticales différentes), l'origine du mot (homonymie par étymons différents) et l'éloignement de ses acceptations contemporaines (homonymie de sens synchronique). Toutefois, ce dernier critère est peut-être discutable en lexicographie. L'unité lexicale est pour sa part tributaire à la fois du signifié et du signifiant, avec des variations de finesse dans sa description en fonction des choix du lexicologue.

Pour la suite du présent document, nous choisissons de suivre *grosso modo* les définitions proposées par Mel'çuk et alii. Par conséquent, nous utiliserons *lexème* ou *terme* pour désigner le verbe en tant qu'entité polysémique non marquée ; par ailleurs, nous emploierons *lexie* pour faire référence au verbe pour un sens particulier¹⁹. Les lexiques inversés que nous avons construits et que nous définissons au chapitre II sont des ensembles de *lexèmes* à *candidats* multiples. Les bases lexicales qui leur sont associées sont des paradigmes de *lexies*, chaque ligne faisant référence à un type d'emploi bien déterminé (traductions, collocations et données extralinguistiques²⁰ à l'appui).

La définition de la *lexie* selon Mel'çuk n'exclut pas la prise en compte des éléments polylexicaux. Étant donné la grande variété de ces derniers, nous allons essayer de fixer les limites qui permettront à un terme constitué de plusieurs unités d'entrer dans nos modèles.

2.3.2 Peut-on être polylexical et homogène en même temps ?

Après avoir passé en revue un certain nombre de définitions « non superposables » de différents auteurs (Jespersen, Pottier, Benveniste, Martinet, Larousse, Darmesteter, Fonagy, Mathieu-Colas, Guiraud, Catach) concernant *le figement, les expressions idiomatiques, le mot composé, la composition, etc.*, Gaston Gross propose une classification simple, issue de la synthèse des premières, en

¹⁹ À savoir pour un ensemble de traits de sélection suffisants pour aboutir à une traduction unique.

²⁰ Les données extralinguistiques sont les éléments non langagiers de la situation d'énonciation qui interviennent dans la sélection des candidats à la traduction. On en rencontre peu dans les dictionnaires car leur nombre potentiel est trop élevé par rapport aux capacités de ces derniers.

distinguant deux grandes classes de mots, les mots *simples*²¹ et les mots *construits*²², ces derniers donnant lieu à une nouvelle distinction entre mots *dérivés* et mots *polylexicaux* :

« Parmi les mots construits nous distinguons :

-les mots **dérivés**, ceux qui sont formés à l'aide d'un affixe : préfixe (*refaire*) ou suffixe (*travailleur*) ;

-des mots **polylexicaux** (ou mots **complexes**), correspondant à toute unité (ou catégorie grammaticale ou partie de discours) composée de deux ou de plusieurs mots simples ou mots dérivés préexistants (...). Cette définition s'applique à des termes comme *jardin d'agrément* mais non à des suites du type *le jardin de ton père*, qui est groupe nominal comprenant un substantif simple et sa détermination. Alors que les mots dérivés sont toujours soudés, les mots polylexicaux (ou complexes) peuvent être soudés (*outrépasser*, et alors, du point de vue informatique, ils peuvent être assimilés à des mots simples) ou comporter un séparateur (blanc, apostrophe, trait d'union, etc.). »

G. GROSS (1996 : 7)

Comme le précise l'auteur un peu plus loin, le problème de la soudure n'a pas la même portée ni la même signification d'une langue à l'autre. En effet, le système lexical de l'allemand intègre la soudure comme procédé de dérivation systématique en une seule unité monolexicale construite, là où en français il est possible - et généralement plus fréquent - de proposer un syntagme nominal non soudé (voir l'exemple de l'allemand *ölmühle* traduisible par *moulin à huile*). De ce point de vue, nous allons être confronté à un choix délicat : Devrons-nous traiter systématiquement les mots complexes non soudés dès lors qu'ils ne sont pas des groupes nominaux libres ? Ou nous contenterons-nous d'intégrer comme lexies les ensembles composés à sens opaque ? Les deux solutions posent problème. En effet, d'une part, traiter *moulin à huile* comme un lexème normal induit la prise en compte d'un nombre incalculable de mots polylexicaux en français ; d'autre part, les ignorer dans le champ d'étude risque de poser un problème de correspondance interlexicale au moment où nous entendrons traduire *ölmühle*.

La solution provisoire que nous préconisons est d'abord d'admettre l'extension de la notion de *lexie* à des compositions non opaques comme *moulin à huile*, *moulin à farine*, etc. D'une certaine façon, le phénomène de dérivation décrit plus haut pour *ölmühle* et *moulin à huile* ne pose pas vraiment de problème de correspondance si les processus de dérivation sont stables et donc aisément repérables en français et en allemand. En revanche, en l'absence d'une stabilité totale dans le processus de genèse des mots construits, comme nous pouvons nous y attendre, nous

²¹ G. Gross appelle **mot simple** ou mot racine « toute unité qui n'est susceptible d'aucune décomposition, comme *lune*, *travail* ou *petit*. »

²² Toujours selon G. Gross, tout mot qui ne rentre pas dans la catégorie des mots simples est dit **construit**.

devrons nous contenter d'intégrer à la liste des lexies les mots polylexicaux que les dictionnaires bilingues font apparaître comme vedettes (traduits par des mots monolexicaux ou polylexicaux) ou comme candidats (traductions de mots monolexicaux ou polylexicaux). Dans ces conditions, il suffirait juste qu'un mot polylexical soit présent dans le dictionnaire pour devenir une lexie.

En ce qui concerne le statut et les limites des locutions verbales traduites par des verbes monolexicaux, nous rencontrons un autre problème : comment déterminer le degré de figement d'une locution verbale ?

Au-delà des critères de détermination du degré de figement des expressions polylexicales (opacité sémantique²³, blocage des propriétés transformationnelles²⁴, non-actualisation des éléments²⁵), Gaston Gross (GROSS 1996 :9-23) précise le statut des locutions par comparaison de leurs propriétés avec celles des termes monolexicaux :

« Quand on examine les listes de séquences contraintes élaborées au LADL et au LLI²⁶, on constate que les suites totalement figées sont très minoritaires par rapport à celles qui ont des restrictions partielles. On peut d'abord se demander si les suites totalement figées doivent être assimilées aux catégories simples. *Cordon(-)bleu* est-ce un nom ? *Casser sa pipe* est-ce un verbe ? La réponse n'est pas simple.

Du point de vue syntaxique, ces suites ont le même fonctionnement que les catégories simples correspondantes. *Cordon bleu* peut avoir la même distribution que (*bonne*) *cuisinière* : (...)

Cependant, on ne peut pas assimiler le nom composé à une catégorie simple, pour des raisons de nature diverse. Comme il y a deux mots différents, la mise au pluriel exige deux marques (*des cordons bleus*) et le traitement informatique se trouve de plus confronté à des difficultés spécifiques. Comme il y a un séparateur entre les deux mots, il faut informer le système que ces deux mots ne forment en fait qu'une seule unité fonctionnelle et, dans une perspective de génération automatique, que les relations habituelles entre un substantif et l'adjectif qui l'accompagne sont bloquées. Des réflexions analogues peuvent être faites à propos de *casser (sa) pipe*. La distribution est celle d'un verbe :

Paul est mort dans un accident de voiture

²³ « L'opacité sémantique » intervient lorsque la lecture d'un énoncé n'est pas compositionnelle.

²⁴ « Le blocage des propriétés transformationnelles » désigne pour une locution l'impossibilité de lui appliquer les processus de transformations des énoncés simples tels que la passivation, la pronominalisation, etc.

²⁵ Gross précise que l'on peut parler de locution lorsque aucun des éléments lexicaux qui la composent ne peut être actualisé.

²⁶ Respectivement : « Laboratoire d'Analyse Documentaire et Linguistique » de l'Université Paris 7 et « Laboratoire de Linguistique Informatique » de Paris 13.

Paul a cassé sa pipe dans un accident de voiture.

Mais on ne peut pas considérer *casser (sa) pipe* comme un verbe simple. La première raison, commune à celle des noms composés, est la polylexicalité et, par conséquent, la présence de séparateurs, dont le blanc, et les problèmes de traitement informatique induits. Il y a ensuite le fait que la séquence comporte une variable : le possessif est coréférent au sujet et varie de ce fait. Enfin, il y a des insertions possibles dans la séquence, ce qui n'est jamais le cas d'un verbe simple. On peut, en effet, insérer certains adverbes ou incises :

Fais attention, sinon tu vas vraiment casser ta pipe.

Si on identifiait la séquence *casser (sa) pipe* à un verbe simple, alors on générerait incorrectement la négation :

Paul ne meurt pas

**Paul ne casse sa pipe pas*

Paul ne casse pas sa pipe.

Une locution n'est donc pas une catégorie primaire. Elle se sert de catégories primaires qui, dépourvues d'actualisation, se combinent pour former des unités polylexicales qui ont le même statut syntaxique que les catégories simples. (...) Les locutions sont donc des unités intermédiaires entre les catégories simples dont elles ont les fonctions syntaxiques et les syntagmes dont elles ont perdu l'actualisation. »

GROSS (1996 : 22-23)

Observons tout d'abord le problème par la différence entre caractéristiques sémantiques et syntaxiques. En termes de construction syntaxique, une locution verbale n'est pas un verbe, elle est constituée de plusieurs mots, dont un verbe recteur et plusieurs mots d'autres catégories grammaticales (ou parfois contenant un autre verbe subordonné au premier), et elle ne respecte pas strictement les mêmes contraintes syntaxiques que le verbe seul.

Gaston Gross pose le problème de la locution verbale d'un point de vue informatif. Sémantiquement parlant, une locution est opaque dans le sens où son interprétation ne peut être réduite à celle de ses parties. Ainsi « *casser sa pipe* » n'est pas réductible à l'addition d'un des sens de *casser* à une acception de *pipe*. Certes, ce point de vue n'exclut pas que la syntaxe d'une locution verbale présente quelques différences avec les verbes monolexicaux, la négation en étant l'exemple le plus évident, bien qu'il nous semble discutable : la négation est douteuse dans l'exemple invoqué par G. Gross autant que dans d'autres auxquels nous avons pensé :

15)?? *Pierre ne casse pas sa pipe*

16)? *Pierre n'a pas cassé sa pipe*

17)?? *Vous n'allez pas voir de quel bois je me chauffe*

Selon nous, ces exemples sont tellement opaques qu'ils ne paraissent plus autoriser la négation, au risque d'en affecter le contenu sémantique et aspectuel (ce

dernier point est important, la négation semblant mieux marcher au passé dans (16) par rapport à l'exemple (15) que nous trouvons plus que discutable).

Pour résumer, une locution verbale est une expression polylexicale dont la tête syntaxique est un verbe, à forte opacité sémantique et au fonctionnement syntaxique globalement équivalent à celui des verbes monolexicaux. Conséquemment, une locution verbale peut être considérée comme un terme à part entière, auquel nous affecterons une étiquette catégorielle différente des verbes habituels. De cette façon, nous espérons associer à cette catégorie « complexe » les caractéristiques syntaxiques qu'elle ne partage pas avec la catégorie simple. Dans nos bases de données, une locution ou une périphrase²⁷ polylexicale constitue une lexie à partir du moment où elle est proposée comme traduction ou si elle est elle-même traduite dans le dictionnaire. Bien entendu, la sous-représentation de ces expressions dans le dictionnaire pose problème.

En définitive, toutes les expressions polylexicales posent des problèmes de typologie et de description. La forme de nos bases de données devra concilier la grande variété des types de ces expressions tout en simplifiant au maximum leurs conditions d'apparition.

3. Le verbe au centre de l'étude

Après avoir passé en revue les principales notions de sémantique lexicale que nous avons manipulées, à savoir la polysémie, la synonymie et les lexies, nous allons à présent nous intéresser à la catégorie grammaticale qui est au centre de nos préoccupations : le verbe.

Nous allons en premier lieu proposer une définition aussi large et précise que possible de ce qu'est un verbe afin de préciser les limites de l'étude et de justifier notre choix de traiter les verbes plutôt que toute autre catégorie.

À la suite de cela, nous présenterons brièvement un certain nombre de travaux ayant trait aux verbes qui nous ont inspiré peu ou prou pour les analyses décrites par la suite. La richesse des informations et des analyses potentielles à partir des prédicats verbaux sera illustrée par divers aspects dont les caractéristiques aspectuelles ou les structures argumentales induites par le verbe.

3.1 Définition

Si tout francophone est capable de reconnaître un verbe dans un énoncé, définir ce qu'est précisément un verbe n'est pas aussi simple. Touratier (1996) met en évidence les contradictions et les approximations inhérentes aux définitions de la classe grammaticale des verbes proposées dans les grammaires, de Aristote à la fin du vingtième siècle.

²⁷ Une périphrase verbale est un type de locution qui véhicule des informations souvent aspectuelles et qui jouent un rôle de pseudo-auxiliaire d'un autre verbe.

Tout d'abord, les grammairiens ont de tout temps présenté le verbe comme un mot susceptible de varier morphologiquement en fonction de la personne, du temps, du mode, etc. et donc d'appartenir « à un paradigme de conjugaison. » On retrouve déjà cette définition chez le grammairien grec Denys le Thrace, puis chez Lyons (1970), Bonnard (1976) ou Grévisse-Goosse (1993).

Selon Touratier, cette définition est contestable dans la mesure où toutes les langues ne possèdent pas de morphologie du verbe permettant de distinguer celui-ci des autres catégories grammaticales (comme en chinois et en indonésien, par exemple). Certes, le phénomène de conjugaison est commun à toutes les langues indo-européennes mais cela ne suffit pas non plus à en arrêter la définition pour cette famille de langues, car les auxiliaires sont les éléments porteurs des morphèmes de conjugaison dans les constructions à temps composés. Ces auxiliaires, modaux ou non, posent un problème de définition autonome (peuvent-ils être considérés comme des verbes ?) autant que de définition syntaxique du verbe en général (si c'est l'auxiliaire qui prend les morphèmes de conjugaison, que sont les mots qui suivent et que l'on met au participe passé ?).

Pour dépasser cet écueil, certains auteurs proposent une définition sémantique de la classe verbale, à commencer par Aristote et Gustave Guillaume, lesquels présentent la spécificité du verbe en tant qu'élément « sous-tendu de temps » (Guillaume, 1965), « qui implique et explique le temps » (id, 1964) ou qui « ajoute à sa propre signification celle du temps » (Aristote, in Tricot, 1994). Comme l'explique Touratier, cela ne permet pas non plus une caractérisation du verbe puisque d'autres morphèmes et d'autres éléments non verbaux peuvent également impliquer, expliquer et porter le temps dans les langues considérées. Il ne s'agit pas par conséquent d'une caractéristique spécifique de la catégorie verbale.

Puisque le verbe n'est pas seul à véhiculer du temps dans le discours, il faut en chercher la définition sémantique à un autre niveau. La première à être apparue correspond à ce qu'on trouve « ordinairement » dans les grammaires scolaires, comme le précise Touratier :

« (...) mettant en avant la seconde partie de la définition de Denys de Thrace, elles prétendent que le verbe exprime une action faite ou subie par le sujet. À vrai dire, elles sont généralement obligées d'allonger la liste des significations du verbe, en disant par exemple :

« Le verbe est un mot de forme variable, qui exprime une action faite ou subie par le sujet, ou qui indique un état du sujet. » (J. Dubois, G. Jouannon, R. Lagane, *Grammaire française*, 1961, Paris Larousse, p.76, § 148.)

(...) »

(Touratier, 1996 :7)

Meillet, puis Grevisse et bien d'autres, donne à cette spécificité supposée du verbe le terme synthétique de *procès* mais celui-ci n'est pas pour autant propre aux verbes car, comme le souligne Jean Fourquet, un nom ou adjectif peuvent tout autant être désignés par le terme de *procès* de nombreux cas, par exemple pour « *bombardement* [qui] évoque la même représentation que *ils bombardent* » (Fourquet, 1950, p.76). Cette solution sémantique n'est donc pas non plus la bonne pour caractériser le verbe.

Dans « L'aventure des langues en Occident » (1994), Henriette Walter note avec un certain humour que si les Allemands semblent être des gens très polis qui ne se coupent jamais la parole, c'est probablement parce que le verbe est habituellement en fin de phrase dans leur langue. Bien que présentant ce point de vue sous forme de boutade, Henriette Walter met tout de même en avant l'importance sémantique du verbe dans la langue, comme si l'attente d'un élément spécifique dénotant le *procès* était indispensable pour l'interprétation des énoncés.

En définitive, la meilleure définition pour une classe verbale universelle semble être celle qui désigne le verbe comme « étant simplement ce qui a vocation à être le constituant ou le noyau central du constituant de l'énoncé qui remplit la fonction informative de rhème » (Touratier, 1996 :9). L'auteur ajoute que cela en fait le point d'ancrage privilégié des « indices de temps » et des « indices d'aspect » (Fourquet, 1950 :78), ce qui explique qu'il est « affecté de ce qu'on appelle une conjugaison » dans de très nombreuses langues.

À cette définition, nous souhaiterions ajouter que si le verbe est à la fois le rhème (ou le noyau de rhème), le siège privilégié du temps et de l'aspect (mais pas exclusivement) et le porteur d'une marque de conjugaison (au moins dans les langues qui nous préoccupent), alors il faut considérer le prédicat verbal dans son ensemble et pas seulement son noyau lexical. En effet, les auxiliaires participiaux ou modaux et les périphrases aspectuelles sont porteurs, lorsqu'ils sont présents, de certains indices temporels et aspectuels ainsi que des marques de conjugaison en question. Autrement dit, le verbe est l'ensemble de la construction verbale, c'est-à-dire le verbe lui-même plus ses auxiliaires.

Pour aller encore plus loin dans la définition avec l'aide de ce que disait Jean Fourquet au sujet du *procès*, nous nous demanderons si absence de verbe signifie vraiment absence de verbe ? Autrement dit, un énoncé sans verbe est-il forcément dépourvu de *procès* ou de noyau rhématique central ? Citons d'abord le cas de l'ellipse du verbe. Dans certaines situations, comme les réponses communicatives, rien n'oblige à ce que le verbe exprimé dans la première phrase de la communication soit systématiquement répété dans toutes les phrases de tous les locuteurs, par exemple :

18) *'Qu'est-ce qu'on mange ce soir ?*
Du poulet.

Du poulet ?
Oui, du poulet.'
NDSH : '[= (On mange) du poulet]'

Ce que ces occurrences ont de particulier est que le verbe non exprimé ne remplit manifestement plus la fonction informative de rhème, celle-ci étant occupée dans la deuxième phrase par l'objet 'du poulet'. C'est sur la base de cette notion de rhème que la grammaire générative considère que les énoncés construits sur cette forme elliptique ne constituent pas des phrases.

À partir de ces considérations sur « l'absence du verbe », nous constatons que les cas apparents de thématization du verbe sont caractérisés par l'ellipse, la répétition, la paraphrase (ce qui revient presque au même) ou par l'emploi d'un déverbal, que nous allons définir.

Dans un autre ordre d'idées, nous rapprocherons les verbes et les substantifs déverbaux : c'est-à-dire les substantifs comme éléments prédicatifs non verbaux issus de verbes. Les déverbaux sont des cas de nominalisation du prédicat dans des constructions non verbales ou à verbes supports. Contrairement aux formes elliptiques, les déverbaux peuvent parfaitement remplir la fonction de rhème, et surtout celle de procès, et ainsi remplacer le verbe dans son expression sémantique habituelle. L'exemple de *bombardement* proposé par Fourquet que nous citons plus haut illustre bien le phénomène d'appropriation du procès par un élément non verbal. Dans les travaux linguistiques plus approfondis que nous ferons avec *traverser* et les verbes de parole, nous avons mentionné la possibilité pour un prédicat de se fonder sur une occurrence nominale (*traversée, discussion, conversation* et de nombreux autres avec un verbe support *effectuer, faire, avoir, etc.*) sans avoir approfondi la question faite de temps.

3.2 Pourquoi les verbes ?

Dans sa définition même, nous trouvons une justification forte au fait que nous nous intéressons davantage aux verbes qu'aux autres catégories grammaticales. Avant toute chose, le verbe est l'élément pivot essentiel des phrases bien formées, y compris lorsqu'il est absent, comme nous l'avons évoqué au sujet de l'ellipse et des déverbaux. C'est également le seul élément lexical à sens plein qui échappe généralement aux chaînes référentielles et anaphoriques²⁸ ainsi qu'à la thématization.

²⁸ Hormis dans certains cas d'anaphores associatives à déclencheur verbal (« il coupa du pain et rangea le couteau ») ou à verbe vicair (« on n'agit point comme vous le faites »), selon le terme proposé par Riegel, Pellat & Rioul (1996). Pour les premières, le référent est à chercher dans la structure argumentale du verbe, couteau étant l'instrument habituel de l'action de couper.

De même, il ne s'agit pas d'un déictique pur car il ne possède pas de référent dans la situation d'énonciation, contrairement à ses arguments.

Ensuite, et surtout, l'analyse du verbe ne fait pas l'économie de l'étude de tous les éléments qui l'entourent. Nous allons largement insister sur le rôle du co-texte immédiat du verbe, et notamment sur l'importance de ses arguments, dans le processus de désambiguïsation de celui-ci. En étudiant de près un verbe comme *traverser*, nous nous apercevons également que la saturation de tous les traits sémantiques qui lui sont liés est tributaire d'un co-texte large qui dépasse souvent le cadre de la prédication.

C'est en raison du rayonnement tout particulier des éléments de la catégorie verbale, et aussi en raison de sa forte dépendance envers ses arguments et le co-texte, que nous avons entrepris d'en faire le matériau de base de toutes nos recherches doctorales.

3.3 La classification des verbes en linguistique

Sous ce titre, nous n'allons pas exposer, loin s'en faut, la totalité des travaux effectués sur les verbes en linguistique car le nombre et la variété de ces études dépassent de très loin ce que nous sommes capable de connaître. Dans les paragraphes suivants, il ne sera pas question d'études morphologiques, large domaine que nous n'avons pas exploré, ni de travaux syntaxiques ou sémantiques portant sur des verbes précis, hormis dans les chapitres II et III où nous ferons référence aux recherches et publications dont nous avons eu connaissance au sujet des verbes que nous avons étudié, *traverser* et les verbes de parole en particulier.

Tout d'abord, nous pouvons citer deux classifications majeures, celles de Maurice Gross pour les verbes français (et les tableaux pour la traduction qui s'en sont ensuivis) et celle de Beth Levin pour un grand nombre de verbes anglais répartis en de nombreuses catégories :

- Les tables de Maurice GROSS. Le travail de classification des verbes effectué par Maurice GROSS est probablement à ce jour le plus riche et le plus complet en ce qui concerne le français. D'abord dévolue à la description de toutes les formes syntaxico-sémantiques possibles²⁹

²⁹ Des travaux déjà largement décrits dans son ouvrage de 1975, *Méthodes en syntaxe. Le régime des complétives*, dans lequel Maurice Gross précise les constructions de nombreux verbes avant d'en proposer une classification en fonction de critères syntaxiques. Cette somme de données considérable servira ensuite de socle à ses recherches en traduction automatique fondées essentiellement sur des dictionnaires et des bases de données très fournies. Dans une moindre mesure, nos propres travaux s'apparentent aux siens en cela qu'ils se fondent sur une base de données détaillée et directement exploitable.

(constructions, types de complément, etc.), l'œuvre de Gross s'est orientée vers la traduction automatique.

- Les catégories de Beth LEVIN. Beth Levin a recensé 3000 verbes parmi la totalité des verbes anglais en les classant et les regroupant en fonction de leurs caractéristiques syntaxiques et de leurs propriétés sémantiques. Chaque verbe peut appartenir à plusieurs classes en même temps, en fonction de ses acceptions ou de ses constructions. Ce travail considérable a deux intérêts immédiats : chaque verbe parmi les verbes anglais les plus fréquents est comparé aux autres verbes qui partagent avec lui les mêmes caractéristiques argumentales et sémantiques ; d'autre part, Beth Levin propose une bibliographie très riche répartie entre chacune des classes qu'elle présente. Toutefois, nous regrettons dans ses travaux l'absence des verbes à particule fusionnée ou détachée, Levin réduisant la classe des verbes étudiés à l'ensemble des unités monolexicales. Ce parti pris limite énormément la vision d'un lexique anglais évolutif dans lequel les prépositions jouent un rôle important et subtil dans la création de sens nouveaux ou dans la nuance des sens du verbe d'origine.

Au-delà des considérations lexicales et sémantiques portant sur le lexème verbal, un courant majeur de la linguistique a mis en exergue la capacité des verbes à exprimer un « sens aspectuel », c'est-à-dire une ou des temporalité(s) spécifique(s) non explicite(s) au moyen des désinences de conjugaison. Ces dernières sont habituellement perçues comme « porteuses du temps de la phrase » mais ne rendent pas nécessairement compte du temps réel ou profond, lequel est connu sous le nom *d'aspect*. A ce sujet, nous notons que l'aspect dans un énoncé peut résulter de la combinaison de plusieurs facteurs propres ou externes au verbe, et que le sens aspectuel ne se rapporte pas seulement au temps *stricto sensu*, c'est-à-dire à la distinction formelle entre passé, présent et futur.

Nous présentons des considérations générales sur l'aspect au début du chapitre III, lequel sera presque entièrement consacré à la reconnaissance des propriétés aspectuelles des occurrences contenant le verbe *traverser* ainsi qu'au rapport existant entre ces propriétés et le sémantisme du verbe. C'est pourquoi nous n'entrerons pas ici dans le détail des travaux effectués sur cette notion.

3.4 La polysémie verbale

La polysémie verbale est une polysémie à géométrie variable, et ce pour plusieurs raisons.

Tout d'abord, il s'agit de bien montrer la différence entre les emplois premiers du verbe et ses emplois métaphoriques, lesquels découlent des premiers. Selon le

principe de lexicalisation, les sens d'abord métaphoriques d'un verbe ont tendance à perdre leur caractère figuré pour devenir peu à peu des sens propres susceptibles d'être métaphorisés à leur tour. En ce sens, les verbes sont plus que la plupart des unités des autres classes grammaticales soumises à la polysémie en tant que processus de création lexicale et pourtant – ou précisément à cause de cela – elle a été très peu étudiée par rapport aux polysémies nominale, adjectivale voire prépositionnelle.

Ensuite, on ne peut pas parler de polysémie verbale au sens large sans mentionner le rôle des collocations dans la caractérisation de la polysémie. Nos analyses contrastives montrent que le sens de certains verbes est avant tout une question de reconnaissance de la structure argumentale, et qu'une seule acception est susceptible de se subdiviser en acceptions plus fines en fonction du sujet et des compléments qui l'accompagnent. Même si dans le fond le sens est le même, on constate des nuances par le jeu du transfert interlangue. Il est en ainsi pour *abandonner* comme nous l'avons déjà relevé au cours de ce chapitre.

Pour la visibilité de la polysémie verbale au moyen des graphes, notons à nouveau les travaux effectués par Bruno Gaume, Karine Duvignau, Sabine Ploux, Jacques François et quelques autres. Pour l'essentiel, ces chercheurs réussissent à caractériser la polysémie des verbes en se fondant sur les calculs automatiques de la synonymie et/ou de la distance lexicographique. Les résultats qu'ils observent en se fondant sur l'automatisation des données sont généralement proches de la réalité du lexique telle que nous le percevons tous. Mais dans tous les cas, la démarche est avant tout mathématique et stochastique.

3.5 Transition

Après avoir posé les notions et concepts dont ces travaux sont tributaires, nous allons à présent décrire succinctement quelques recherches majeures dans le domaine de la sémantique lexicale, et plus particulièrement les travaux qui ont trait à la description et à la modélisation du lexique et de la polysémie.

Ceux qui nous ont le plus inspiré et dont nous allons parler maintenant sont les modèles de traduction selon Jacques François (par la prise en compte de critères situationnels précis), les graphes conceptuels de John Sowa, les dictionnaires électroniques des synonymes du CRISCO et de l'ISC Lyon, le Dictionnaire Intégral de Dominique Dutoit, le Lexique Génératif de James Pustejovsky, le DEC de Mel'çuk et ses collaborateurs sans oublier les classes objets développées par l'équipe de Gaston Gross.

4. Modèles théoriques

4.1 Modèles de traduction selon J. François

Largement présentés dans « Changement, Causation, Action », les modèles de traduction de Jacques François (tirés de son « multilexique contrastif ») illustrent une représentation du sens à partir de la description langagière la plus fine possible pour le passage du français vers l'allemand. Il adopte une démarche onomasiologique pour la structuration de champs sémantiques. À première vue, notre propre démarche est purement sémasiologique car nous n'étudions d'emblée que la polysémie par allers et retours successifs à l'intérieur des dictionnaires. Toutefois, il est manifeste que nous avons voulu nous rapprocher d'une démarche onomasiologique par la prise en compte des liens de synonymie généralisés d'une part et par la constitution de graphes multilingues d'autre part. De ce point de vue, les travaux de Jacques François ont une forte résonance au travers des nôtres. La banque d'équivalences à laquelle il participait dans le cadre d'EUOTRA dans les années 80 est également semblable à celle que nous avons produite pour notre recherche.

Pour en revenir au multilexique, une prise en compte aussi fine et absolue des variations contextuelles n'est que rarement conforme à l'usage, dans le cadre du discours rapporté. Sans la connaissance extrinsèque de la situation d'énonciation, il n'est sûrement pas possible de rassembler ces informations par l'étude du co-texte seul. Il s'agit donc avant tout de modèles descriptifs, sous la forme d'un formalisme très précis.

Jacques François décrit le langage à partir des critères situationnels les plus fins. C'est le langage étendu à la connaissance du monde énonciatif dont voici un exemple avec le verbe *étouffer* et l'un de ses équivalents en allemand :

ÉTOUFFER, 1
étouff- [N:x2^{anim} N:x1^{anim}]
 caus/agent ___v patient

avec V a: +Dyn, +Agent, +Trans (+Cause), +Tel,
 (-CentreTel), -Period /2/
 b: RespLibre à Déb-T (TYPIQUE) , Resp0 à Fin-T /E/
 c: méthode : x2 empêche les voies respiratoires de x1 de rester en
 contact avec un mélange gazeux propre à la respiration /3/
François (1989:470)

L'avantage de la méthode est qu'elle met en valeur l'importance des variations contextuelles et aspectuelles pour la traduction. Chaque lexie est codée par Jacques François avec toutes les valeurs aspectuelles et sémantiques qu'il juge utile pour sélectionner le juste correspondant en langue cible, en l'occurrence le verbe allemand *ersticken* dans le cas que nous citons, lequel verbe a également été encodé dans sa base de la même manière que le verbe français *étouffer*.

Dans la mesure du possible, nous avons le même degré d'exigence dans la lecture des dictionnaires dans le but d'extraire les données qui enrichissent notre base. Cette exigence est primordiale dans l'optique de rassembler toutes les informations nécessaires pour le transfert et, au-delà, pour la discrimination de tous les candidats d'un même terme. Bien sûr, le travail de Jacques François va beaucoup plus loin par la modélisation de toutes les données aspectuelles, ce qu'un dictionnaire d'usage courant ne peut pas faire de manière explicite. Mais la connaissance fine de l'aspect revêt une telle importance que nous en ferons une étude complète pour un verbe qui en joue énormément : *traverser*, au chapitre III.

4.2 Graphes lexicaux, graphes conceptuels

4.2.1 La genèse et la création des formalisations du langage et des graphes conceptuels : d'Aristote à John Sowa

Les relations entre concepts et les ontologies ont retenu l'attention de nombreux philosophes et linguistes depuis l'Antiquité. Le premier à avoir élaboré des semblants de réseaux conceptuels fut Aristote. Beaucoup plus tard, au dix-septième siècle, Leibniz imagina un système de combinaison mathématique qui donnait à un assemblage de concepts un nouvel indice par le produit des indices des deux précédents³⁰, le considérant comme un nouveau concept. Plus près de nous, au vingtième siècle, un certain nombre de scientifiques, au demeurant plus souvent mathématiciens que linguistes, ont élaboré des formalismes susceptibles de rendre compte par la logique formelle la complexité du langage.

John Sowa fut un des pionniers de la représentation du lexique et du langage sous la forme de graphes dits conceptuels. Mathématicien d'origine, son postulat était que la langue était formalisable par le biais de la théorie des graphes et que les liens conceptuels n'étaient ni plus ni moins que des liens entre les sommets d'un graphe.

Dans son article de 1992, « Logical Structures in the Lexicon », John Sowa propose une lecture critique de certains formalismes qui ont précédé le sien, à commencer par celui de Leibniz puis en poursuivant avec une réflexion sur le « feature-based system » de Katz et Fodor (1963) :

« In their feature-based system, Katz and Fodor (1963) factored the meaning of a word into a string of features and an undigested lump called a *distinguisher*. Following is their representation for one sense of bachelor:
bachelor → *noun* → (*Animal*) → (*Male*) → (*Young*)
→ [*fur seal when without a mate during the breeding time*].

³⁰ «If rational were represented by 2 and animal by 3, then their product 6 would represent rational animal or human. (...)», (Sowa, 1992 :39). Sowa précise que Leibniz essaya d'appliquer son système à la mécanisation des syllogismes d'Aristote, mais avec un succès mitigé.

(...)”
Sowa (1992:40)

Dans cet exemple, *bachelor* est pourvu d'un marqueur syntaxique (*noun*) et de divers traits sémantiques (*Animal*, *Male* et *Young*) qui sont des classes sémantiques dont la combinaison est censée caractériser l'entrée principale. La phrase entre crochets, le fameux *distinguisher*³¹, est la glose destinée à rendre compte du sens véritable du terme étudié.

Malgré les nombreuses critiques essuyés par Katz et Fodor, notamment à propos de leur refus de voir derrière les traits sémantiques l'existence de primitives ou de structures profondes qui caractériseraient beaucoup mieux la pertinence de leurs modèles, Sowa tempère quelque peu la fronde en indiquant que les traits ne sont pas pour autant inutiles dans la représentation du sens mais qu'ils ne peuvent exclure une réalité du sens plus profonde :

« (...) These criticisms do not imply that features are useless. But they indicate that features are derived from some deeper, more fundamental representation.”
Sowa (1992:40)

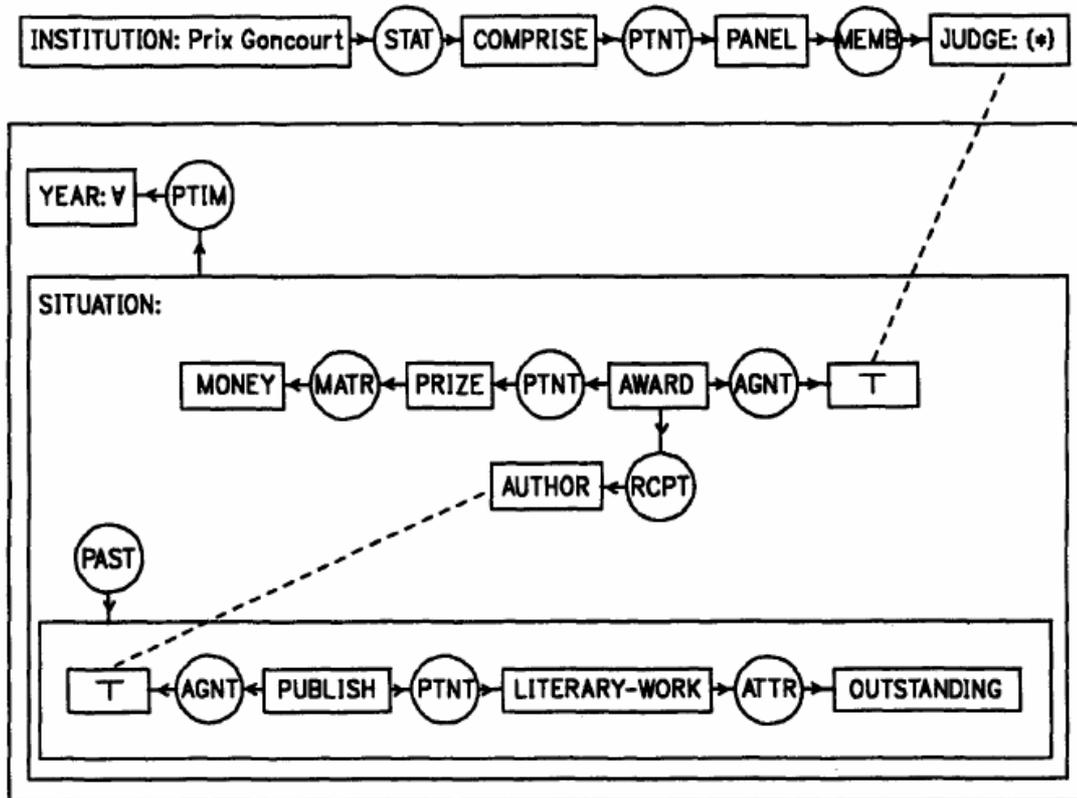
Sowa relève également certaines insuffisances dans les modèles proposés par Montague et Schank. Selon lui, leur principale faiblesse est de ne pas profiter des forces des autres : Montague a élaboré un formalisme efficace mais particulièrement rigide tandis que celui de Schank est attractif en raison de sa plus grande flexibilité qui le rapproche plus de la réalité de la langue, ce qui constitue également son plus grand défaut puisque les modèles qu'il préconise sont instables et peu réutilisables d'une personne à l'autre, comme Sowa le souligne à propos des étudiants de Schank. Le premier est plus efficace pour une analyse exhaustive d'un point linguistique précis et ponctuel tandis que le second peut embrasser plus largement mais de manière moins nette des emplois variés d'une même expression.

En définitive, John Sowa entend essayer de garder le meilleur des deux, sans omettre la théorie des traits initiée par Katz et Fodor, afin de concevoir un formalisme qui allie rigueur et flexibilité, reconnaissance des traits et prise en compte des représentations profondes comme les liens conceptuels, les cadres temporels et d'autres types de relations sémantiques qui participent du sens lexical. Le background de ses graphes conceptuels est bien meilleur car c'est de la rigueur de ses modèles que peut jaillir la malléabilité du langage, comme en témoigne sa représentation du *Prix*

³¹ Sowa semble ne pas beaucoup apprécier le *distinguisher* qu'il qualifie d'« indigested lump », littéralement « (gros) morceau non digéré » avant de préciser que Katz et Fodor ne l'analysent pas. La question sous-jacente à son sujet est donc « à quoi sert-il ? »

Goncourt (1992 :50) qui met en avant le mécanisme de métonymie comme processus de création polysémique :

Figure 1 : Graphe de « Prix Goncourt » par John Sowa (1992)



Dans ce graphe, Sowa fait co-habiter des traits avec des relations sémantiques qui les relient, des cadres temporels qui permettent une lecture récursive de l'ensemble ou de certaines parties du graphe : Certaines opérations représentées dans le graphe peuvent se répéter à certains moments du temps, l'œuvre qui gagnera le prix est antérieure au moment où le prix est décerné (cadre PAST), le prix est décerné chaque année (cadre principal) et l'institution se situe en dehors de ces deux sous-cadres car il possède une dimension « intemporelle ».

En l'illustrant par les processus métonymiques qui autorisent la polysémie sous-jacente de *Prix Goncourt*, l'objectif avoué de John Sowa est de parvenir à une représentation exhaustive des relations et des sens lexicaux au sein d'un graphe unique. Son ambition assumée est que tout ce qui participe à la construction du sens doit être formalisable et qu'il parviendra lui-même à la formaliser au sein de ses graphes conceptuels.

Même si l'entreprise est louable et ses travaux fort convaincants, il est fort probable qu'il n'y parviendra jamais.

4.2.2 Les dictionnaires de synonymes modélisés

4.2.2.1 Les réalisations de l'ELSAP et du CRISCO³²

Dans leur cahier du CRISCO n°17 de juillet 2004, les membres de l'équipe présentent leur dictionnaire électronique des synonymes :

« le dictionnaire des synonymes du CRISCO (<http://www.crisco.unicaen.fr/>) est d'un genre particulier ; c'est en effet l'un des deux dictionnaires des synonymes du français en accès libre sur Internet³³ ; il contient 49000 entrées, chacune suivie de ses synonymes, comme par exemple :

démarcation, distinction, délimitation, frontière, ligne, limitation, limite, lisière, marque, séparation

Chaque synonyme est lui-même une entrée, de sorte que l'on peut considérer notre dictionnaire comme un ensemble de 49000 mots-vedettes (entrées) reliés les uns aux autres par un réseau de relations. Il y a relation lorsque les deux vedettes sont synonymes, et l'on dénombre un peu plus de 199000 relations pour tout le dictionnaire.

Ces relations de synonymie proviennent de sept dictionnaires classiques : deux dictionnaires analogiques (le Grand Larousse et le Grand Robert), deux dictionnaires des synonymes du 19^{ème} siècle (Lafaye et Guizot), et trois dictionnaires du milieu du 20^{ème} siècle (Bailly, Bénac et Du Chazaud). Toutes les informations contenues dans ces dictionnaires nous ont été fournies sous forme de fichiers informatiques par l'INALF³⁴ ; le travail effectué au laboratoire a consisté à harmoniser les mots-vedettes de ces sept fichiers, puis à fusionner toutes ces données. (...) »

Manguin, François et alii (2004 :1-2)

De la même façon que Jean-Luc Manguin, Jacques François et leurs collaborateurs, notre propre mode d'extraction et de représentation des données s'appuie sur une fusion des dictionnaires ou des différentes parties d'un même

³² CRISCO = Centre de Recherches Inter-langues sur la Signification en COntexte, UMR de l'Université de Caen et du CNRS. L'ELSAP est l'ancienne équipe de recherche en linguistique de l'université de Caen, remplacée depuis par le CRISCO. C'est dans le cadre d'une collaboration entre l'ELSAP et l'INALF que Sabine Ploux, Bernard Victorri et Catherine Fuchs ont initié les travaux d'informatisation des dictionnaires de synonymes.

³³ Les auteurs précisent que l'autre dictionnaire des synonymes disponible est celui de l'ISC que nous décrivons dans le paragraphe suivant.

³⁴ À ce titre, les auteurs précisent qu'outre les données des sept dictionnaires mentionnés, leur regroupement de liens synonymiques contient « un ensemble important de corrections et d'ajouts » (CRISCO, 2004 :2).

dictionnaire lorsqu'il s'agit d'ouvrages bilingues. Certes, il en résulte en apparence une perte d'informations :

« la fusion des dictionnaires a fait disparaître les commentaires (quand ils existaient) et la structure des ouvrages d'origine ; cet appauvrissement apparent, nécessaire en raison des disparités entre les différents dictionnaires de départ, constitue en fait un tremplin pour des méthodes automatiques qui, comme nous le verrons, permettent d'analyser de manière fine et objective l'espace sémantique des mots. »
ibid. (p.2)

Notre méthodologie personnelle et nos conclusions sont donc très proches de celles des membres de cette équipe. Toutefois, nous avons toujours essayé, dans la mesure du possible, de conserver les commentaires qui accompagnent éventuellement un parcours de traduction dans les dictionnaires bilingues, en partant du principe qu'un commentaire est souvent un critère discriminant indispensable ou à tout le moins l'explicitation d'un critère discriminant nécessaire.

4.2.2.2 Autre volet du projet à l'ISC Lyon³⁵ : les graphes multilingues

Les dictionnaires de synonymes modélisés auxquels nous nous référons sont les champs lexicaux élaborés à partir des fichiers informatiques des synonymes réalisés à l'INALF et mis en graphiques à l'ELSAP et au CRISCO puis bilingués par Sabine Ploux et Hyungsuk Ji, lesquels sont en train de développer leurs modèles en espagnol en attendant d'y intégrer d'autres langues.

Notre vision de la co-représentation multilingue est un peu dans le même esprit. Tout ici repose sur le principe de proximité sémantique par synonymie (calcul à partir de la synonymie issue de données lexicographiques croisées). Le principe est à première vue le même que pour Wordnet Multilingue mais l'opacité de ce dernier ne permet pas d'en être certain. Mais surtout, les relations de synonymie au sein des dictionnaires subit un « traitement de symétrisation » :

« Dans les dictionnaires initiaux, la relation de synonymie n'est pas toujours attestée symétrique : un terme peut être synonyme d'une vedette sans que la réciproque soit vraie. Pour symétriser cette relation, le programme cherche, pour une vedette donnée, tous les articles contenant cette entrée en partie gauche et ajoute les vedettes de ces articles à la liste des synonymes, quand elles n'y figurent pas déjà. »
Ploux (1997:5)

C'est à peu de choses près la même méthodologie que nous avons utilisée pour la création des Lexiques Inversés à partir des dictionnaires bilingues. Dans le chapitre

³⁵ ISC Lyon = Institut des Sciences Cognitives de Lyon.

IV qui traite de l'algorithme de transformation des dictionnaires, nous discuterons de la question de la symétrisation dans une perspective beaucoup plus critique et dans le souci d'adapter nos outils aux langues et non pas le contraire. En ce qui concerne la synonymie intra-langue, nous admettons plus facilement le rétablissement d'une symétrie qui paraît être à la base de la définition de la synonymie.

Le problème des « patates » (nom usuel des graphiques), c'est qu'il n'existe pas d'espace autonome pour les unités lexicales étudiées. Chaque espace du graphe est occupé par le mot sélectionné, un ou plusieurs synonymes (généralement plusieurs) ainsi qu'une ou plusieurs traductions (généralement plusieurs) en anglais ou en français selon la langue d'origine. A chaque portion du graphe à informations constantes correspond une « clique », et inversement. Une clique est à l'origine un ensemble de synonymes qui sont réputés interchangeable dans un contexte donné (cf. définition de la synonymie). Elles sont calculées automatiquement par un algorithme de détection de la distance lexicographique à partir de la fusion des dictionnaires de synonymes fusionnés que nous avons décrite précédemment.

De même, la prolifération des unités lexicales dans une même clique ne permet pas de déterminer quelles sont les relations privilégiées entre elles deux par deux. En admettant que ces dernières puissent être calculées automatiquement (avec des outils semblables à ceux qui ont permis de construire les graphes), rien ne les valide d'emblée.

Il est à noter que les graphes bilingues réalisés par Ploux et Ji (Ploux & Ji, 2003) sont le résultat de la mise en relation de deux dictionnaires monolingues de synonymes, un français et un anglais, et d'un dictionnaire bilingue français-anglais. L'automatisation de cette fusion sous la forme d'un graphe appelé « atlas sémantique » conduit au même mode de représentation que le graphe monolingue d'origine. La différence majeure et l'apport considérable par rapport à celui-ci sont qu'à chaque espace du graphe sont associées toutes les traductions potentielles de l'item considéré dans des emplois bien délimités. Les mêmes auteurs vont encore plus loin en incluant dans leurs modèles ce qu'ils appellent des « contextonymes », c'est-à-dire des termes qui sont sémantiquement corrélés aux items visés. Ces contextonymes sont intégrés à un espace sémantique multidimensionnel. Selon eux, la prise en compte de ces termes aide à la compréhension des textes ainsi qu'au repérage des connotations et des sens contextuels pour des unités polysémiques (Ji, Ploux & Wehrli, 2003). C'est dans un même esprit que nous avons souhaité étudier certains verbes dans leur environnement langagier, notamment *traverser*, pour montrer ce que les dictionnaires ne peuvent pas révéler de manière exhaustive quant au sens des verbes.

4.2.3 Le Dictionnaire Intégral (DI©, MEMODATA©)³⁶ de Dominique Dutoit ou le Petit Robert en réseau

Le DI élaboré par Dominique Dutoit à Caen, au sein de l'entreprise MEMODATA, est sans doute l'un des travaux lexicologiques informatisés les plus aboutis en terme de volume et du nombre de liens implantés. Nous allons lui consacrer un sous-chapitre conséquent, en raison du fait qu'il s'agit probablement du travail le plus proche de nos aspirations et de nos résultats tangibles. D'autre part, nous avons eu l'opportunité d'avoir accès aussi bien à l'outil qu'à son mode de fonctionnement interne.

Dans la perspective d'effectuer un travail d'ensemble sur les lexiques multilingues en évitant l'écueil de la terminologie, il nous paraît intéressant voire indispensable de se pencher sur les potentialités des réseaux sémantiques. En partant du principe que la traduction automatique impose la prise en compte des travaux texte→sens et sens→texte, l'utilisation d'un réseau sémantique conçu pour permettre la compréhension aussi bien que la génération nous semble judicieuse, à la condition de disposer d'un outil construit déjà implanté. Dans les paragraphes suivants, nous présenterons les caractéristiques du Dictionnaire Intégral élaboré par Dominique Dutoit à Caen ainsi que la façon dont nous aurions souhaité l'exploiter.

Nous présenterons séparément le Dictionnaire Intégral, l'algorithme qui exploite les données de ce dernier ainsi que les outils activés par cet algorithme. Pour chacune des trois parties, nous proposons un diagnostic sommaire et quelques remarques « critiques » pour étayer notre point de vue. En guise de conclusion, nous donnerons notre vision personnelle au sujet de l'intérêt que représente l'apport de ce type de réseaux sémantiques dans le domaine de l'industrie de la langue. Nous tenterons également d'exposer notre vision de l'outil pour son exploitation en traduction automatique ou semi-automatique (TAO³⁷). Ce dernier point sert de passerelle à l'explicitation de notre propre exploitation des données dans une perspective de TAO.

4.2.3.1 Structure générale

Le DI possède une structure pyramidale dont le sommet non lexical est appelé « dictionnaire ontologique » duquel partent un certain nombre de liens vers des catégories ontologiques prédéfinies (*être* et *non-être* à la racine, etc.). La base de la pyramide est constituée des unités lexicales saturées, à savoir les mots ou locutions pour lesquelles suffisamment de traits sémantiques et syntaxiques ont été implantés en amont pour les désambiguïser. A partir de cette description, on peut concevoir le DI

³⁶ Le Dictionnaire Intégral est une marque déposée par l'entreprise Mémodata dont Dominique Dutoit, par ailleurs Directeur de Recherche associé au CNRS et au CRISCO, est le PDG.

³⁷ TAO = Traduction Assistée par Ordinateur

comme un hybride de réseau sémantique et d'arbre³⁸ puisque à part la racine (le *dictionnaire ontologique*) et les unités saturées, chaque terme ou concept possède un ou plusieurs lien(s) vers des archilexèmes de type générique ou conceptuellement englobants, un ou plusieurs lien(s) vers des lexèmes qu'il englobe et des liens de parenté lexicale (synonymie, para-synonymie, appartenance de classe) avec les autres éléments présents dans la même case³⁹.

Globalement, le Dictionnaire Intégral se présente sous la forme d'un réseau multidimensionnel de 250 000 mots, recouvrant de manière quasi exhaustive le lexique du Français. On doit comprendre par « multidimensionnel » qu'il existe plusieurs phases lexicales dans le DI, que chaque lexème instancié possède des liens multiples et de différentes natures avec les autres et qu'il est possible de paramétrer ces liens sans limite *a priori* hormis les contraintes liées à l'informatisation et au stockage en mémoire machine.

4.2.3.2 Les trois dimensions du DI

La première dimension est constituée d'un dictionnaire classique de la langue épuré de presque tous ses éléments non discriminants ou définitoires. Chaque lexème est dégroupé selon les définitions disponibles dans le dictionnaire d'origine (en l'occurrence le Petit Robert) ; les citations, renvois et contraires, exemples et parties lexicales des définitions qui sont jugées non utiles ne sont pas conservés. En revanche, il est possible de préciser le type sémantique et syntaxique des arguments éventuellement sous-catégorisés par l'unité lexicale considérée, sur la base des syntagmes exemples du Robert ou en fonction de l'intuition de l'implanteur (de la même façon que les concepteurs du DEC).

Ainsi, l'entrée du verbe *aborder* se présente grosso modo comme suit dans le DI (impression visuelle fidèle à la page du vrai DI) :

Aborder 1⁴⁰
*/se mettre bord à bord avec*⁴¹ [*arg1*⁴²=*navire*⁴³]
/éperonner [*arg1*=*navire*]
/heurter (accidentellement) [*arg1*=*vaisseau*]

Aborder2

³⁸ C'est-à-dire un treillis ou une classification à héritage multiple

³⁹ *Case* signifie ici toute portion du graphe visualisable à l'écran sous la forme de listes de termes corrélés.

⁴⁰ Entrée lexicale deuxième niveau.

⁴¹ Définition épurée d'une lexie.

⁴² Une des fonctions syntaxiques : argument objet 1.

⁴³ Type ou mot, ou les deux, pour l'argument objet 1.

/arriver au rivage⁴⁴, arriver sur le bord [arg1(dans NP)⁴⁵=lieu en bord de mer ou de plan d'eau]
 /atteindre, toucher [arg1=rivage]
 /arriver à [arg1=lieu (inconnu)⁴⁶]
 /s'approcher de, aller à [arg1=qqn (inconnu)]
 /en venir à [arg1=thème abstrait⁴⁷]
 /arriver au bord de [arg1=situation]
 /commencer à s'occuper de [arg1=qqchose]
 /s'engager dans [suj=véhicule ou sujmetony⁴⁸=conducteur, passagers ; arg1=partie de voie]

Le signe ‘/’ désigne un lexème qui possède un lien vers un autre lexème plus spécifique ou vers une unité lexicale saturée. C’est en pratique le signal qu’il est encore possible de naviguer dans le graphe vers la base de la pyramide à partir de la lexie précédée du signe.

A chaque ligne du DI est associée une fonction grammaticale et une seule, une fonction lexicale et une seule, un code d’identification, une marque catégorielle. Les fonctions seront rapidement explicitées ultérieurement dans ce document, mais sans entrer dans les détails car il est impossible d’en dresser la liste.

Lorsqu’on aborde le DI au moyen de la recherche d’un mot ou d’un lexème, la page obtenue se présente sous la forme d’une liste de définitions tel que présenté plus haut. Il est à noter que cette liste est ordonnée arbitrairement en fonction de la fréquence d’emploi supposée de chacune des définitions. Il s’agit alors systématiquement d’un parti pris du concepteur de l’outil car la construction de ce dernier se fait en dehors de toute utilisation de corpus ; Dominique Dutoit part du principe qu’il n’y a pas de corpus suffisamment étendu ou représentatif pour être pertinent dans sa démarche, à moins de constituer un corpus ontologique universaliste lequel, sauf s’il pût exister à l’état naturel, est inenvisageable car il serait immanquablement soumis à l’arbitraire humain. Cette position est d’ailleurs paradoxale puisque l’intégralité des travaux de Dominique Dutoit est soumise à son propre arbitraire.

Dans une seconde phase lexicale, sont pris en compte les renvois comme synonymes potentiels (*accoster* pour *aborder1.1*, *aborder2.4* ; *draguer* pour *aborder*

⁴⁴ Double définition, induction d’une relation synonymique entre les deux définitions.

⁴⁵ Dans la fonction syntaxique [argument objet 1], c’est l’intégration d’une contrainte sur la structure syntaxique de l’objet (« dans + [Arg1] »).

⁴⁶ Donnée de la situation d’énonciation, servant à préciser l’objet. Peut faire l’objet d’une fonction lexicale spécifique.

⁴⁷ Type d’argument générique sur un plan ontologique (qui se réfère au contenu du DI lui-même). Fonction lexicale également.

⁴⁸ Précision sur le type de sujet (en l’occurrence métonymique vs. non métonymique), fonction lexicale qui détermine ce type par superposition avec la fonction syntaxique *Suj.*

2.4, *entamer pour aborder, etc.*), et éventuellement les contraires comme antonymes, ces derniers étant assez rares dans le Petit Robert. On y retrouve également un grand nombre de collocations (comme arguments sujet ou objet) ainsi que certains termes associés avec les fonctions appropriées. Il faut à ce stade bien marquer l'opposition entre définitions et renvois : les définitions se veulent les meilleurs équivalents du terme à expliciter ; les renvois sont des synonymes contextuels, eux-mêmes soumis au même phénomène de définition et renvois. Pour éviter la circularité des définitions (un mot défini par un autre, lui-même défini par le premier), celles-ci sont essentiellement périphrastiques.

En dehors de ces deux dimensions 'originelles', le DI est un réseau sémantique fermé, i.e. il est toujours possible de trouver un chemin entre deux unités lexicales quelles qu'elles soient sans être obligé de « sauter » une case, c'est-à-dire d'enjamber un lien qui n'existe pas. C'est grâce à cette propriété essentielle du réseau fermé qu'il est possible de systématiser le calcul de la distance sémantique entre deux termes.

4.2.3.3 Fonctionnement apparent et inapparent des liens au sein du réseau

A l'écran, chaque page visible du DI regroupe des unités hiérarchiquement équivalentes d'un point de vue sémantique. Ainsi, les mots *moineau*, *autruche*, *mésange*, *vautour* peuvent apparaître sur une même page parce qu'ils sont tous les éléments d'une même classe (celle des oiseaux). De même, *autruche* pourra apparaître sur d'autres pages comme celles portant l'étiquette « oiseau qui ne peut voler » ou « oiseau élevé pour sa chair » voire également « animal » ou « être vivant. » En clair, chaque unité lexicale peut être appelée à différents niveaux du réseau sémantique en raison de son appartenance théorique à plusieurs classes en même temps. Le principe de base du réseau est la recherche de l'exhaustivité : par exemple, chaque unité doit être directement rattachée à tous ses hyperonymes ainsi qu'à tous ses hyponymes identifiés. Ce qui est intéressant dans ce principe est qu'une unité lexicale n'est implantée qu'une seule fois dans le graphe, alors que ses liens sont multiples et l'autorisent à apparaître dans des pages différentes. Il y a donc ici un décalage entre l'implantation (un terme = un lieu informatique pour le programmeur) et la consultation par un utilisateur quelconque pour lequel chaque terme a le don d'ubiquité.

Dans l'esprit du DI, la notion d' « unités hiérarchiquement équivalentes d'un point de vue sémantique » n'est pas limitative à partir du moment où les fonctions évoquées plus haut sont correctement définies et différenciées. Ainsi, outre les autres oiseaux, *autruche* pourra être associée à d'autres paradigmes lexicaux en vertu de la fonction lexicale ou syntaxique appropriée⁴⁹ :

⁴⁹ En l'occurrence, *oiseaux qui ne volent pas*, *thèmes associés*, *localisation* sont des fonctions lexicales et *phraséologie* une fonction syntaxique.

- Autruche --> émeu, nandou, casoar, courvite (oiseaux qui ne volent pas)
- Autruche --> élevage, course d'animaux (thèmes associés)
- Autruche --> Afrique, savane (localisation)
- Autruche --> se cacher, faire le mort (phraséologie : « faire l'autruche »), etc.

Les liens établis peuvent être de nature hyperonymique dit lien père-fils (rapport de la classe à l'objet : « oiseau » est hyperonyme de « autruche »), hyponymique ou fils-père (rapport de l'objet à la classe), synonymique ou frère (quasi-équivalence de sens), parasyonymique (rapport de sens évident mais pas équivalent), morphologique ou intrinsèque (paradigmes de conjugaison, forme plurielle, féminin) et la liste n'est pas close. En réalité, tout dépend de la façon dont sont instanciées les fonctions lexicales et grammaticales.

4.2.3.4 Enrichissement et exploitation du réseau

Le DI est régi par un programme qui permet de naviguer d'un mot à l'autre en fonction des liens activés. La plate-forme informatique permet en outre d'introduire et de modifier à loisir un très grand nombre de traits sémantiques, morphologiques et autres. Il est également très aisé (en théorie) d'ajouter des traits nouveaux selon les besoins. La principale faiblesse du système est de ne pas autoriser plus d'une fonction d'un type donné par terme ou lexème (sans doute par souci d'éviter le risque d'explosion combinatoire). En outre, l'ajout d'une nouvelle fonction entraîne la nécessaire réévaluation des algorithmes associés si l'on souhaite l'activer. Autrement dit, le coefficient arbitrairement affecté à chaque fonction et à chaque type de lien est susceptible de varier en permanence par l'entremise du programmeur. Une telle absence de cohérence de calcul conduit infailliblement à la méthode du « moins mauvais résultat. » Autrement dit, si la réponse attendue à une certaine requête offre le meilleur coefficient parmi tous les résultats possibles, rien n'est modifié ; en revanche, tout coefficient insuffisant ou toute réponse erronée conduit à la réévaluation précitée.

4.2.3.5 Diagnostic du DI en tant que réseau

En théorie, le Dictionnaire Intégral est un prototype possible du lexique idéal dans la perspective du traitement automatique de la langue. Jusqu'à présent, son existence relevait de la science-fiction aux yeux des linguistes. Dominique Dutoit a eu

la constance et l'énergie nécessaires pour mener à terme un tel travail, lequel a nécessité près de dix ans pour la seule implantation des données.

Pourtant, son développement est relativement simple. Il est ainsi extrêmement facile d'implanter un nouveau mot : il suffit de lui définir un lien de parenté avec un terme déjà instancié pour lui permettre de trouver immédiatement sa place dans le réseau, d'autres liens étant automatiquement activés à partir du premier. Rien n'empêche également d'activer ou de désactiver de nouveaux liens entre des lexies déjà présentes dans le graphe. D'autre part, tout le réseau est paramétrable à merci tant le programme qui le gère est malléable et complet. Enfin, la nature même du DI permettrait d'envisager son extension à différentes langues, à condition d'implanter tous les termes des langues visées mais sans qu'il soit nécessaire de modifier le programme d'aucune façon.

Au rayon des écueils, on peut citer la grande désorganisation générale du réseau, chaque lexème étant choisi arbitrairement et implanté de même, avant d'être réécrit ou déplacé le cas échéant sur le mode empirique : si les résultats du calcul de la distance sémantique sont bons, on laisse les choses en l'état ; s'ils sont mauvais, on réévalue.... Par conséquent, la validité générale du DI peut être remise en cause car elle ne repose sur aucun cadre théorique prédéfini, hormis la méthodologie et le cadre ontologique choisis par les concepteurs eux-mêmes.

D'autres difficultés subsistent comme celles précédemment évoquées concernant la sélection et l'ajout des fonctions lexicales et grammaticales ainsi que la décision de 'fermer' le réseau : à quel moment décide-t-on qu'une unité lexicale est saturée ou non ?

4.2.3.6 L'algorithme

Nous n'avons pas eu accès à la forme mathématique de l'algorithme utilisé pour exploiter les données du DI mais nous savons qu'il est fondé en grande partie sur le calcul de la « distance » sémantique entre les unités lexicales étudiées. Nous savons que cette distance se calcule principalement en fonction du nombre de liens qui séparent deux termes dans le réseau avec un coefficient différent pour chaque lien. Mais nous ne pouvons pas vraiment en dire plus sinon que les outils développés autour du DI utilisent tous le même algorithme, d'après les propres dires de Dominique Dutoit.

Il faut rappeler que ce système est à « corpus zéro » car il est entièrement fondé sur les connaissances issues du DI et ne nécessite aucun apprentissage pour l'activation et l'amélioration de ses performances. Cette méthode comporte un risque, qui est de reproduire en les amplifiant les erreurs qui ont pu être commises lors de l'implantation des unités lexicales dans le réseau. Toutefois, le pourcentage dérisoire de ces erreurs n'excède sûrement pas celui mesurable dans un corpus, d'autant qu'il est possible de les corriger facilement dès qu'elles sont détectées.

Voici maintenant un aperçu du fonctionnement de ces algorithmes sur un exemple de réduction lexicale, i.e. la paraphrase d'un syntagme nominal en un seul mot. L'objectif de la manœuvre sera ici d'effectuer la « réduction lexicale » de *Personne qui vend des fleurs*, c'est-à-dire de proposer pour ce syntagme la lexie-synonyme la plus appropriée.

Le Sémiographe, outil généraliste de TAL associé au DI, effectue le calcul simultané de la distance entre les différents termes qui composent le syntagme instancié et **toutes** les unités lexicales répertoriées dans le réseau. Pour chaque unité, trois résultats sont renvoyés, correspondant à trois modes de calcul différents :

Concernant le premier algorithme, il s'agit du résultat du calcul du 'plus court chemin' entre l'entrée (mot seul ou partie de syntagme) et la sortie (mot). Il est appelé « processus d'activation » par D. Dutoit. Supposons que nous souhaitions activer le chemin entre les lexies A et D.

Figure 2 : Chemin dit « d'activation »



Le processus d'activation consiste à calculer le rapport entre A et D sans passer par le calcul des segments AB, BC ou CD afin d'éviter des obstacles potentiels entre les arcs intermédiaires. Pour un syntagme comme *personne qui vend des fleurs*, il s'agit d'effectuer le calcul d'activation entre tous les mots « pertinents » du syntagme (à savoir ici les mots non grammaticaux *personne*, *vendre* et *fleurs*) et la totalité des mots du graphe.

Dans cet ordre d'idées, on considère que $A \subset D$, c'est tout ce qui est jugé utile pour ce calcul : le calcul du chemin est implicite, chaque lien intermédiaire se voyant assigné une valeur par défaut identique à celle de tous les autres liens, quelle que soit leur nature.

L'intérêt majeur de cet algorithme réside dans la rapidité de calcul qu'il permet. L'inconvénient, qui en découle directement, est qu'il ignore la spécificité de certains liens intermédiaires. A titre d'exemple, on peut citer le cas du rapport de proximité sémantique entre *eau* et *oxygène* d'une part, *eau* et *liquide* d'autre part. Selon le processus d'activation, le résultat sera le même dans les deux cas car il ne fera pas la différence entre une relation de type « élément atomique ; composant » et une autre comme « spécifique ; membre d'une classe », d'où le renvoi de résultats incorrects sinon incomplets.

Pour les deuxième et troisième algorithmes, les résultats vont de 0,1 (meilleur score d'activation) à 140 (plus mauvais) et se répartissent généralement de la façon suivante :

- Pour le deuxième algorithme, le résultat du processus d'activation est pondéré par les coefficients affectés à chaque type de lien. Une relation d'hyponymie se verra par exemple affectée d'un coefficient 0.5, une relation d'appartenance à une même famille 0.8 et ainsi de suite. Le produit de tous les coefficients des liens qui séparent deux lexies donne le score du chemin considéré. On comprend dès lors pourquoi le produit de liens proches (à valeur < 1) mène à un résultat proche de 0,1 alors que celui de liens plus distendus (à valeur > 1) fait augmenter le coefficient vers la limite « autorisée », i.e. 140.
- Le fonctionnement du troisième algorithme ne nous est pas apparu très clair, en tout cas pas directement perceptible lors de nos visites chez Dominique Dutoit qui n'a pas souhaité nous le dévoiler.

4.2.3.7 Les outils associés au DI

L'ambition de MEMODATA est d'élaborer et de proposer tous les outils possibles et imaginables en TAL à partir d'une seule interface. On peut citer comme déjà opérationnels à partir du DI le dictionnaire de synonymes, le routeur de messages, le lemmatiseur, le résumeur de textes, le phonétiseur (seul à ne pas fonctionner exclusivement sur le DI), le dictionnaire thématique, le paraphraseur et probablement d'autres à cette heure. Un analyseur syntaxique était récemment en projet. Pour celui-ci, l'équipe de Memodata recherchait activement une méthode efficace de représentation (cf. les grammaires de propriété), étant entendu qu'aucun modèle proposé à ce jour n'a fait ses preuves sur l'ensemble des formes syntaxiques d'une langue donnée.

Comme nous l'avons déjà mentionné, tous les outils fonctionnent sur le même algorithme et nous ne pourrions en valider le contenu qu'à la double condition que les formules utilisées soient cohérentes et que les outils seront capables d'avoir des résultats probants sur un corpus large, étant entendu que les expérimentations menées sur place portent essentiellement sur des exemples choisis par les concepteurs du système eux-mêmes.

4.2.3.8 Conclusion et perspectives autour du DI : un bilan mitigé

A notre avis, le Dictionnaire Intégral justifie à lui seul tout l'intérêt que l'on peut porter aux travaux de Dominique Dutoit. Les données qu'il renferme sont peut-

être uniques au monde et la plate-forme ergonomique qui gère le réseau est d'excellente facture. Il vaut évidemment mieux rester prudent avant d'encenser le système, mais en volume et en implantation informatique, le commentaire est pertinent.

L'autre atout de l'entreprise est d'être parvenue à élaborer un seul et même algorithme pour l'ensemble des outils développés. D'un point de vue linguistique, l'idée de mesurer la distance sémantique entre les termes (voire entre les syntagmes) nous paraît constituer une avancée majeure dans le domaine du TAL et nous y souscrivons avec d'autant plus d'intérêt que cette démarche est tout à fait conforme aux idées que nous entendons appliquer dans nos propres recherches sur les lexiques multilingues, lesquelles sont la justification du présent document. En revanche, nous entendons faire preuve d'une plus grande rigueur dans certains domaines, en particulier pour ce qui concerne la réévaluation des données douteuses.

Au rayon des faiblesses, nous pouvons citer à la fois le manque de visibilité et de lisibilité du système. Si nous sommes tout disposé à louer le DI pour sa seule existence, nous nous montrerons beaucoup plus circonspect en ce qui concerne la partie mathématique. Par ailleurs, la complexité et la lourdeur du réseau sémantique le rendent presque inutilisable par quelqu'un d'autre que les membres de l'équipe sans formation adaptée (d'où la nécessité d'un séjour sur place, que nous avons d'ailleurs effectué).

En guise de conclusion provisoire, nous insisterons à nouveau sur la richesse exceptionnelle du DI ainsi que sur les réserves que nous émettons au sujet de ses extensions. Pour une bonne exploitation du Dictionnaire Intégral, il nous semble indispensable de tester les différents outils opérationnels en situation « réelle. » En cas d'échec de l'algorithme, il sera plus facile de modifier ce dernier que de réécrire le réseau sémantique dans son intégralité.

En ce qui concerne la partie spécifique de la traduction multilingue, notre souhait fut initialement d'exploiter et d'aménager le réseau pour l'adapter à nos besoins sans être obligé d'en modifier la structure.

L'idée que nous eussions souhaité mettre en pratique était la suivante : étant donné que chaque unité lexico-sémantique (notion à préciser, chaque unité lexicale étant divisible en plusieurs unités sémantiques et vice-versa) du français est déjà implantée, peut-être pouvons les transformer en cases lexico-sémantiques multilingues où la lexie de chaque langue possède les mêmes caractéristiques sémantiques que les autres lexies de la même case. Il s'agirait simplement d'un lexique multilingue intégré à un réseau sémantique. Si cette éventualité trouvait confirmation dans la pratique, il resterait à effectuer le transfert syntaxique d'une langue à l'autre. Mais pour l'instant, notre objectif demeure exclusivement la désambiguïsation lexicale.

Le DI est dit universaliste apriorique, c'est-à-dire qu'il est censé englober toute (ou presque toute) la connaissance langagière d'un locuteur du français autant qu'il est

capable de prédire les énoncés valides. En d'autres termes, il est sous-tendu par la théorie $\text{texte} \rightarrow \text{sens} \rightarrow \text{texte}$ et autorise ainsi les opérations de compréhension et de génération avec la même facilité, comme nous l'avons évoqué en introduction.

Contrairement à des théories plus normatives ou conduisant naturellement à une certaine rigidification du langage, la TTST entend décrire la langue telle qu'elle peut (pourrait) être et non pas (ou plus seulement) telle qu'elle est (position indéfendable) ou telle qu'elle doit (devrait) être (vision normative non conforme à l'existant). Ce vœu pieux trouve partiellement un écho au travers du DI.

4.3 Quelques théories et applications en sémantique lexicale

4.3.1 *Théorie Sens/Texte par Mel'çuk, Polguère et alii.*

Les tenants de la théorie Sens/Texte posent le problème des limites de la description lexicale (« on crée une fonction lexicale à chaque fois qu'une lacune est rencontrée dans la description »). Le mode de construction du lexique à partir de cette théorie a pour effet de rigidifier le cadre d'emploi des unités lexicales et rend du même coup le langage « comme il doit (devrait) être » et plus seulement « comme il est » malgré la grande richesse des modèles proposés. Cela rappelle un peu le « locuteur idéal » imaginé par Chomsky pour les modèles syntaxiques, sans parler de la complexité des lexèmes et surtout du temps considérable nécessaire pour modéliser chaque acception.

Au demeurant, l'utilisation de fonctions lexicales comme le définissent les concepteurs du DEC peut aider avantageusement à la correspondance interlangue, comme précisé dans la leçon inaugurale d'Igor Mel'çuk. Mais de notre point de vue, elle risque de brider excessivement l'expressivité de la langue.

Une des caractéristiques les plus importantes du DEC est que les données sont ajoutées en fonction de l'intuition des concepteurs, et non pas par le truchement de données attestées. Le travail n'en est que plus long, même si nous ne le considérons pas comme invalidé par la méthode.

4.3.2 *Texte-sens-texte (TTST) selon Dutoit.*

C'est l'idée a priori la plus la plus proche de nos désirs mais pas la plus satisfaisante scientifiquement parlant. Le but est de modéliser le langage plus (ou pas seulement) comme il doit (devrait) être ni comme il est mais « comme il peut (pourrait) être. » Du côté mathématique, le problème réside dans l'approximation 'humaine et individuelle' des algorithmes employés. Du côté linguistique, on peut également douter de la validité de la structure sémantique du réseau conçu sur la base de cette théorie, à partir du moment où elle ne repose que sur les représentations mentales et sur la connaissance du monde du concepteur.

L'application de la TTST à la génération automatique de corpus est l'écueil le plus dangereux. La méthodologie consiste à représenter la langue telle qu'elle pourrait

être afin de générer un nombre très élevé d'énoncés servant à leur tour de corpus de base à d'autres représentations et applications du langage. Le péril est là évident. Il n'a d'ailleurs pas été évité.

4.3.3 Le lexique génératif (LG) ou le traitement de la polysémie par la représentation "pustejovskienne"

À tout le moins en apparence, le lexique génératif rend très imparfaitement compte de la polysémie au-delà des questions de polymorphie ou de polysémie logique. Les principes de coercion, co-composition ou de type pointé semblent inapplicables (ou non généralisables) en dehors d'exemples ad hoc, mais il est vrai une fois de plus que le formalisme est complexe pour qui n'a pas participé à son élaboration.

4.3.3.1 Présentation générale et critique d'ensemble

L'un de nos souhaits initiaux était de proposer une modélisation de la polysémie verbale avec l'aide de la théorie du Lexique Génératif élaborée par James Pustejovsky. Avec l'appui du LG, il paraissait possible à première vue d'imaginer une représentation commune des différents sens d'un verbe polysémique.

Or, nous estimons à présent que cela n'est pas tellement le cas et que le LG est un formalisme prioritairement voué au traitement de la polymorphie des structures syntaxiques induites par les mots du lexique et non pas à la polysémie de ces termes eux-mêmes. Pour renforcer ce point de vue, nous pouvons dire que Pustejovsky regroupe sous le terme de polysémie ce que nous appelons (pour ce travail) polymorphie, et qu'il nomme homonymie ce que nous entendons par polysémie.

Pierrette Bouillon corrobore ce point de vue dans sa thèse en présentant le LG comme l'outil idéal pour identifier les alternances, verbales ou nominales, les différences dans la forme syntaxique des compléments ainsi que les différences aspectuelles (opposition accomplissement/procès, par exemple). Il est cependant dommage qu'elle s'attache exclusivement à expliciter le principe de coercion, sans évoquer le potentiel de la co-composition ou de liage sélectif.

Le LG semble être également efficace en ce qui concerne l'acceptation ou le rejet des énoncés métonymiques :

19) Marie croit le livre

20) Marie prend le livre

*21) *Marie prend le livre puis croit tout ce qu'il dit*

Grâce à l'interaction entre les trois niveaux de la structure du mot 'livre', les deux premières phrases sont acceptées par le LG. Le processus consiste ici à présenter le lexème livre comme pouvant être soit une (source d') information, soit un artefact :

$$\text{ARGSTR} = \left[\begin{array}{l} \text{ARG1} = y : \text{information} \\ \text{ARG2} = x : \text{objet_physique} \end{array} \right]$$

Cette représentation de la polymorphie n'est pas sans rappeler les graphes conceptuels de Sowa. Ceci nous inspire qu'une entrée du LG n'est ni plus ni moins qu'une modélisation de la structure conceptuelle parcellaire d'un terme. La structure des qualia elle-même confirme ce point de vue en faisant le lien entre les deux types qui forment le type 'livre' (le livre est un artefact qui contient des informations) ainsi qu'en distinguant les différents prédicats intervenant dans sa structure conceptuelle.

Bien qu'offrant des perspectives plus larges pour le traitement de l'interface sémantique/syntaxe, le LG paraît encore loin de maîtriser la question globale de la conceptualisation. En effet, on ne doit pas ignorer que le livre est non seulement écrit et lu (ou susceptible de l'être), mais également édité, imprimé, acheté, récompensé, etc. Mais ces informations sont de toute façon probablement sans intérêt pour le traitement automatique en raison du double principe d'évitement des informations superflues et d'économie des traits implémentés.

4.3.3.2 Les mécanismes génératifs

- Le mécanisme de la coercion

Si l'on s'en tient au qualia de 'livre', certains rameaux du graphe conceptuel sont ignorés :

$$\text{QUALIA} = \left[\begin{array}{l} \text{information.objet-physique-lcp} \\ \text{FORMAL} = \text{contient} (x,y) \\ \text{TELIC} = \text{lire} (e2,w,x,y) \\ \text{AGENTIVE} = \text{écrire} (e1,v,x,y) \end{array} \right]$$

En vertu du principe de coercion, les formes syntaxiques qui ne sont pas prévues dans la structure des arguments mais qui respectent la structure prédicative contenue dans les qualia sont recevables par le LG. En lisant entre les lignes, nous sommes tentés de comprendre que les formes syntaxiques acceptables mais non prévues sont validées implicitement par le lexique.

Après réflexion, le processus est probablement beaucoup plus efficace que cela. Il est évident qu'une modélisation purement implicite d'un phénomène linguistique devient hasardeuse lorsqu'il s'agit de l'explicitier aux yeux des lecteurs profanes.

Pour mieux comprendre notre démarche réflexive, voici un exemple tiré de la thèse de P. Bouillon, phrase destinée à nous éclairer sur le principe de coercion :

- 22) *Jean est triste de lire ce livre*
(*objet=ÉVÉNEMENT*)
23) *Jean est triste à cause de ce livre*
(*objet=INDIVIDU NON-ANIME*)

D'après Pierrette Bouillon :

« Dans LG, la coercion agit au niveau sémantique. Elle permet à un prédicat de changer le type sémantique de ses arguments en celui qu'il requiert, pour autant que ce dernier soit présent dans la structure des qualia de cet argument. »
Bouillon (1996)

Pour les phrases (22) et (23), on note que *triste* ne sous-catégorise dans le lexique qu'un complément de type ÉVÉNEMENT. Et Pierrette Bouillon ajoute :

« Quand le type de l'argument est différent, il y a alors coercion : le prédicat a le pouvoir de transformer le type de l'argument en celui qu'il exige. »
Bouillon (op. cit.)

Étant donné que les miracles existent peu dans ce domaine, il nous a paru probable qu'il y avait autre chose derrière cette théorie de la coercion. Cette autre chose doit nécessairement être un phénomène mécanique pour permettre au LG de fonctionner au moins en théorie.

Observons ce qui se produit au niveau de la phrase (24). On sait déjà que 'triste' ne sous-catégorise qu'un complément de type ÉVÉNEMENT, et que 'à cause du livre' n'est pas reconnu comme un événement *a priori*. En regardant les qualia de 'livre', on s'aperçoit qu'il est susceptible d'être lu ou d'être écrit. Les traits *lire* et *écrire* renvoient à des unités lexicales répertoriées comme des événements à part entière. C'est ainsi que s'opère la magie du principe de coercion. Dans ce cas précis, c'est l'utilisation de l'objet (la possibilité d'être lu) qui autorise la construction de la phrase « Jean est triste à cause du livre » et permet ainsi l'ellipse du prédicat événementiel, devenu superflu. En fin de compte, il n'est pas nécessaire de préciser les concepts inhérents à la fabrication et à la diffusion du livre puisque les qualia prévus suffisent à rendre compte de la polymorphie observée en (22) et (23). On pourra juste ergoter sur l'approche conceptuelle du trait AGENTIVE en proposant une distinction entre les activités de création (*write*), de fabrication (*print, bind*) et de diffusion (*publish*). Ces trois opérations interviennent presque toujours dans le processus de production d'un livre.

Considérons maintenant les exemples suivants :

24) *Jean est triste à cause des délais d'impression du livre*

25) **Jean est triste à cause du livre (en parlant de son impression)*

Si l'on accepte le prédicat « impression » comme faisant partie de la structure conceptuelle et fonctionnelle de l'objet 'livre', alors la coercion doit autoriser la phrase « Jean est triste en raison du livre (en parlant de son impression). » A notre sens, les limites des principes génératifs énoncés par Pustejovsky se révèlent dès lors que l'on essaie d'étendre ses applications, que ce soit à des structures grammaticales nouvelles ou à de nouveaux maillons de la chaîne conceptuelle des termes étudiés. Observons l'exemple suivant :

26) « *Jean est triste à cause du livre.*

- *Quoi, il n'est pas bon ?*

- *Si, mais ça traîne à l'impression. »*

L'exemple (26) montre qu'un minimum de précision est apparemment nécessaire mais suffisant pour autoriser l'emploi de *Jean est triste à cause du livre* pour le cas où ce serait l'impression du livre qui était la cause de la tristesse. Manifestement, ce phénomène fonctionne également avec l'édition, la critique, la vente, etc. En ce sens, on peut dire que *le livre* peut se rapporter en contexte à tout élément de la chaîne conceptuelle qui s'y rapporte, sur le même modèle que le *prix Goncourt* décrit par Sowa.

Pour en revenir à la coercion, et si nous en avons correctement compris le principe, le Lexique Génératif fonctionne grâce à l'interaction entre les différentes entrées lexicales. Si l'analyseur rencontre un obstacle lié au contenu de l'ARGUMENT-STRUCTURE, il lui est possible de consulter la structure des qualia afin d'y trouver des ouvertures vers les concepts qui y sont exprimés.

Dans leur article « le traitement lexical de la coercion » (Cahiers de Linguistique Française, n°14), D. Godard et J. Jayez proposent une critique des mécanismes du principe de coercion énoncés par J. Pustejovsky. Ils lui reprochent en particulier d'énoncer des principes généraux, d'en démontrer leur efficacité sur des entrées lexicales dont il détermine lui-même la structure de traits et de prétendre qu'ils sont applicables dans tous les cas similaires.

Selon eux, ce parti pris ne rend pas compte des règles complexes de lexicalisation inhérentes à chaque unité lexicale. Nous allons maintenant discuter l'exemple qu'ils proposent pour illustrer leur propos :

27) *L'auditoire commence à écouter la symphonie*

28) *Jean commence à lire le livre*

29)**L'auditoire commence la symphonie*

30)*Jean commence le livre*

Les phrases (28) et (30) sont présentées par Pustejovsky pour illustrer le principe de coercion : Étant donné que 'lire' est la fonction habituelle (son rôle télique) attribuée à un livre, on peut en déduire la validité de la phrase (30) à partir de la (28).

Jayez et Godard contestent ce point de vue en indiquant que ce qui est valable dans ce cas précis ne l'est pas nécessairement pour d'autres constructions prédicat+objet tel que celui exprimé en (27). Partant du principe que la fonction d'une symphonie est d'être écoutée aussi bien qu'un livre est destiné à être lu, les auteurs réfutent donc la validité de la coercion telle qu'elle est conçue pour le Lexique Génératif.

On pourra objecter à ces derniers que les rôles attribuables à [*lire_livre*] et [*écouter_symphonie*] sont de natures différentes dans la mesure où le lecteur d'un livre a un rôle actif tandis que l'auditoire d'une symphonie n'a qu'un rôle passif. Dans ce dernier cas, ce sont les interprètes de l'œuvre musicale qui remplissent un rôle intermédiaire entre la création et la découverte. Le passage d'une œuvre à un public n'est donc pas un processus identique pour tous les objets concernés :

- écrire (x, livre) → lire (y, livre) ;
- écrire/composer (x, symphonie) → interpréter (y, symphonie) → écouter (z, interprétation_symphonie)⁵⁰.

Ceci nous ramène à une question importante : que doit-on choisir comme rôle télique pour le nom 'symphonie' ? La fonction d'une œuvre musicale est-elle d'être interprétée ou d'être écoutée ? Quelle qu'en soient les réponses, ces interrogations nous ramènent à une constatation d'ordre plus général : à notre avis, les mots du lexique sont des unités intrinsèquement trop complexes pour être représentées systématiquement selon un modèle unique.

□ Le principe de co-composition

Pustejovsky présente ainsi son principe de co-composition :

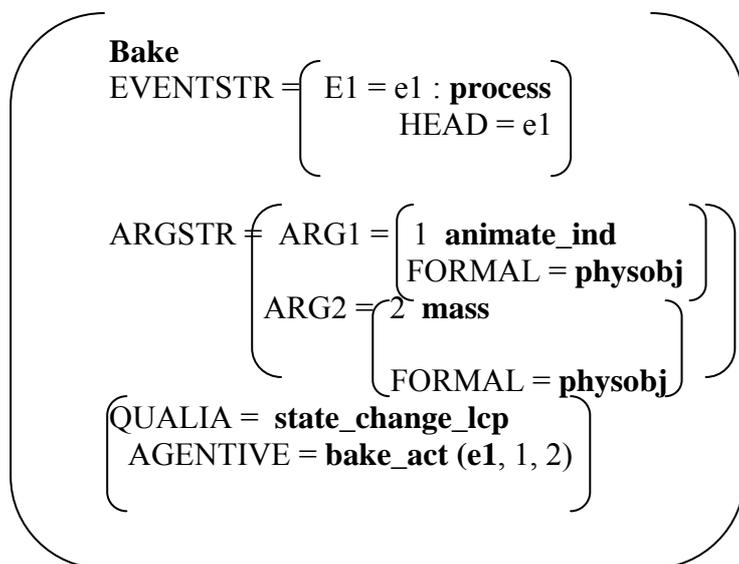
« Briefly, co-composition describes a structure which allows, superficially, more than one function application. We begin by studying the polysemy of baking verbs to illustrate this point. In (45) below, the verb 'bake' has two meanings, both a 'change of state' sense and a 'creation' sense.

(45) a. John baked the potato.

b. John baked the cake.»

⁵⁰ Où x, y et z correspondent à des individus ou à des ensembles d'individus effectuant une certaine action sur l'objet en second paramètre.

La structure lexicale proposée par J.P. pour le verbe ‘to bake’ est la suivante :



Ce qui est exprimé pour ce principe, et que Pustejovsky précise d’ailleurs, c’est que *bake* ne prend qu’un seul sens. Ce n’est donc pas le verbe qui varie ici, mais la relation logique entre le prédicat et son objet.

La co-composition fonctionne par unification d’un ou plusieurs traits de la structure des qualia du prédicat et de son objet. Ce principe a été nommé en anglais *cospecification* par Pustejovsky (1991). Dans l’exemple présenté ici, le trait *bake-act* désigne à la fois l’AGENTIVE de *bake* et celui de *cake*.

On peut imaginer sans peine que ce même trait est partagé par *potato*. A première vue, rien ne permet de distinguer *cake* de *potato* en termes de contraintes de lexicalisation. Dans ce cas précis, la co-composition nous apparaît plus comme un outil d’analyse linguistique mais semble moins pertinent (nous voulons dire utile) dans une perspective de modélisation.

L’exemple de *bake a cake/potato* nous paraît sous-tendre une propriété sémantique originale que l’auteur évoque à demi mot. Si la cuisson d’une pomme de terre induit un changement d’état de cette dernière, elle n’en reste pas moins une pomme de terre, on n’assiste dans ce cas qu’au passage d’une ‘pomme de terre crue’ à ‘une pomme de terre cuite’. En revanche, il n’en est pas de même pour un gâteau qui n’est pas un gâteau avant la cuisson. C’est en cela que le principe de co-composition devient intéressant dans ce cas précis, à savoir que l’on autorise le remplacement d’un lexème qui va être modifié (la pâte à cuire) par le lexème représentant le résultat du procès exprimé par le prédicat. La traduction sémantique littérale de l’expression ‘cuire le gâteau’ devrait donc être : ‘cuire la pâte afin de créer le gâteau’.

Ainsi, la co-composition va chercher dans le trait AGENTIVE de *cake* la légitimité de l'utilisation de ce dernier comme argument du prédicat. La transition est ici exprimée par le résultat.

4.3.4 Les classes d'objets, les paradigmes collocationnels (Véronis, Gross, etc.).

4.3.4.1 Les classes d'objets de Gaston Gross

Un des fils conducteurs des recherches de Gaston Gross est sa volonté constante de « rendre les dictionnaires plus actifs »⁵¹. Parmi les thèmes qu'il a développés, le plus important est sans doute la notion de classes d'objets ainsi que les « opérateurs appropriés » qu'il leur associe :

« Prenons deux substantifs comme *bus* et *manteau*. Les dictionnaires usuels ne précisent pas qu'il s'agit dans les deux cas de substantifs *concrets*, parce que c'est une évidence pour le lecteur. Les dictionnaires électroniques⁵² portent cette indication, puisqu'elle permet de « prédire » un grand nombre de phrases correctes. Par exemple, tous les concrets peuvent être compléments de verbes comme *voir*, *regarder*, etc., ou être sujets de prédicats comme *avoir (telle ou telle) couleur*, *peser (unité de mesure de poids)*, *coûter (unité de mesure de prix)*, etc. Au regard du codage, ces verbes sont compatibles avec les deux substantifs que nous avons envisagés.

Mais, ces indications, bien que nécessaires, ne sont pas encore suffisantes pour rendre compte du fonctionnement syntaxique de ces noms. En position d'objets du verbe *prendre*, ils sélectionnent des sens différents de ce prédicat.

Une phrase comme :

Luc a pris (le, son) manteau

Signifie à peu près qu'il ne l'a pas laissé à la maison. Tout autre est le sens de *prendre* avec *bus*. Dans la phrase :

Luc a pris (le, son) bus

Le verbe *prendre* peut être paraphrasé par *emprunter*, *voyager en*. On voit donc que si nous codons les deux substantifs avec le seul trait *concret*, nous ne sommes pas en mesure de prévoir le fonctionnement syntaxique de l'un et de l'autre, et par conséquent, de générer automatiquement toutes les phrases que l'on peut construire autour d'eux. Ajoutons donc une information supplémentaire, en codant le mot *bus* comme un *moyen de transport en*

⁵¹ De l'article du même nom : Gross G., 1996. « Rendre les dictionnaires plus actifs » in *Lexicographie et informatique. Autour de l'informatisation du TLF*. Actes du Colloque International de Nancy (29, 30 et 31 mai 1995) publiés par David PIOTROWSKI. Didier Érudition, Paris.

⁵² Gaston Gross oppose les « dictionnaires électroniques » destinés au traitement automatique de la langue aux « dictionnaires informatisés » qui se contentent de transformer les dictionnaires papier en version informatisée à l'identique.

commun et *manteau* comme un *vêtement*. Nous pouvons maintenant établir la liste des prédicats qui s'appliquent à l'une ou l'autre classe et, par voie de conséquence, à tous les éléments de chacune. Nous appelons ces sous-ensembles des **classes d'objets** et les prédicats qui les définissent des prédicats **appropriés**. »

Gross G. (1996b :196)

Comme le postulent Mel'çuk et ses collaborateurs, Gaston Gross propose la généralisation de traits spécifiques pour des substantifs dont la structure de traits permet de prédire à quels prédicats ils peuvent être associés. Cette méthode pose deux problèmes majeurs :

- Comme le rappelle Sowa dans sa critique de Katz et Fodor (voir dans ce chapitre en 4.2.), un ensemble de traits ne suffit pas pour rendre compte de la complexité des éléments du lexique. Il est nécessaire pour compléter ce dernier de prendre en compte des critères plus profonds comme les rôles et relations entre éléments du graphe conceptuel de chaque unité lexicale.
- En réduisant chaque terme à un membre d'une classe, on lui interdit dans une certaine mesure de faire jouer son potentiel d'essaimage polysémique vers d'autres classes, y compris si cette polysémie s'exerce au présent vers des classes ou il n'est pas (encore) lexicalisé.

Grâce aux classes d'objets, Gross caractérise une polysémie des prédicats en fonction de la description des traits de ses arguments. Outre sa réelle efficacité du processus en traitement automatique (au moins par rapport à un système simple sans classes d'objets), la méthode semble ignorer la polysémie potentielle des objets eux-mêmes. Reprenons l'exemple de *bus*, nous pouvons dire que ce mot est en contexte un *moyen de transport en commun*, un *moyen de transport individuel*, un *jouet miniature*, un *conducteur électronique* voire un *objet quelconque* ou un *projectile*⁵³ (en se mettant « à la place de King Kong » dans ces deux derniers cas). Pour chacune des classes précédentes, *prendre le bus* aura donc un sens différent qui ne peut être caractérisé que par le contexte. Dans notre propre méthodologie, nous utilisons les classes d'objets à la mesure de ce que les dictionnaires bilingues proposent comme collocations associées à chaque prédicat. Toutefois, nous espérons ouvrir nos bases à des objets et à

⁵³ Ce dégroupement des classes d'objets tient moins à la polysémie de *bus* qu'à la vision que chacun peut avoir de l'entrée lexicale. Pour cette dernière vision, un même objet *bus* peut rentrer dans des classes d'objets différentes et donc s'accoler à des prédicats différents selon Gross. En ce sens, le dictionnaire de Dutoit est plus proche de cette réalité par le jeu de son multidimensionnalisme que les classes d'objets.

des critères plus abstraits par l'observation des corpus littéraires sans préjuger du sens des lexies hors contexte.

Les théories sur les classes d'objets posent également une fois de plus le problème de l'origine des données : en général, celles-ci sont issues de corpus en majorité journalistiques ou encore nées de l'intuition des concepteurs. Cela étant admis, l'objectif des travaux de G.Gross sur les classes d'objets est de constituer une typologie des « objets du monde » au sens le plus large possible, afin d'obtenir une ontologie complète. Même si cette ontologie repose sur une vision figée de la langue, il s'agit d'un travail considérable.

4.3.4.2 Les paradigmes collocationnels

Initié par Jean Véronis, un grand projet de recensement des paradigmes collocationnels pour un grand nombre de mots du français n'a à notre connaissance jamais vraiment « vu la lumière. » Dans son projet d'origine, Véronis proposait la création d'une gigantesque base de collocations associées à chaque mot, le tout présenté sous la forme d'un dictionnaire monolingue. Toutes les informations devaient être recueillies à partir de journaux.

Au demeurant, cette entreprise est équivalente à la réalisation de John Sinclair, le COBUILD, pour l'anglais.

Dans le même esprit que la recherche des contextonymes effectuée par l'équipe de Sabine Ploux et dont nous venons de parler, Peter Blumenthal a également réalisé de très importants travaux pour constituer un dictionnaire des collocations du français dont on peut apercevoir quelques bribes dans le numéro 150 de Langue Française en juin 2006. Son évocation des collocations des termes d'affect ne sont d'ailleurs pas sans rappeler les recherches d'Eveline Martin dont on trouve trace dans son ouvrage de 1993.

Ces différents travaux nous amènent à effectuer une distinction importante entre les collocations qui se rapportent à des éléments fréquemment présents dans le contexte d'un terme et les arguments sujet et objet d'un verbe ou d'un élément prédicatif d'une autre catégorie grammaticale que le verbe. Pour le présent document, nous employons le terme collocations quasi exclusivement pour les seconds. Les collocations du co-texte proche, aussi nommées « déclencheurs » ou « contextonymes » dans certains cas, ne nous ont pas préoccupé pour l'instant de par l'ampleur et la complexité de leurs paradigmes. Le domaine des collocations est lui-même assez vaste et difficile à délimiter, comme le montrent Manguin et François (2006) en confrontant les « collocations étroites, asymétriques et lâches » à leurs modèles de synonymie. Les quasi-phrasèmes les plus établis ont selon leurs conclusions un caractère purement conventionnel et sont dans leur ensemble des « emplois combinatoires libres ».

4.4 La tentation de toucher à tout

John Sowa précise pour ses propres représentations qu'elles doivent faire apparaître toutes les informations pertinentes d'un mot pour un emploi donné. Dans ses critiques des formalismes qui ont précédé le sien, il en signale à chaque fois les limites en mettant en avant que certaines informations importantes ne pouvaient y être codées car ces formalismes ignoraient un ou plusieurs des mécanismes de relation sémantique, structurelle ou conceptuelle inhérents au langage.

La « tentation de toucher à tout » mène à la prise en compte de toutes les données lexicographiques, textuelles et intuitives possibles. Elle conduit également à l'étude de toutes les données « utiles » pour un transfert interlangue, qu'elles soient syntaxiques, sémantiques ou les deux à la fois. Enfin, elle cherche à aboutir à des modèles qui confrontent ceux proposés par les autres théories (fonctions lexicales, distance sémantique, champs sémantiques, primitives, classes ontologiques, fonctions syntaxiques, constructions, collocations, classes-objets, liens sémantiques multiformes, co-représentation multilingue, variables contextuelles, etc.).

Cette pseudo-théorie demande naturellement à être nuancée. Pour commencer, il n'est pas raisonnable d'envisager des modèles qui engloberaient la totalité des connaissances nécessaires à la « juste » interprétation d'une occurrence ou d'un énoncé. Ensuite, la recherche systématique de la précision la plus fine conduit à une formalisation tellement pointue qu'elle en devient inexploitable. Enfin, le simple recueil de toutes les données linguistiques et extra-linguistiques est tout simplement impossible.

Le mieux que nous ayons à proposer dans l'immédiat est de recenser l'ensemble des données pertinentes visibles dans les ouvrages lexicographiques, de les ordonner conformément à des modèles polymorphes, et de les enrichir le cas échéant par des exemples issus de sources non dictionnairiques.

Autant que faire se peut, nous éviterons d'avoir trop souvent recours à l'intuition, non pas parce que celle-ci n'est pas digne d'intérêt, mais avant tout parce qu'elle conduit à un travail de trop grande ampleur qui demande à être vérifié constamment dans le plus strict arbitraire ou pas du tout, au risque de se tromper gravement sur la réalité de la langue. Le DEC de Mel'çuk et de ses collaborateurs est le meilleur exemple connu de travail intuitif de longue haleine nécessitant de nombreuses « petites mains » pour la description d'un nombre de lexies très limité.

5. Transition

Après avoir discuté de nos plus importantes sources d'inspiration, nous allons initier dès le chapitre suivant la description du parcours que nous avons fait effectuer aux lexèmes pour les amener jusqu'à la transformation en lexies génératrices de liens synonymiques avec les autres lexies de toutes les langues.

Avant d'aborder les analyses linguistiques et les tentatives de modélisation que nous concluons par l'algorithme final, nous allons revenir sur l'origine de nos travaux,

à savoir sur la réalisation, l'observation et l'exploitation des Lexiques Inversés ou LI tels que nous les avons déjà définis en introduction à cette thèse.

CHAPITRE II
Outils et analyse lexicographiques : Lexiques Inversés et Construction de
Champs Lexicaux et Sémantiques

1. Introduction de partie

1.1 Présentation

Ce deuxième chapitre présente l'essentiel de nos travaux de dépouillement de données lexicographiques, la méthodologie sur laquelle nous nous sommes appuyé pour les mener, les outils d'analyse ainsi que l'ensemble des principes, matériaux et conclusions corrélatifs.

En premier lieu, nous expliquons le fonctionnement des Lexiques Inversés (LI), création originale qui constitue l'un des fondements de nos recherches, trouvant leur aboutissement dans la définition d'un algorithme général de construction de graphes sémantiques. Les LI et leur prolongement sont dirigés par un certain nombre de principes que nous détaillerons avec précision.

Ensuite, nous présentons les analyses à partir de plusieurs verbes différents pour illustrer un certain nombre de phénomènes inhérents aux correspondances entre langues :

- *aborder* et ses équivalents italiens comme étalon de la dissymétrie entre les deux parties d'un dictionnaire bilingue ;
- *abandonner* et ses traductions en tant que verbe représentatif de la distributionnalité sémantique (faible polysémie apparente, vaste champ synonymique en français et paradigme de traductions très important), ce verbe illustrant également le rôle important des collocations dans la sélection des candidats ;
- la périphrase verbale *se mettre à* pour mettre en avant que les auxiliaires et les périphrases aspectuelles fonctionnant morphologiquement comme des verbes ne peuvent pas être occultés dans notre modélisation ;
- les verbes de parole, ou plus exactement les verbes qui impliquent une composante explicite et obligatoire de communication verbale, pour montrer comment un corpus littéraire remet en question les données strictement lexicographiques.

Les analyses ainsi détaillées constituent les prémisses de la constitution des Champs Sémantiques Multilingues Unifiés (CSMU) dont le contenu et le fonctionnement seront présentés au Chapitre IV, après enrichissement par le corpus précisé au chapitre III et à la fin de celui-ci. La méthodologie de base de ces Champs sera partiellement explicitée à la fin de ce chapitre ; elle sera accompagnée de fragments des bases de données telles que celles-ci sont réellement saisies.

1.2 Quelques difficultés à surmonter

En premier lieu, un travail de constitution de bases de données « propres » suppose l'emploi d'un dictionnaire infallible ou considéré comme tel. En utilisant les dictionnaires bilingues comme ouvrages exclusifs de référence pour la construction de notre lexique, nous supposons qu'ils ne contiennent pas d'erreurs. D'autre part, nous partons également du principe qu'ils sont exhaustifs ou au minimum représentatifs de l'usage le plus fréquent et le plus adapté.

Bien sûr, ce minimum est insuffisant au regard de nos principes méthodologiques de base, dont précisément celui d'exhaustivité mais rien ne nous empêche d'ajouter ou de modifier des données en fonction des besoins.

Ensuite, nous avons besoin d'un dictionnaire à jour, c'est-à-dire qui tient compte de l'évolution naturelle du lexique de chaque langue et par extension des changements qui interviennent entre les langues. Commencés il y a sept ans, nos travaux de récupération de données ont partiellement subi l'érosion du temps par la faute de cette évolution du lexique. Plus encore que ce dernier, ce sont les dictionnaires publiés qui présentent énormément de modifications d'une édition à l'autre. L'un des arguments souvent évoqués pour expliquer ces rapides changements est la nécessité de renouveler les données pour justifier des rééditions quasi annuelles.

Comme nous le rappelons au chapitre premier, l'écueil de l'obsolescence est la principale source de problèmes en lexicographie, en particulier pour l'actualisation de dictionnaires comme le TLF. Ce dernier est une immense source d'informations sur le lexique de l'origine du français « à nos jours » mais pâtit presque de sa taille eu égard à la nécessité de veille qu'il induit. Par surcroît, il s'agit d'un dictionnaire spécialisé en littérature française et son enrichissement procède de fait d'une exégèse spécialisée pratiquée par de véritables lexicographes.

En ce qui concerne les LI, nous avons commencé l'extraction des données à partir de dictionnaires contemporains ou presque de la période où nous les avons initiés. Par la suite, nous nous sommes rendu compte que des dictionnaires plus récents proposaient de profondes modifications par rapport à leurs éditions antérieures ; les LI « primitifs » devenaient du même coup en partie obsolètes même si les changements en question ne remettent pas en cause notre méthodologie générale. Pour pallier cette donnée nouvelle, nous avons complété et enrichi le LI français-anglais à partir d'un dictionnaire de la même édition plus récent. Toutefois, il s'agit d'un travail considérable et il n'a pas été effectué pour les LI français-espagnol et français-italien, lesquels existent par conséquent dans des versions non actualisées bien que suffisamment récentes pour être représentatives.

Heureusement, nous pouvons nuancer notre propos en reconnaissant que la majorité des données lexicales qui étaient exactes il y a dix ou quinze ans le sont toujours aujourd'hui. D'autre part, nous pouvons espérer que la probable numérisation

des dictionnaires bilingues permettra un jour prochain de construire automatiquement des LI, ne serait-ce que terme par terme, au lieu d'avoir à les constituer manuellement comme nous l'avons fait.

Grâce à la méthodologie que nous allons développer tout au long de ce document, les modifications pourront être reportées directement dans la base de données sans nécessité aucune de repasser par une réécriture des LI. Sans vouloir anticiper sur les derniers chapitres, cette réactualisation facilitée représente l'objectif principal que nous entendons atteindre.

2. Le lexique inversé

2.1 Genèse

L'élaboration des Lexiques Inversés (LI) provient d'une observation critique des dictionnaires bilingues abrégés et de notre façon de les consulter. D'ordinaire, lorsque nous cherchons la traduction d'un mot, et plus particulièrement d'un verbe polysémique, nous avons l'habitude de vérifier notre choix en consultant l'entrée du mot candidat dans l'autre partie de l'ouvrage. Cette méthode nous a rapidement permis de constater que les traductions d'un mot ne proposent pas systématiquement ce dernier en traduction lorsqu'il se retrouve en position de vedette. Par ailleurs, nous avons découvert d'autres liens de traduction et de synonymie par ce biais.

En effet, il apparaît qu'il n'y a jamais de symétrie entre les deux parties d'un même dictionnaire en ce qui concerne les unités polysémiques, seuls les termes techniques échappant généralement à ce principe, même si ce n'est pas toujours le cas. En confrontant nos dictionnaires de référence à leur « reflet dans le miroir », nous parvenons à réduire en partie les carences éventuelles de leur forme non biunivoque.

2.2 Construction des lexiques inversés (LI)

Tout lexique bilingue normalement constitué présente deux parties distinctes, langue1 / langue2 et langue2 / langue1. Certains lecteurs pourraient croire qu'il existe une biunivocité systématique entre les deux. L'expérience prouve que ce n'est pas le cas et c'est donc ce constat qui nous a donné l'idée d'élaborer un lexique inversé pour chaque partie de chaque dictionnaire bilingue. Dans cette optique, nous proposons 3 dictionnaires bilingues revisités pour les verbes : anglais / français, italien / français et espagnol / français. Seul le manque de temps nous a empêché d'en construire davantage, en particulier les LI avec le français comme nouvelle langue source.

Le principe général des LI ou « lexiques miroirs » est d'écrire la partie Langue1 / Langue2 en n'utilisant que la partie Langue2 / Langue1, et réciproquement. Pour ce faire, nous plaçons en entrée toutes les vedettes rencontrées et nous faisons glisser les mots-sources correspondants en vedettes.

Nous allons expliciter la méthode à l'aide de l'exemple suivant qui reprend des fragments d'entrées de *décocher* et *lancer* dans le dictionnaire français-italien Bordas :

Décocher, V

- 1⁵⁴ *scoccare*(« une flèche »)⁵⁵
- 2 *sferrare*(« un coup de pied »)
- 3 (*fig*) *scoccare, lanciare*

Lancer, V

- 1 *lanciare*
- 2 *gettare*

Entrées inversées directement issues de la précédente :

- *gettare* (1) = *lancer*
- *lanciare* (2) = *décocher (fig)-lancer*
- *scoccare* (1) = *décocher*(« flèche » | *fig*)
- *sferrare* (1) = *décocher*(« coup de pied »)

En vertu d'un principe d'accumulation, les nouvelles entrées sont disposées à accueillir d'autres vedettes. Dans les lexiques que nous avons construits, le chiffre précédant le signe '=' est le compteur non-automatique des vedettes.

Le résultat de l'opération est la création d'un nouveau dictionnaire ou lexique italien / français. Autrement dit, nous disposons de deux entrées par mot de chacune des deux langues grâce à ce procédé. Dès lors, il nous est possible de comparer les divergences entre les deux parties, comme l'illustrent les exemples avec les verbes *aborder*, *abandonner* et *traverser*, et la périphrase *se mettre à*, pour lesquels nous observons une forte dissymétrie entre le dictionnaire français / langue-cible « réel » (que nous appelons « droit ») et l'inversé, le LI.

3. Exploitation des données des dictionnaires

3.1 Retour sur les principes d'origine

À partir des règles que nous venons de décrire, le premier travail que nous proposons est d'effectuer le relevé détaillé des informations apparentes pour un lexème de chaque dictionnaire bilingue que nous utilisons. L'objectif de ce

⁵⁴ La numérotation correspond au regroupement des acceptions tel qu'il apparaît dans les dictionnaires bilingues de référence.

⁵⁵ Les collocations habituelles, tout comme les informations liées au registre de langue, doivent apparaître dans un lexique complet tel que nous l'envisageons.

recensement est avant tout de répertorier les éléments essentiels pour un rapport de traduction « complet », ainsi éventuellement que d'en isoler les critères non pertinents.

Nous ferons émerger l'exhaustivité au travers d'un relevé minutieux. Au nom de la limpidité, la liste des informations nous obligera à réfléchir à un mode de classement et de représentation clair. Les informations syntaxiques, sémantiques et autres traits nécessaires au transfert nourriront les principes de valeur de transfert en général (y compris la sous-catégorisation). Enfin, nous défendrons l'étude du reflet du dictionnaire dans le miroir pour ajouter à l'existant perceptible les informations dont l'existence est masquée par leur emplacement dans la « mauvaise » partie du dictionnaire.

3.2 Détail des entrées dans nos dictionnaires de référence

3.2.1 Une entrée du R&C français-anglais

L'entrée de *to leave* dans le Robert & Collins français-anglais est présentée telle quelle, sans ajout ni omission d'aucune sorte. Chaque discriminateur sera relevé une fois et une seule à partir de cet extrait et nous en donnerons la description précise :

leave⁵⁶ /li:v/⁵⁷ (vb⁵⁸ : prêt, ptp **left**⁵⁹)
VT⁶⁰ **a**⁶¹ (= *go away from*⁶²) [+ *town*⁶³] *quitter*⁶⁴, *partir*
*de*⁶⁵; (*permanently*⁶⁶) *quitter*; [+ *room, building*⁶⁷]
sortir de, quitter; ⁶⁸ [+ *person, job, one's husband, wife*] *quitter*; [+ *one's children*] *abandonner* ♦ **he left Paris in 2001**⁶⁹ il a quitté
 Paris en 2001⁷⁰ ♦ **we left Paris at 6 o'clock**
 nous sommes partis de Paris *or* nous
 avons quitté Paris à 6 heures...
 (...)
e (*Math*⁷¹) **three from six ~s**⁷² **three** six moins

⁵⁶ Entrée lexicale

⁵⁷ transcription phonétique

⁵⁸ catégorie grammaticale

⁵⁹ formes fléchies et/ou conjuguées

⁶⁰ indicateur de transitivité

⁶¹ marqueur de polysémie

⁶² équivalent en langue source

⁶³ collocation objet

⁶⁴ traduction en langue cible

⁶⁵ double traduction en langue cible

⁶⁶ information complémentaire pour la langue source

⁶⁷ double collocation objet

⁶⁸ changement de contexte langagier

⁶⁹ syntagme ou phrase exemple

⁷⁰ traduction du syntagme exemple

trois égalent⁷³ trois...

f (in will) [+ money] laisser {to à⁷⁴} ; [+ object, property] laisser, léguer {to à⁷⁵}...

VI (= go away) [person, train, ship etc]⁷⁶ partir, s'en aller ; (= resign) partir, démissionner

♦ **to ~ for Paris** [person, train] partir pour⁷⁷

Paris ; [ship] partir or appareiller⁷⁸ pour Paris...

▶ **leave off**⁷⁹ **VI** (*⁸⁰ = stop) s'arrêter ♦ **where did**

we ~ off? (in work, reading⁸¹) où nous sommes-nous arrêtés? ♦ **~ off !** arrête !, ça suffit ! *

VT SEP a (* = stop) arrêter (doing sth de faire qch)...

▶ **leave out VT SEP**⁸² **a** (= omit) (accidentally)

oublier, omettre ; (deliberately) exclure ;

[+ line in text] (also Mus)⁸³ [+ note] sauter...

Chaque élément constitutif des entrées possède des caractéristiques et des justifications que nous allons à présent détailler :

1) Lexème

C'est le lexème, le mot qui est traduit (ici *leave*). Dans les lexiques inversés, tous les lexèmes se retrouvent en position de candidats des traductions proposées.

2) Forme phonétique

La forme phonétique serait utile dans le cadre d'un système de compréhension par la reconnaissance vocale et est indispensable aux apprenants pour la juste prononciation des mots, mais cela ne fait pas partie de nos préoccupations pour le présent travail.

⁷¹ domaine d'emploi

⁷² remplacement du mot source dans le texte

⁷³ traduction en langue cible dans le syntagme exemple

⁷⁴ constructions syntaxiques liées au contexte

⁷⁵ idem que le point précédent pour deux verbes en même temps

⁷⁶ collocations sujet

⁷⁷ construction syntaxique supplémentaire

⁷⁸ double traduction en langue cible dans les syntagmes exemples

⁷⁹ verbe à particule

⁸⁰ marqueur de registre de langue

⁸¹ indicateur de situation d'énonciation spécifique

⁸² séparation possible du verbe et de la préposition

⁸³ marqueur d'utilisation éventuelle dans un registre spécifique

3) Catégorie grammaticale

La catégorie grammaticale est un critère essentiel pour la reconnaissance des unités lexicales en contexte. C'est d'autant plus vrai en anglais où de nombreux mots peuvent aussi bien être noms, adjectifs ou verbes, sans marque affixale particulière. C'est notamment le cas de *leave* qui peut être aussi bien un nom (= congé, permission, etc.) qu'un verbe dans sa forme lemmatique de base.

La catégorie doit apparaître dans un graphe complet mais n'apparaît pas forcément dans nos données car seuls les verbes y figurent, la valeur par défaut étant donc « V(erb). » Toutefois, il peut arriver qu'un verbe soit traduit par une unité lexicale d'une autre catégorie grammaticale, et vice-versa. Dans ce cas, la catégorie « déviante » devrait être mentionnée.

4) Formes conjuguées

Il s'agit des formes conjuguées particulières de l'entrée, en l'occurrence le prétérit et le participe passé d'un verbe anglais irrégulier. C'est un critère non utilisé pour la construction du graphe, mais il serait nécessaire de l'implanter en contre-champ pour une éventuelle application de nos lexiques à des corpus (lemmatiseur).

Remarquons qu'à l'instar de la forme infinitive du verbe, les formes conjuguées sont volontiers homographes de mots appartenant à d'autres catégories (*left* comme adjectif, par exemple).

5) Indicateur de transitivité

L'indicateur de transitivité sert à préciser si le verbe en entrée est habituellement transitif ou intransitif ainsi que pour séparer les traductions des verbes qui peuvent être les deux selon la norme.

Ce trait n'est pas reproduit tel quel dans notre base mais est tout de même explicite : la présence de collocations objet y indique que l'on traite une forme transitive (avec une distinction entre la transitivité directe et indirecte, la bitransitivité cumulant les deux, la bitransitivité directe et la bitransitivité indirecte), leur absence dénote les formes intransitives.

6) Dégrouper typographique des traductions

La numérotation est propre à chaque dictionnaire pour dégrouper les paradigmes de traduction en fonction des acceptions de l'entrée. Ce dégrouper est assimilable au dégrouper polysémique ou homonymique selon le cas.

7) Synonyme en langue source

Mot équivalent en langue-source de l'entrée dans un contexte linguistique donné. Apparaît dans la colonne « synonyme LangueSource » générateur d'un lien de synonymie entre ce terme et l'entrée.

8) Collocation objet apparente

Collocation objet exprimé dans la langue du mot vedette. Apparaît dans une des colonnes « collocations objet » ; il est commun à l'entrée et au mot traduit, bien que seulement exprimé dans la langue source.

9) Traduction littérale

Traduction de la vedette.

10) Concurrence de traductèmes

Présence d'au moins deux traductions en langue cible de la vedette dans des contextes linguistiques équivalents.

11) Informations complémentaires

Information complémentaire d'ordre sémantique portant sur la vedette et ayant une influence sur la traduction (le mot cible étant de son côté porteur de l'information → *leave permanently = quitter*).

12) Concurrence de collocations objets

Présence d'au moins deux collocations objet distinctes pour un même paradigme de traductions. Ces collocations sont elles-mêmes des classes (et non des fragments de discours littéral) induisant des paradigmes lexicaux sous-jacents.

13) Marqueur typographique de changement de contexte langagier

Le point-virgule indique qu'on passe à un autre contexte langagier. Dont acte, il est important de ne pas confondre ce marqueur avec une virgule simple, laquelle est une disjonction qui indique que le contexte est le même pour les traductions qu'elle sépare. Virgule et point-virgule sont également séparateurs pour d'autres éléments que les traductèmes.

14) Syntagme exemple

Syntagme exemple. Très important car il peut faire apparaître d'autres traductions, d'autres collocations, de nouvelles constructions et d'autres liens de synonymie en langue source ou en langue cible que ceux proposés précédemment. Leur analyse oblige à un travail de fourmi fastidieux mais nécessaire dans un souci d'exhaustivité.

15) Traduction du syntagme exemple

Traduction du syntagme exemple. Voir point 14 pour sa prise en compte.

16) Indicateur de domaine d'emploi

Indicateur de domaine d'emploi : typographie presque identique aux « infos complémentaires » à l'exception de la majuscule initiale. Présent dans la colonne

« registre » qui inclut le niveau de langue, le domaine d'emploi et éventuellement la zone géographique où l'acception courante a cours.

17) Marqueur typographique de remplacement du mot source

Le ~ sert à remplacer la graphie littérale du mot source. Ressemble à un pointeur sur une case-mémoire.

18) Nouvelle traduction

Nouvelle traduction repérée à l'intérieur de la traduction d'un syntagme-exemple, élément ajouté dans la colonne « mot-cible », informations langagières ajoutées par ailleurs si elles sont repérables (collocations, constructions, etc.).

19) Constructions syntaxiques

Constructions syntaxiques particulières associées à l'acception traitée : si le verbe est transitif indirect, on ne fait apparaître que la préposition précédant l'objet, lequel est présent juste avant. L'information portant sur la langue source est en premier et en italique ; la construction langue cible est en second en écriture normal.

20) Constructions syntaxiques communes

La construction en langue cible de cet exemple concerne les deux verbes candidats proposés.

21) Collocations sujet

Collocations sujet ou restriction sur la sélection sémantique du sujet pour des emplois donnés. Comme pour les objets, l'information est commune aux deux langues⁸⁴ bien qu'il ne soit exprimé qu'en langue source.

22) Construction syntaxique dans les syntagmes exemples

Construction syntaxique particulière non prévue dans le paradigme de traductions principal. A repérer et à ajouter le cas échéant dans nos lexiques, si l'on considère qu'il s'agit d'une forme syntaxique pertinente (i.e. fréquente et/ou complément, par opposition aux circonstants).

23) Nouvelle traduction dans les syntagmes exemples

Nouvelle traduction non présentée d'emblée et mise au même niveau qu'un autre terme-cible dans les syntagmes exemples.

⁸⁴ Sauf avis contraire dans l'article du dictionnaire : les constructions peuvent être inversées d'une langue à l'autre (voir la différence entre *manquer à* et *to miss*)

24) Verbe à particule

Verbe à particule. Pour de nombreux verbes anglais, ces formes sont des lexies à part entière et nous souhaitons les traiter comme telles au sein du graphe. Dans ce cas *leave off* fera peut-être l'objet d'un traitement différencié par rapport à *to leave* en devenant un « mot-source ».

25) Marqueur typographique de niveau de langue

Le * sert à signaler un emploi familier, en l'occurrence celui de *leave off* avec le sens de *s'arrêter*. Plus le nombre d'étoiles est élevé, plus l'expression est considérée comme grossière. Peut faire l'objet d'une colonne à part, non prévue pour l'instant, ou bien être codé dans la colonne *registre*.

26) Information pragmatique

Information pragmatique sur le contexte langagier, même typographie que les domaines d'emploi ou les informations complémentaires de type adverbiales. On suppose qu'il peut être implicite aussi bien qu'explicite selon les circonstances. Utile lorsque le contexte d'énonciation peut permettre de distinguer deux acceptions.

27) Verbes à particule séparables et fusionnés

SEP indique que le verbe et la préposition sont séparables dans certaines constructions du verbe à particule considéré. Lorsque le verbe et la préposition ne sont pas séparables, le dictionnaire le mentionne par FUS (comme fusionné). Critère non répertorié mais important, à étudier.

28) Domaines spécifiques supplémentaires

« Also » est présent pour préciser que la traduction générique proposée s'applique également à un domaine spécifique, en l'occurrence *la musique*. Dans ce cas, deux lignes de tableau sont nécessaires pour distinguer les collocations alors que la traduction reste la même, car nous ne fusionnons pas d'emblée les données d'emplois différents.

3.2.2 Une entrée du Bordas français-italien

lasciare⁸⁵, v.t.⁸⁶ 1.⁸⁷ lâcher⁸⁸: *lasciami*⁸⁹!, lâche-moi !; _⁹⁰
le briglie, lâcher les rênes ; _ *la presa*, lâcher prise ;
non lasciare il cane, ne lâche pas le chien || *ha lasciato*

⁸⁵ Entrée lexicale (cf. 1 pour l'anglais)

⁸⁶ Marqueur de transitivité (cf. 5 pour l'anglais)

⁸⁷ Indicateur de dégroupement (cf. 6 pour l'anglais)

⁸⁸ Traduction (cf. 9 pour l'anglais)

⁸⁹ Phrase exemple en langue cible suivie de sa traduction (cf. 14 et 15 pour l'anglais)

⁹⁰ Le _ remplace le verbe en entrée, afin d'éviter d'avoir à le répéter (cf. 17 pour l'anglais)

tutto per dedicarsi alla letteratura, il a tout lâché pour se consacrer à la littérature || _ *il gruppo, (ciclismo)*⁹¹ lâcher le peloton 2. (*seguito da un infinito*)⁹² laisser :...

Le dictionnaire français-italien Bordas présente une structure plus rudimentaire que le Robert & Collins français-anglais, lequel offre nettement plus d'informations et de variations typographiques pour les mettre en exergue. À première vue, la seule nuance typographique notable est l'emploi de l'italique pour tout ce qui concerne la langue source, du lexème aux marqueurs de registre, et d'une typographie normale pour tout ce qui concerne la langue cible. Faute de plus de variété dans la manière de présenter les rubriques, la consultation de ce dictionnaire est objectivement assez fatigante voire fastidieuse.

3.2.3 Une entrée du Larousse français-espagnol

dejar⁹³ v. tr.⁹⁴ Laisser⁹⁵ : *lo he dejado en casa*⁹⁶, je l'ai laissé à la maison ; *le ha dejado algo por ou sin hacer*, il vous a laissé quelque chose à faire ; *dejalo tranquilo*, laisse-le tranquille ; *dejar a uno el cuidado de*, laisser à quelqu'un le soin de ; *dejar improductivo un capital*, laisser dormir un capital. ||⁹⁷ Déposer : *dejè a mi amigo en la estación*, j'ai déposé mon ami à la gare. || ☐ Quitter : *dejar a su mujer*, quitter sa femme ; *dejar un empleo*, quitter sa place. || Laisser, abandonner : *dejar a sus hijos*, abandonner ses enfants. || Cesser, arrêter : *deja de trabajar a las seis*, il arrête de travailler à 6 heures ; *dejò de escribirme*, il a cessé de m'écrire. || Rapporter : *este negocio deja mucho dinero*, cette affaire rapporte beaucoup d'argent. || _ *Dejad que los niños se acerquen a mi*, laissez venir à moi les petits enfants. || FAM.⁹⁸ ; *Déjalo !*, laisse tomber ! | *¡Déjame en paz!*, laisse-moi tranquille !, fiche-moi la paix ! || *Dejar al descubierto*, découvrir (*un ejército*)⁹⁹. || *dejar aparte*, laisser de côté. || FIG. *Dejar a salvo*,

⁹¹ Marqueur de domaine

⁹² Information grammaticale, construction particulière (cf. 19 pour l'anglais)

⁹³ Entrée lexicale (cf. 1 pour l'anglais)

⁹⁴ Marqueur de transitivité (cf. 5 pour l'anglais)

⁹⁵ Traduction (cf. 9 pour l'anglais)

⁹⁶ Phrase exemple en langue cible suivie de sa traduction (cf. 14 et 15 pour l'anglais)

⁹⁷ Indicateur de dégroupement (cf. 6 pour l'anglais)

⁹⁸ Marqueur de registre (cf. 16 et 28 pour l'anglais)

⁹⁹ Collocation objet (cf. 8 pour l'anglais)

faire abstraction de. || *Dejar atrás*, laisser en arrière, lâcher, distancer. || *Dejar a un lado ou de lado*, laisser de côté. || FIG. *Dejar caer*, glisser : *dejò caer el la conversaciòn que querìa irse a España*, il glissa dans la conversation qu'il voulait partir en Espagne. || *Dejar como nuevo*, remettre à neuf (*una cosa*), remettre (*una persona*). || *Dejar correr*, laisser faire, laisser courir. || FAM. *Dejar chiquito*...

Dans le Larousse français-espagnol, l'accent est surtout mis sur les traductions du terme source ainsi que sur la variété des locutions et expressions où celui-ci figure. Les entrées de ce dictionnaire donnent peu d'informations sur les collocations et de ce fait ne permettent pas de manière claire d'associer directement à chaque groupe de traductèmes l'acception considérée, hormis en faisant des inférences sur la relation synonymique entre d'éventuelles traductions communes pour une même acception de la vedette.

La typographie utilisée est presque la même que pour le Bordas, seule la distinction *italique* / pas italique étant utilisée pour séparer visuellement les informations.

4. Un exemple flagrant de dissymétrie : les traductions du verbe *aborder* en italien

4.1 Préparation française du verbe, pour *aborder* *aborder*

*4.1.1 Double nominalisation : *abord* et *abordage**

Dans des travaux anciens (Haton, 1999), nous tentions d'évaluer le degré d'éloignement des sens littéraux et métaphoriques de certains verbes, dont celui qui nous occupe dans le présent chapitre.

En vertu du principe énoncé par Todorov (1967), nous entendons plutôt donner à cette problématique une « réponse synchronique », ce qui n'exclura pas de rechercher dans l'histoire de ce verbe quelques éléments de réflexion. Considérons tout d'abord les substantifs déverbaux associés à *aborder* :

31) *L'abordage du navire a été difficile*

32) **L'abordage du problème a été difficile*

33) **Ce navire a eu un abord facile*

34) *Ce problème est d'un abord facile*

Ces exemples semblent montrer que le verbe *aborder* se nominalise par *abordage* dans son sens littéral tandis qu'il admet seulement *abord* comme équivalent

nominal dans son acception métaphorique. Ce constat immédiat remet apparemment en cause un des facteurs premiers de la polysémie, à savoir le lien fort de parenté entre les différentes acceptions d'un même terme, ainsi que la mise en commun des termes dérivés (substantifs, adjectifs, etc.).

Pourtant, les sens métaphoriques et non métaphoriques de ce verbe sont étymologiquement corrélés, les premiers ayant un jour découlé des seconds. Un petit tour par la diachronie n'est donc pas complètement superflu pour essayer de comprendre la polysémie de ce verbe en synchronie. L'hypothèse étymologique intuitive selon laquelle *abordage* est la première nominalisation du verbe *aborder* en tant que terme maritime (son emploi initial), puis que le nom *abord* a été choisi par la suite pour caractériser les emplois métaphoriques ne tient pas selon nos sources. En effet, *abord* est apparu au 15^{ème} siècle, 85 ans avant *abordage*, comme nominalisation du sens de « fait d'avoir accès à une personne ». Aujourd'hui, *abord* est désuet dans ses acceptions de « action d'aborder au rivage », alors que *abordage* n'a jamais servi dans le sens métaphorique.

Abordage dans le TLF

1553 « action d'entrer dans un port (d'un navire) » terme mar. (P. BELON, *Observations*, III, 13 ds *R. Philol. fr.*, XLIII, 174 : Les juifs ... mettent de plus haut pris à la recette du revenu des provinces, affermans les gabelles, et l'**abordage** des navires et autres choses de Turquie); 1616-20 « *id.* », *id.* (D'AUBIGNÉ, *Hist.*, II, 179 ds LITTRÉ : Les Rochellois, dès l'**abordage** [des navires], se jettent sur le pont de corde); **2.** 1660 « assaut donné à un navire dans un combat » *id.* (CLEIRAC, *Explic. termes mar.*, 53 ds JAL² : On distingue l'**abordage** de franc-étable (éperon contre éperon), l'**abordage** en belle (éperon contre flanc)); **3.** 1687 « collision volontaire ou non de 2 navires » *id.* (CHOISY, *Voy. de Siam*, 141, *ibid.* : Cependant il [le navire] arrivoit sur nous, et nous alloit aborder à bâbord... on a donné un coup de gouvernail pour éviter l'**abordage**).

Dér. de *aborder** : 1 de *aborder* 2, terme mar.; 2 et 3 de *aborder* 1, terme mar.; suff. *-age**.

Abord dans le TLF

a) 1440-1475 « fait d'avoir accès à une personne » (CHASTELLAIN, *Chron.*, IV, 379, éd. Kervyn ds *R. Hist. litt. Fr.* I, 180 : Tandis doncques que ces anglois estoient devers le duc besognant pour leur premier **abord**); 1611 « accueil fait par la personne abordée » (COTGR. *s.v.* *abord* : De doux **abord**. Gentle, courteous, affable : of open accesse, easie to be spoken with, or come into); **b)** 1636 « attaque » (CORNEILLE, *Cid.* 1087 ds DUB-LAG. 1960 : De ces vieux ennemis va soutenir l'**abord**), d'où 1575 *de premier abord* « dès le commencement » (MARG. DE FRANCE, *Mém.* ds *D G* : *De premier abord*, l'effroy l'ayant pris); 1607 *d'abord* « *id.* » (HULSIUS, *Dict. fr.-all.*); **2.** 1616-1620 « action d'arriver en un lieu » (D'AUBIGNÉ, *Hist.*, II, 183 ds LITTRÉ : Il vint à la cour en poste, et deux heures après son **abord**, Pellicar, ...); **3. a)** *ca* 1530 « lieu (où l'on est arrivé) »

(C. BUCHER, *Poésies*, éd. Denais, 87 ds *R. Hist. litt. Fr.*, I, 180 : Que cerches-tu ores en ces **abords?**); 1556 « lieu où l'on aborde, port » terme mar. (SALIAT, *Trad. d'Hérodote*, éd. chez E. Groulleau, III, 5 ds HUG. : Plusieurs ports et **abords** de mer ou se fait grand trafic); **b**) ^{xvi^e} s. *abords* « rive le long d'un cours d'eau » terme mar. (*Cout. de L'Angle, Nouv. Cout. gén.* I, 312 ds LACURNE, t. 1, 1875 : Est ordonné... a un chascun ayans **abords** contre la grande rivière... qu'ils ayent à les entretenir).
 Dév. de *aborder** : 1 a de *aborder* 3; 1 b de *aborder* 1; 2 et 3 a de *aborder* 2; 3 b dér. de *bord** 2 b.

HISTORIQUE

I. — Hist. des sens. — A. — « fait d'avoir accès auprès d'une pers. », 1^{re} attest. 1440-75 (*cf.* étymol.) et attesté surtout dans les accept. suiv. : **1.** « possibilité d'avoir accès auprès d'une pers. », apparaît ds *Ac.* 1718 et se trouve durant toute notre période; **2.** « accueil, aspect de la pers. abordée », 1^{re} attest. 1611 (*cf.* étymol.) et constamment attesté. **B. —** « lieu où l'on est arrivé », attesté en 1530 (*cf.* étymol.), sans attest. post., sinon dans les accept. suiv. : **1.** « lieu où l'on aborde, port », attesté en 1556 (*cf.* étymol.) et encore en 1610 : Le Roy... le comble [l'hermitage d'Honneur] de toutes les singularitez dont il se peut aviser, l'ayant fait un **abord** de toutes sortes de gens d'honneur. BEROALDE DE VERVILLE, *Voyage des Princes fortunez*, 322 (Hug.). Pas d'attest. jusqu'en 1908, où cette accept. reparait dans un emploi plus restreint (*cf.* ex. 6); **2.** « rive le long d'un cours d'eau », attesté au ^{xvi^e} s. Qq. rares attest. jusqu'au début du ^{xix^e} s. (peut-être ds *ex.* 2). **C. —** « arrivée en un lieu », 1^{re} attest. 1616-1620 (*cf.* étymol.). En ce sens gén., *abord* aurait pu prendre la place d'un autre mot, *arrivée*, de même orig. mar., mais il est condamné comme vieilli au ^{xviii^e} s. (*cf.* BRUNOT t. 6, pp. 1335-1336; *cf.* également BESCH. 1845, *s.v.*), est relayé par certaines accept. partic. : **1.** fig. « action d'aborder une entreprise », noté ds *Ac. Compl.* 1842 (le commencement d'une action), et qui possède qq. attest. au ^{xx^e} s., mais qui se trouve déjà dans les loc. apparues dès la fin du ^{xvi^e} s. : *de premier abord*, 1575; *d'abord*, 1607 (*cf.* étymol.); **2.** à noter, de même que pour les pers., la nuance entre le fait et la possibilité : « possibilité d'accéder à un lieu, spéc. par mer », 1^{re} attest. ds *FUR.* 1690 et attesté durant toute la période; **3.** p. ext. « arrivée massive, affluence », 1^{re} attest. 1687 : Les autres n'étaient que des hôteliers que le grand **abord** des étrangers enrichissait. FONTENELLE, *Orac.*, I, 14 (Littré). Attesté dans les dict. jusqu'à la 1^{re} moitié du ^{xix^e} s. Noté vieilli ds *Ac.* 1835 et BESCH. 1845; **4.** « présence », de 1701 (*FUR.*) à 1832 (*cf.* ex. 15); **5.** « lieu par où l'on accède », ds *FUR.* 1701; attest. constantes et avec plus d'ext. dep. BESCH. 1845 « environs »; **6.** « action d'aborder à une côte », apparaît ds *Ac.* 1798 et constamment attesté. **D. —** « attaque par terre ou par mer ». Apparaît en 1636 (*cf.* étymol.). Bien attesté jusqu'à la fin du ^{xviii^e} s. (*cf.* ex. 23), encore attesté en 1837. Pas d'attest. post. **II. — Vitalité des sens. — A. —** Sens et accept. disparus : **1.** av. 1789 : néant; **2.** apr. 1789 : le sens B dans sa totalité; le sens C dans ses accept. 3 et 4; le sens D (sous réserve de certains emplois de loc.). **B. —** La vitalité des sens et accept. restants n'est rien en compar. de celle des loc. (notamment *d'abord*) où se manifeste surtout l'accept. C 1 (*cf.* sém.).

Cette situation paradoxale d'un verbe à double nominalisation semble indiquer que les deux substantifs en concurrence sont moins sémantiquement éloignés qu'il n'y paraît. Or, la (sur)vie du second (*abord*) pour désigner l'action de rentrer au port n'a pas excédé deux occurrences (voir ci-dessus), si l'on excepte quelques résurgences à la fin du 19^{ème} siècle. En réalité, la distinction de sens n'a pas eu lieu entre le terme concret et le terme métaphorique, mais plutôt entre les deux sens littéraux du verbe :

- *Abordage* se rapporte à l'action de se mettre au bord d'un rivage ou bord à bord avec un autre navire.
- *Abord* nominalise préférentiellement le fait d'être au bord du rivage ou d'avoir accès à une personne.

Ce qui les distingue manifestement est la différence entre l'action et l'état, le second ayant vu ses emplois s'étendre métaphoriquement à « l'action de débiter quelque chose ». Aspectuellement parlant, *l'abordage* est un climax¹⁰⁰ et on peut parler d'achèvement pour le caractériser ; *abord* renvoie à un état dans ses acceptions premières et possède un aspect ingressif¹⁰¹ pour son emploi métaphorique.

Cette non-concurrence sous forme de divergence lexicale est certainement le fait du hasard de l'évolution du français. Rien dans l'historique complexe que nous avons présenté ne semble expliquer un tel dégroupement, hormis peut-être les divergences aspectuelles que nous venons de décrire. Et par conséquent, rien ne justifie réellement de traiter les différents sens de ce verbe comme de véritables homonymes. En revanche, la distinction est d'importance pour l'avenir si nous décidons d'inclure les prédicats substantifs à nos modèles. En tant que déverbaux tous les deux, *abord* et *abordage* opèrent une opposition *de facto* entre les différents sens du verbe *aborder* auxquels ils se rapportent. La question est de savoir si nous associerons systématiquement les substantifs aux verbes dans le graphe ou s'ils feront l'objet d'entrées autonomes. Dans le premier cas, *aborder* ferait l'objet d'une séparation entre ses deux ensembles d'acceptions liés aux substantifs.

4.1.2 Forte polysémie

Dans notre terminologie, la polysémie d'un verbe est dite forte lorsqu'il est devenu difficile de faire un rapprochement conceptuel entre ses différentes acceptions, quels que fussent les liens historiques entre ces dernières. Il conviendrait dans ce cas

¹⁰⁰ C'est-à-dire le point culminant d'une action qu'on peut considérer comme achevée lorsqu'il est atteint.

¹⁰¹ L'aspect ingressif correspond à la période qui précède de peu ou initie le procès dénoté par le verbe (voir Cosériu 1979, Laca 2004).

de parler d'homonymie synchronique, cette nuance terminologique expliquant pourquoi la polysémie est « forte ».

Entre certains des divers sens du verbe *aborder*, il ne subsiste plus qu'un lien ténu qu'il est possible de capter par l'intuition, mais qu'il n'est apparemment plus possible de verbaliser par une primitive sémantique. Ainsi, *aborder un problème* et *aborder le/au rivage* nous paraissent avoir conservé une proximité sémantique sans que nous puissions la décrire autrement que par une paraphrase maladroite : « venir/arriver au bord de », paraphrase qui d'ailleurs ne résout pas le problème du décalage entre le sens littéral et la métaphore puisque *venir* et *arriver* peuvent tous deux être employés métaphoriquement. En l'occurrence, le bord d'un problème, d'un thème ou d'une question renvoie au commencement du traitement de ceux-ci, alors que le bord du rivage est le rivage lui-même¹⁰². D'autre part, la distinction entre les deux substantifs associés à *aborder* accentue encore davantage la sensation de distance entre les acceptions.

Conséquemment, nous conservons au verbe *aborder* le statut de terme polysémique, en tant qu'unité homographique de catégorie constante (verbe) et à origine commune. Toutefois, nous suggérons de le faire entrer dans la catégorie des verbes « à forte polysémie », sous réserve de la pertinence de créer cette dernière dans l'optique de construire des graphes. Le relevé des traductions de *aborder* en italien confirme partiellement ce point de vue en présentant des vedettes différentes pour les sens les plus éloignés comme nous allons le montrer maintenant.

4.1.3 Dégrouperment des acceptions

Nous allons illustrer notre propos initial au sujet de la dissymétrie par l'étude des traductions proposées pour le verbe *aborder* en italien. Cet exemple est le premier que nous avons commencé à traiter dans l'ordre chronologique, il y a six ans, et nous le réutilisons aujourd'hui car il illustre parfaitement le phénomène.

Afin de présenter clairement les deux lexiques et de les comparer efficacement, nous avons utilisé les définitions du verbe *aborder* tirées du dictionnaire Larousse de langue française¹⁰³, et nous les avons synthétisées :

- 1- **arriver au rivage** ([intransitif])
- 2- **approcher de** (quelqu'un (pour lui parler))¹⁰⁴

¹⁰² Cette différence a encore à voir avec l'aspect du verbe, achèvement dans un cas, ingressif dans l'autre.

¹⁰³ Nous n'avons pas utilisé ici les ressources du TLF. Le dégroupement effectué par ce dernier est d'une grande richesse, bien supérieure à ce que les autres dictionnaires monolingues ou bilingues sont en mesure de proposer. C'est donc avant tout par souci de simplification qu'un autre dictionnaire de référence a été choisi pour illustrer notre propos.

- 3a- **arriver à** (*un lieu*)
- 3b- **en venir à traiter** (*problème-question*)
- 3c-**commencer à étudier** (*matière*)
- 4a-**accoster** (*un navire* (pour l'assaillir))
- 4b-**heurter** (*un navire* (par accident))
- 5-**prendre** (*virage*)
- 6-**attaquer** (*un ennemi*)
- 7-*s'aborder* : **entrer en collision** (pour deux navires)

Nous ne conserverons pour la suite que les indices figurant au début des lignes précédentes.

4.2 Traductions

4.2.1 Illustration de la dissymétrie dans les lexiques français-italien

Voici maintenant nos deux entrées du verbe *aborder*, la première étant celle du dictionnaire Bordas français-italien et la deuxième étant celle du lexique inversé constitué à partir de la partie italien-français du même dictionnaire. Pour le LD comme pour le LI, les traductions ont été classées ci-dessous en fonction de l'acceptation d'*aborder* auxquelles elles correspondent :

Lexique droit (LD)

- 1-attraccare, approdare, sbarcare (exemple¹⁰⁵)
- 2-fermare (exemple)
- 3a-approdare (exemple)
- 3b-abbordare,affrontare (exemple)
- 3c
- 4¹⁰⁶-abbordare (exemple), arrembare

¹⁰⁴ La définition sous forme de paraphrase verbale est en gras, l'objet prototypique est proposé en italique, éventuellement accompagné d'une extension en écriture normale qui décrit la situation d'énonciation.

¹⁰⁵ Lorsque la traduction est accompagnée d'un syntagme ou d'une phrase exemple, nous le mentionnons avec la marque *exemple*. Quand aucun exemple n'est associé à une traduction, nous avons dû déduire par nous-même le rapport traductème-acceptation ; cela n'a pas été trop difficile pour cette entrée car *attraccare* et *approdare* sont proposés pour l'emploi intransitif du verbe (acceptation 1 uniquement), *entrare in collisione* n'est valable que pour la forme pronominale (acceptation 7) et *arrembare* est proposé pour l'acceptation 4 au même niveau que *abbordare* qui lui est muni d'une phrase exemple.

¹⁰⁶ Nous n'avons pas trouvé de distinction dans le dictionnaire bilingue entre les traductions des acceptations 4a et 4b, c'est pourquoi nous les avons fusionnées ici.

- 5-prendere (exemple)
- 6-attaccare (exemple)
- 7-entrare in collisione

Lexique inversé (LI)

- 1-accostare (*attraccare*¹⁰⁷), affiancare (v.**accostare**¹⁰⁸), approdare, arrembare
- 2-abbordare (*avvicinare*), accostare (*personne*), appressare (*deriv letter*¹⁰⁹ per **avvicinare**), appropinquare (*letter ant* per **avvicinare**), avvicinare (*una persona*)
- 3a
- 3b-abbordare (*affrontare*), accennare¹¹⁰, affrontare
- 3c-accostarsi a
- 4-abbordare, arrembare
- 5-abbordare
- 6¹¹¹
- 7

Nous allons maintenant effectuer deux représentations comparatives des deux lexiques afin d'en extraire les différences et les points communs.

Le premier graphique est constitué de trois colonnes, la première est la liste ordonnée des vedettes du LD, la seconde contient les indices des sens du verbe *aborder* en français tels qu'ils sont définis ci-dessus et la troisième contient la liste ordonnée des vedettes du LI. Le principe consiste à mettre en relation les différentes traductions du verbe *aborder* avec les acceptions auxquelles elles correspondent.

¹⁰⁷ *Attraccare* n'est pas explicitement traduit par *aborder* pour cette acception mais on le retrouve mentionné comme synonyme proche de *accostare* à l'entrée de ce dernier quand il est lui-même traduit par *aborder*. Dans les bases de données que nous avons finalement constituées, le verbe *attraccare* possède un statut particulier de « traduction indirecte » ou plus exactement de « synonyme en langue cible » de *aborder* via *accostare*.

¹⁰⁸ v.**accostare** signifie qu'on a un renvoi vers *accostare* depuis *affiancare*, celui-ci n'étant pas directement traduit par *aborder* à sa propre entrée.

¹⁰⁹ Marques de registre (*letter* pour littéraire, *ant* pour vieux)

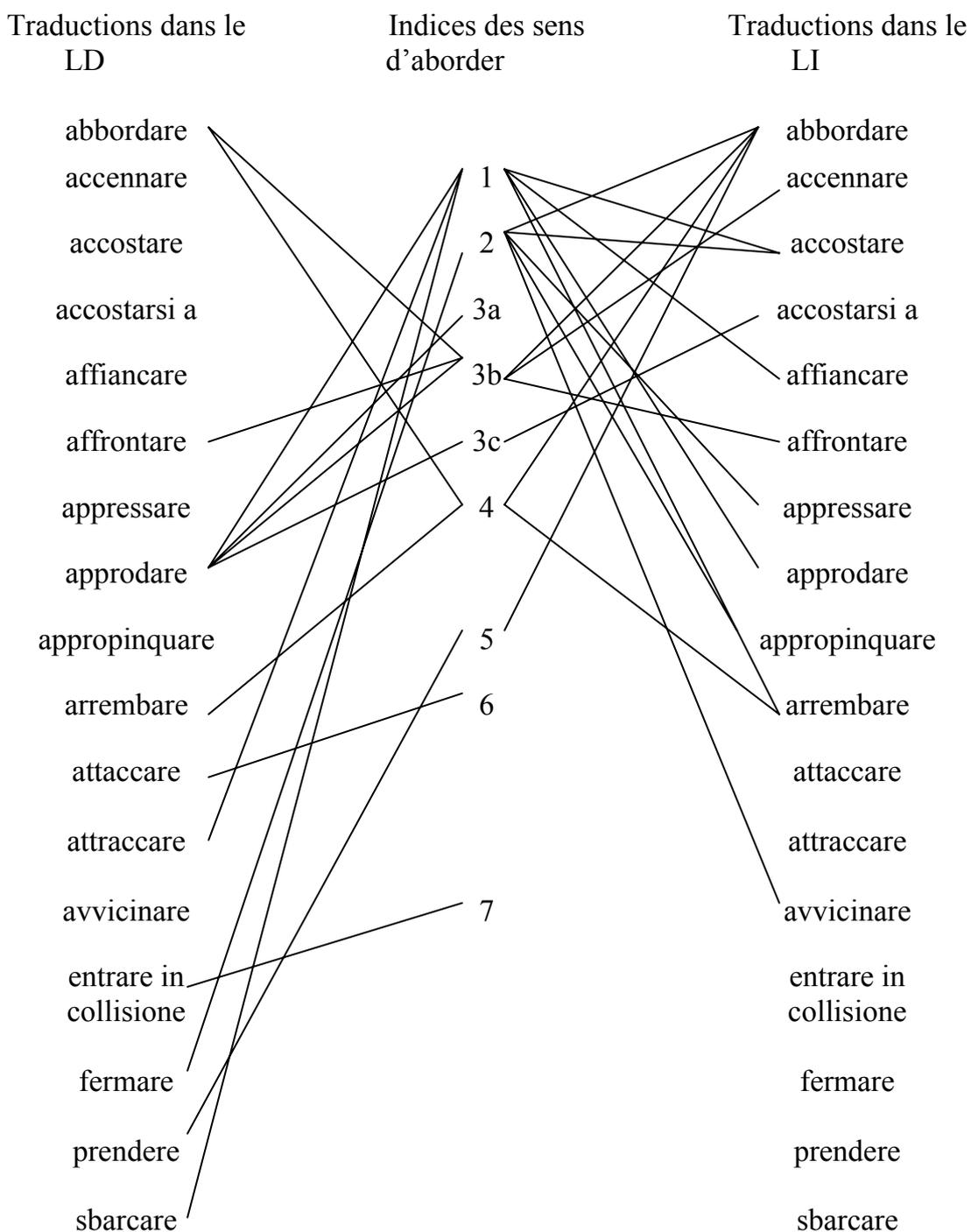
¹¹⁰ Dans certains cas, il faut penser aux influences des sens premiers du terme sur la traduction du mot-cible proposé. Ici, *accennare* en tant que transitif direct signifie *ébaucher* ou *esquisser*. C'est par analogie qu'il pourrait être traduit par *aborder* dans un contexte très précis.

¹¹¹ L'absence de verbes traduits par *aborder* dans certaines acceptions est caractéristique des décalages que nous avons observés dans les dictionnaires bilingues. La lacune la plus fréquente concerne les verbes pronominaux français, lesquels font trop rarement office de traductions dans la partie inversée du dictionnaire.

La seconde figure comporte à la base les mêmes informations que la première mais elle fait ressortir plus nettement la dissymétrie dont nous avons déjà parlé. Il s'agit d'un tableau à quatre colonnes où chaque ligne concerne un (et un seul) verbe italien. Les cases vides correspondent ainsi aux verbes qui n'apparaissent que dans un des deux lexiques. Plus exactement, les cases vides indiquent que le verbe apparaît dans l'autre lexique mais pas dans celui-là. La cinquième colonne a été établie à partir des traductions les plus proches du verbe '*aborder*' pour les verbes présents dans le LD et pas dans le LI.

Sous cette forme, le processus de recherche de symétrie n'est pas correctement rendu. C'est pourquoi nous proposons une autre représentation sous forme de tableau, en rétablissant en vis-à-vis, de part et d'autre des indices de sens, les verbes italiens identiques, qu'ils soient ou non illustrés par un ou plusieurs candidats à la traduction.

Figure 3 : Graphique récapitulatif, seconde version. Mise en exergue de la dissymétrie



Très nette dans ce dernier schéma, la dissymétrie globale entre le LD et le LI est également plus directement accessible dans le tableau suivant, lequel met en valeur les décalages par acceptions.

Figure 4 : Tableau récapitulatif du verbe *aborder*

Dico droit	Indice T1	Indice T2	Dico inversé	Traduction préférentielle ¹¹²	Traduction inversée du sens proche ¹¹³
<i>abbordare</i>	3b-4	2-3b-4-5	<i>abbordare</i>		
		3b	<i>accennare</i>		
		1-2	<i>accostare</i>		
		3c	<i>accostarsi a</i>		
		1	<i>affiancare</i>		
<i>affrontare</i>	3b	3b	<i>affrontare</i>		
		2(litt)	<i>appressare</i>		
<i>approdare</i>	1-3	1	<i>approdare</i>		
		2(litt,ant)	<i>appropinquare</i>		
<i>arrembare</i>	4	1-4	<i>arrembare</i>		
<i>attaccare</i>	6			<i>attaquer</i>	<i>attaccare</i>
<i>attraccare</i>	1			<i>accoster</i>	<i>attraccare</i>
		2	<i>avvicinare</i>		
<i>entrare in collisione</i>	7			entrer en collision	entrare in collisione ; scontrarsi
<i>fermare</i>	2			<i>arrêter</i>	<i>fermare</i>
<i>prendere</i>	5			<i>prendre</i>	<i>prendere</i>
<i>sbarcare</i>	1			<i>débarquer</i>	<i>sbarcare</i>

Quelques observations s'imposent :

- **Flagrante dissymétrie :** Ce travail comparatif nous permet d'observer une forte dissymétrie dans la représentation des vedettes, même en ignorant leur répartition par acception. Cinq vedettes du LD n'apparaissent pas dans le LI (*attaccare, attraccare, fermare, prendere, sbarcare*) et sept vedettes du LI ne sont pas présentes dans le LD (*accennare, accostare, accostarsi a, affiancare, appressare, appropinquare, avvicinare*). Seulement quatre vedettes (sur seize) sont communes aux deux lexiques (*abbordare, affrontare, approdare, arrembare*). À la lecture de nos résultats, il nous paraît désormais évident, voire logique, qu'une dissymétrie plus ou moins importante

¹¹² Lorsqu'un autre verbe que *aborder* traduit mieux que celui-ci le verbe italien placé en entrée, il est précisé dans cette colonne sous la mention « traduction préférentielle ».

¹¹³ Bien entendu, « la traduction inversée du sens proche » est dans la grande majorité des cas le verbe italien d'origine. Il faut comprendre par-là que la « traduction préférentielle » du verbe italien est préférable à *aborder* dans le contexte approprié.

est inévitable entre les deux parties d'un dictionnaire bilingue. Toutefois, l'ampleur du phénomène sur cet exemple dépasse nos attentes en la matière. C'est pourquoi il serait sans doute nécessaire de compléter cette analyse par la consultation systématique d'un dictionnaire monolingue italien (ou anglais ou espagnol pour d'autres cas) afin de déterminer la validité des hypothèses et des remarques induites par notre étude. La première hypothèse pertinente est la confirmation de la validité des quatre vedettes communes comme traductions d'*aborder*. En apparaissant deux fois, elles semblent constituer une espèce de « noyau dur » infrangible dans l'optique d'une fusion filtrée des deux lexiques.

- **De la dissymétrie dans la symétrie :** Nous notons une dissymétrie moins marquée concernant la répartition des sens pour les vedettes communes aux deux lexiques. Ainsi, le LI montre que *abbordare* est une traduction probablement plus complète du verbe *aborder* que ne le laisse apparaître le LD (cinq acceptions contre trois pour le second). Ceci tend à prouver qu'*abbordare* n'est pas systématiquement « prioritaire » pour toutes les acceptions du verbe français même s'il en est une traduction possible. Cette seconde observation vient un peu nuancer la première. Elle ne constitue pas pour autant une remise en cause assez pertinente du principe du noyau dur dans la mesure où cette dissymétrie n'est que partielle et offre plus de compléments d'informations que de contradictions. Ici encore, c'est le dictionnaire monolingue qui nous fournirait la meilleure réponse.
- **Dissymétrie cachée :** Il existe de grosses différences dans les traductions proposées pour chacun des sens du verbe *aborder*. Les acceptions 1 et 2 en particulier font apparaître les plus grosses différences en ce domaine (voir la première page de l'étude d'*aborder*). Notre troisième observation mérite quelques commentaires qui dépassent le cadre de cette présentation sommaire. Le relevé de la dissymétrie se passe de commentaires. À propos du sens 5 '*prendre (un virage)*' le dictionnaire nous offre *prendere* comme traduction exclusive. Or la traduction de *prendere* dans cette acception est *prendre* comme on pouvait s'y attendre. C'est *abbordare* qui figure dans le LI pour le sens 5 alors que le LD ne le mentionne pas. Dans ce cas précis, nous pensons que c'est le LI qui a raison car l'action d'aborder un virage (ou une courbe, un carrefour) comporte des traits sémantiques (prudence, danger) qui ne sont pas inclus dans la structure de traits de *prendre* ou de *prendere*, celui-ci étant plutôt à considérer comme une traduction générique non marquée sémantiquement. Pour

cette dernière raison, il est nécessaire d'augmenter la structure de traits aussi bien pour le LD, pour déterminer toutes les paraphrases du mot-source, que pour le LI au sein duquel on souhaite repérer une « hiérarchie de sélection » entre les différentes vedettes.

- **Un cas de symétrie très isolé :** Il est notable qu'une seule véritable symétrie est perceptible dans ce relevé concernant la traduction *affrontare*, qui est la seule dont la présence conjointe aux deux lexiques s'accompagne d'une répartition commune des sens qui lui sont attachés, du sens unique 3b en l'occurrence. Même en prenant en considération que la dissymétrie lexicographique est une donnée inévitable, cet unique cas de correspondance dans le miroir paraît bien isolé. Bien entendu, *affrontare* est confirmé comme vedette d'*aborder* en raison de son appartenance aux deux lexiques, le Droit et l'Inversé.
- **La paronymie dans la traduction : extension de sens ou « zèle » des rédacteurs ? :** Parmi toutes les traductions, nous relevons une incohérence apparente avec la présence de *sbarcare* comme traduction possible du verbe *aborder* au sens 1 alors que le LI ne propose que la traduction proche *débarquer* (*après avoir abordé*¹¹⁴). Or, il semble que le LD ait négligé ce trait en ne mentionnant pas que *sbarcare* ne s'emploie pas pour la seule action de *toucher au rivage*. Nous parlons d'incohérence, voire d'erreur, dans la mesure où l'abordage est un présupposé de l'action de débarquer mais n'en est pas le synonyme. Cette remarque s'inscrit dans une réflexion plus générale sur les limitations nécessaires imposées *de facto* aux dictionnaires concernant le format et la somme d'informations qu'on peut y mettre. Il est possible que *sbarcare* ait été mentionné comme extension du sens d'*aborder* pour montrer que l'action de toucher au rivage pour un navire s'accompagne généralement de l'action de débarquer pour les personnes qui étaient à son bord. Si tel est le cas, le dictionnaire devrait mentionner ce fait en précisant qu'il s'agit d'un emploi métonymique avec extension du sens par implication (l'abordage implique le débarquement). En réalité, il est acceptable de considérer l'abordage comme une « conséquence » implicite du sens premier de *sbarcare*. On pénètre ainsi dans le périlleux territoire de la sémantique pragmatique qui est rarement compatible avec le format très synthétique d'un dictionnaire.

¹¹⁴ Trait ajouté par nous, non apparent dans le dictionnaire

- **Dissymétrie dans les acceptions disponibles :** Chaque partie du dictionnaire bilingue ne donne pas de liste de traductions exhaustive. D'autre part, de nouveaux sens émergent par rapport aux dictionnaires monolingues. Remarquons la disparition ou l'apparition de sens possible du verbe *aborder* en fonction des vedettes proposées. Le LI ne propose aucune traduction pour le sens 6 (*attaquer (un ennemi)*) ni le LD pour le sens 3c (*commencer à traiter (une nouvelle matière)*). Les sens 3c et 6 d'*aborder* ne figurent pas (explicitement) dans le dictionnaire français Larousse sur lequel nous nous sommes appuyé. Ils n'en sont pas moins pris en compte par le dictionnaire français / italien. Comme il s'agit apparemment de sens faibles, il n'est pas surprenant de constater qu'ils n'apparaissent que dans un seul lexique. L'entrée d'*attaccare* dans le LD italien / français a en vedettes les verbes *attaquer* ou *affronter* qui sont des traductions plus littérales et plus courantes qu'*aborder* qui est résiduel dans cette acception. Le sens 3c est plus intimement lié à d'autres acceptions du verbe *aborder*, il fait donc moins office de bruit dans la liste des traductions proposées.
- **Vaste paradigme de traduction :** Nous pouvons être surpris par le grand nombre de candidats recensés une fois le LD et le LI unifiés. L'union des deux lexiques nous en offre seize différents (dont neuf pour le LD, onze pour le LI et quatre pour l'intersection des deux). L'aspect comptable de ce travail ne doit pas nous amener à une vision accumulative ni restrictive de la lexicographie bilingue. Ni l'union ni l'intersection des deux lexiques ne nous offrent un résultat satisfaisant dans la perspective d'une réécriture de l'entrée d'*aborder*. Pour optimiser celle-ci, nous envisageons de réaliser une 'fusion filtrée' issue de nos remarques, de la prise en compte de facteurs plus fins que ceux apparents (degrés sémantiques, registres de langue, cohérence, complémentation, etc.) et de la consultation d'un lexique italien que nous mentionnons plus haut. Toutefois, c'est d'abord l'union des lexiques que nous exploiterons dans les CSMU.

5. De la polysémie faible à l'explosion combinatoire : le verbe abandonner

5.1 Polysémie faible

5.1.1 Un fort lien sémantique entre acceptions

Pour caractériser le degré de polysémie des verbes d'une langue, il est tentant d'explorer la représentation par primitives conceptuelles et de vérifier dans quelle mesure les différents sens de chaque verbe peuvent être regroupés sous la même primitive. Si nous parvenons à esquisser cette dernière, le verbe analysé pourra être considéré comme faiblement polysémique car tous ses sens seront corrélés, à condition que les traductions le confirment par leur nombre important et la finesse des critères qui permettent de les distinguer les uns des autres. Ceci se rapproche de la notion de polysémie verticale dans la mesure où la primitive englobe toutes les acceptions, comme nous l'évoquions au Chapitre I.

L'objectif est de créer un label formalisé et englobant de classes de prédicats. Pour le champ sémantique de l'abandon sur lequel nous avons travaillé, la primitive utilisée pourrait être :

- Il existe x , y et z tels que : x cause (ou fait), volontairement ou sous la contrainte, que n'existe plus (ou cesse) un lien privilégié z entre lui-même et un objet y ¹¹⁵, ce dernier étant exprimé ou non au moyen du langage¹¹⁶. Le résultat est l'arrêt du processus y (quand y est un processus) ou l'état d'abandon de y (quand y n'est pas un processus).

Une fois formalisé, nous aurions :

- $\exists x, \exists y, \exists z, \text{Exister_lien}(z,x,y)$ [état initial], $\text{Causer}(x,-z)$ [processus], $-z$ [résultat]¹¹⁷.

¹¹⁵ Objet est à prendre ici au sens syntaxique

¹¹⁶ L'argument en position objet peut faire l'objet d'une ellipse comme dans les exemples suivants : « J'abandonne » (la partie, je n'en peux plus), « je renonce » (à t'expliquer, tu es vraiment trop bouché), « je laisse » (les enchères monter sans moi). On peut noter que les différents prédicats du champ n'ont pas le même comportement face à l'ellipse ; si celle-ci paraît naturelle en contexte pour *abandonner* ou *renoncer*, elle est déjà moins immédiate pour *laisser*, *quitter* hormis dans des emplois très précis (*jeu* pour le premier, *session informatique* pour le second) et devient même douteuse pour *cesser*, même en emploi absolu : ?? « Je cesse. »

¹¹⁷ Formalisme auquel il conviendrait d'ajouter des marqueurs pour exprimer l'intensité du lien, sa nature ainsi que les notions de volonté du sujet ou d'expression langagière de l'objet (i.e. ellipse ou pas ellipse).

En primitive lexicalisée, cela pourrait donner « causer que n'existe plus un lien entre le sujet et l'objet ».(4)

En ce qui concerne le verbe *abandonner*, les divergences de sens ne sont pas à rechercher seulement dans les acceptions elles-mêmes, mais également dans l'environnement langagier du prédicat en contexte.

Nous considérons que les primitives sont beaucoup moins sensibles à l'inadéquation interlangue que les unités lexicales. Par conséquent, il nous paraît possible de les écrire dans une langue donnée et de les appliquer à toutes les autres langues.

5.1.2 Pas de dégroupement de substantifs

Abandonner admet un et un seul substantif dans sa forme déverbale, à savoir *abandon*. Ce dernier est utilisable pour toutes les acceptions du verbe, y compris les formes pronominales et les emplois métaphoriques.

Si le noyau de sens est commun à toutes les acceptions, les nuances sont à rechercher prioritairement dans les collocations. Le dégroupement que nous présentons dans le tableau récapitulatif est entièrement motivé par les objets (tous directs) que le verbe prend en fonction des traductions proposées.

Vu sous cet angle, *abandonner* est circonscrit à une seule acception principale, que nous résumons par la primitive élaborée au sous-chapitre précédent.

5.2 Un champ lexical large

5.2.1 Constitution du champ lexical de *abandonner* en français

5.2.1.1 Méthode d'extraction des synonymes et quasi-synonymes

La récupération des données lexicographiques (TLF, Robert, fusion des dictionnaires de synonymes, dictionnaires monolingues et bilingues, etc.) nous donne les paradigmes lexicaux suivants pour le champ sémantique de l'abandon (prédicats verbaux uniquement).

La liste suivante est globalement issue de la fusion des dictionnaires de synonymes de l'INALF exploitée par ailleurs :

abdiquer, abjurer, abolir, accorder, adieu, aliéner, apostasier, balancer, bazarder, caler, caner, capituler, céder, cesser, concéder, confier, déguerpir, délaisser, déloger, démissionner, démordre, déménager, déposer, dépouiller, désertier, dételer, détourner, disparaître, donner, enterrer, évacuer, exposer (anc), finir, flancher, fléchir, fuir, immoler, interrompre, jeter, lâcher, laisser, larguer, léguer, liquider, livrer, mourir, négliger, omettre, oublier, partir,

plaquer, plier, poser, quitter, rabattre, raccrocher, reculer, rejeter, relâcher, reléguer, remettre, renier, renoncer, repousser, répudier, résigner, résilier, rompre, s'éloigner, s'enfuir, s'incliner, sacrifier, se débarrasser, se défaire, se démettre, se départir, se déposséder, se dépouiller, se désintéresser de, se désister, se désolidariser, se dessaisir, se détacher, se détacher de, se livrer, se rendre, se replier, se retirer, se séparer, se séparer de, se soumettre, semer, sortir, succomber, supprimer, suspendre, tomber, vider, violer, baisser les bras, baisser pavillon, battre en retraite, couper là, déclarer forfait, faire cesser, faire la part du feu, fausser compagnie, jeter le manche après la cognée, lâcher pied, laisser aller à vau-l'eau, laisser aller, laisser choir, laisser courir, laisser de côté, laisser périliter, laisser tomber, laisser traîner, mettre bas, mettre les pouces, passer la main, planter là, plier bagage, renier sa foi, s'avouer vaincu, s'en aller, se laisser déposséder de, tourner le dos à, vider la place, vider les lieux

Nous notons en premier lieu la présence éventuelle de verbes polylexicaux au milieu des unités monolexicales. En effet, les dictionnaires de synonymes qui ont servi de base au dictionnaire fusionné ne font pas de différence entre verbes simples, composés ou locutionnels, ce qui est conforme à notre vision du lexique et à notre souci de ne laisser aucun mot hors de notre étude.

5.2.2 Sélection des synonymes et quasi-synonymes de abandonner en anglais

5.2.2.1 Méthode d'extraction

Comme pour la sélection des synonymes en langue source, le recensement des synonymes et quasi-synonymes en langue cible ne saurait être récupérable par la seule intuition, ni d'ailleurs par l'observation, même approfondie, des dictionnaires bilingues. Les Lexiques Inversés que nous défendons proposent, en mode fusion des données, une méthodologie d'extraction des synonymes et quasi-synonymes non directement perceptibles par la lecture des dictionnaires. Toutefois, il ne s'agit que d'une confrontation au sein d'un seul et même dictionnaire, non d'une liste exhaustive de tous les verbes corrélés en langue cible. Il est amusant de constater que bien que la dissymétrie soit un phénomène commun à tous les dictionnaires, elle ne se manifeste pas de la même façon selon les éditeurs.

Dans l'optique de proposer un champ lexical restreint au-delà duquel un terme de quelque langue que ce soit n'est plus considéré par nous comme connexe de *abandonner*, nous nous sommes appuyé sur les travaux de Sabine Ploux qui récupèrent les équivalents en anglais du terme d'origine à partir du même procédé que celui qui régit la fusion des dictionnaires des synonymes en français, sans plus de précision méthodologique néanmoins :

abandon, abdicate, abjure, abnegate, abolish, abort, abrogate, accept, admit, alienate, annul, apostatize, avert, award, be lost, break, break off, bring to an end, capitulate, cast aside, cast off, cease, change, chicken out, chuck away, chuck out, clear off, come off, come to a , alt/standstill, come to an end, concede, confide, denude, depose, deposit, deprive, desert, desist from, despoil, die, die down, die out, disappear, discontinue, dissuade, ditch, divert, divest, do away with, donate, drop, end, entrust, evaporate, fade, fail, fall, flush out, forsake, funk it, get rid of, give, give away, give in, give the push/boot, give to, give up, give way, go missing, grant, hand in one's office, jettison, kick out, knock off, lay aside, lay bare, leave, leave off, leave to, let have, let take, lift, make concessions, make over, move, move out, neglect, not finish, ooze away, opt out, outgrow, pass, plunder, put a stop to, put an end to, put off, quit, recant, recede, relinquish, remove, renounce, resign, retire from, scarper, set down, shed, shift, slip away, sneak away/out, stop, strip, subside, surrender, swallow up, take flight, take to, throw out, transfer, turn, turn away, turn out, vanish, waive, ward off, withdraw from, yield, yield up

Les synonymes et quasi-synonymes de *abandonner* en langue source et en langue cible ont leur propre paradigme de synonymes et quasi-synonymes, comme nous l'avons précédemment indiqué. On y retrouve les verbes proposés par Sabine Ploux ainsi que certains autres ajoutés grâce au relevé des traductions des synonymes d'*abandonner*.

Tous ne sont d'ailleurs pas pourvus de traductèmes du champ d'*abandonner* car la source de nos données bilingues diffèrent de celles qui ont servi à construire la liste des verbes anglais à l'ISC. Une partie de la base de données préparatoire à la construction du graphe d'*abandonner* est théoriquement présente en annexe.

5.3 Les traductions d'*abandonner*

5.3.1 Le LI anglais comme complément riche au LD

Avec les collocations comme facteurs discriminants principaux (une caractéristique avantageuse des dictionnaires Robert & Collins), nous obtenons une répartition des traductèmes beaucoup plus fine après prise en compte du LI. Le tableau de la page suivante met en avant le phénomène de dissymétrie de la même façon que ce que nous avons montré pour *aborder*. Sa structure en est tout de même très différente puisque nous ne nous appuyons non plus sur les acceptions du verbe mais

sur les collocations proposées par le dictionnaire, au nom de l'acceptation du phénomène de faible polysémie accepté pour *abandonner*.

Les objets de la colonne « paraphrase » correspondent le plus fidèlement possible aux collocations suggérées pour chaque acception de l'entrée du verbe *abandonner*.

La question de la procédure peut être posée au sujet de la constitution de ce tableau. Nous sommes obligé de reconnaître que c'est seulement notre propre intelligence synthétique qui est à l'origine de cette vision restructurative des traductions du verbe *abandonner*. Nous avons cependant fait attention à ne pas dénaturer ni surinterpréter le matériau lexicographique ce qui ne nous a pas empêché de réunir « de force » des collocations différentes mais apparentées.

Concernant l'aspect « algorithmisable » de la conception de ce tableau, nous pouvons effectivement dire qu'il est absent en l'occurrence. Toutefois, l'algorithme que nous avons écrit à l'issue de ce travail mène à un résultat restructuratif très semblable à ce qui est présenté ci-dessous. D'une certaine façon, nous pouvons dire que cette restructuration naïve, réalisée bien avant notre algorithme, montre *grosso modo* la même chose que notre programme de construction de graphes, mais de manière lisible pour un être humain peu formé à la programmation informatique.

Figure 5 : Tableau des correspondants anglais du verbe ABANDONNER : première représentation du lexique inversé français-anglais

TRADUCTION DROITE	PARAPHRASE	TRADUCTION INVERSEE
desert , abandon	désertier un lieu	desert
leave, abandon	(qqn) abandonner (qqn)	abandon , give up, desert, forsake
fail, desert , forsake	(qualité, partie de qqn) abandonner (qqn)	break, ooze away, desert
desert, abandon , forsake	abandonner intentionnellement	abandon , quit, give up
abandon, give up	abandonner (technique, appareil)	cast aside, give up
abandon, desert	abandonner (positions, poste)	desert , relinquish
take flight, give up	abandonner le terrain	
give up, relinquish	se retirer de ses fonctions	
give up , abandon	abandonner (études, recherches, travail)	quit, give up
give up , relinquish, renounce	abandonner (droits, privilèges, avantages)	cast aside, waive, give up , renounce
give up , withdraw from, retire from	abandonner (course, lutte, partie)	quit, give up , retire hurt (sur blessure)
give up , abandon	abandonner (projet, espoir, hypothèse)	outgrow [en prenant de l'âge], give up , jettison, relinquish
retire from, abandon, leave	abandonner (le pouvoir)	relinquish
give up	abandonner	quit, opt out, give in, give up , withdraw from the game
	abandonner (mission, opération)	abort [pour raisons de sécurité]
	abandonner (poursuite, soupçons, scrupules)	lay aside
	abandonner (forteresse, territoire)	yield
	abandonner (opinion, mode de vie)	outgrow
	abandonner (biens, propriété)	relinquish
	abandonner (condition, limite d'âge)	waive
	abandonner (vêtements)	leave off
	abandonner (l'étalon-or)	come off
	abandonner (le navire)	take to
give to, leave to	abandonner / laisser à	
let (so) take one's hand	abandonner sa main à	
leave it up to...to	abandonner le soin à...de	
leave / abandon to one's fate	abandonner à son sort	leave to one's fate
give / abandon one's up to	abandonner son corps à	yield up

leave to (be)	abandonner à (un désastre)	
---------------	----------------------------	--

Le tableau ci-dessus fait d'abord apparaître un phénomène inhérent à la faible polysémie du verbe *abandonner* : comme nous l'avons déjà mentionné, les différentes traductions se répartissent en fonction des objets sous-catégorisés par le verbe et pas nécessairement par le dégroupement de ses acceptions, que nous avons finalement ignorées. Si l'on adopte une vision restructurative à partir de ces données, il est possible de restituer une structure de l'entrée *abandonner* du point de vue de ses traductions. Autrement dit, le dégroupement des acceptions devient identique au dégroupement de ses sous-paradigmes de traductions répartis par collocations.

5.3.2 Même les langues romanes...

Alors qu'on pouvait s'attendre à une distributionnalité bien moindre avec les langues romanes, en l'occurrence l'espagnol et l'italien, il se trouve que les traductions de *abandonner* dans ces deux langues sont également présentes en nombre conséquent. Contrairement au R & B, les dictionnaires Bordas et Larousse que nous avons utilisés donnent peu d'informations sur les collocations associées, ce qui ne permet pas a priori de classer les données dans un tableau comme nous avons pu le faire pour les traductions anglaises.

ABANDONNER (ESP) = *abandonar-abdicar (syn)-aborrecer-aburrir-ahorcar-arrimar-arrinconar-ceder-ciar-ciar en-dejar-dejar de la mano de Dios-dejar desamparado-derrelinquir-desamparar-desasistir-desertar de-deshabitar-deshacerse de-desmantelar-desocupar-renunciar a-separarse de*

ABANDONNER (ITA)¹¹⁸ = *abbandonare-buttare a mare-desistere da-dimettere-lasciare-recedere da-non reggere*

Néanmoins, ces traductions fort nombreuses pour un verbe à faible polysémie ne sont que moyennement surprenantes eu égard à la grande diversité des synonymes et quasi-synonymes du verbe *abandonner* pour le français seul. Intuitivement, nous pouvons penser que les paradigmes des synonymes et quasi-synonymes des verbes espagnols et italiens susceptibles de traduire *abandonner* sont aussi riches dans leurs langues respectives.

C'est notamment à partir de l'observation de ces paradigmes pour des mots pourtant peu dégroupés en acceptions que nous est venue l'idée de construire des bases de données implantables qui rendent compte de leur grande richesse.

¹¹⁸ Le LI français-italien n'étant pas achevé à ce jour, cette liste est proposée sous réserve d'enrichissement ultérieur. En revanche, la liste des verbes espagnols est en principe complète.

6. L'intégration des locutions et périphrases verbales dans le champ de l'étude

6.1 Intégration des périphrases verbales dans les champs lexicaux : exemple à partir de *se mettre à* en français.

6.1.1 Arguments en faveur d'une intégration

Dans les représentations lexicographiques habituelles, les périphrases ne font jamais l'objet d'un traitement complètement séparé. Dans un même ordre d'idées, dans le dictionnaire d'Alain REY et Sophie CHANTREAU (1988) dévolu aux locutions du français, les entrées sont exclusivement des mots simples ou dérivés¹¹⁹, chaque expression et locution étant placée derrière l'entrée monolexicale la plus « représentative ».

Au-delà de la question du degré de figement ou d'opacité sémantique des locutions, nous considérons qu'une expression polylexicale a tout autant sa place dans un réseau lexical que les mots simples, pour les raisons suivantes :

- Les définitions lexicographiques sont très souvent de nature périphrastique, ce qui peut apparaître comme un argument faible dans la mesure où il y a fort peu de synonymie parfaite entre unités monolexicales d'une langue donnée. Toutefois, les périphrases utilisées comme définitions peuvent elles-mêmes être considérées comme des synonymes polylexicaux des mots définis, et dès lors acquérir un statut lexical établi.
- Une périphrase verbale possède un sens plein qui renvoie à une univocité sémantique du même type que celle des verbes « à mot unique », avec parfois une dimension aspectuelle supplémentaire. Ainsi, *se mettre à* peut être considéré et défini comme un mot à part entière, apparenté à *commencer* avec une nuance qui peut être *enfin*, *à contre-cœur*, *avec entrain*, etc., selon le contexte d'énonciation.
- Dans un exercice de traduction, l'homogénéité sémantique des périphrases verbales et la non biunivocité interlangue sont bien mises en valeur par le passage d'une langue à une autre. En d'autres termes, une périphrase en langue source est très souvent traduite par une ou

¹¹⁹ La terminologie employée est celle de Gaston Gross dans son ouvrage de 1996, *les expressions figées en français*. Nous distinguons ainsi les mots monolexicaux, dits « simples » s'ils sont constitués d'un seul morphème ou « dérivés » dans les autres cas, des mots « polylexicaux » ou « complexes » qui sont des « unités composées de deux ou plusieurs mots simples et/ou dérivés ».

plusieurs unité(s) monolexicale(s) en langue cible. Ce phénomène n'est ni systématique ni exceptionnel, les Lexiques Inversés étant remplis de locutions et de périphrases impossibles à ignorer et terriblement difficiles à classifier.

- La fréquence d'apparition des périphrases verbales dans le discours est particulièrement élevée et ne saurait être négligée. Cependant, ce dernier point n'exclut pas complètement d'avoir une vision compositionnelle des périphrases.
- Le dernier argument que nous avancerons est que les périphrases verbales sont tout autant sujettes à la polysémie que les verbes monolexicaux, ce que les dictionnaires monolingues mettent peu en exergue. En outre, le nombre de traductions proposées pour certaines d'entre elles peut être très élevé, comme en atteste le relevé que nous avons fait pour *se mettre à*.

Bien entendu, ces arguments en faveur de leur intégration ne peuvent faire ignorer que ces périphrases ne sont pas autonomes, i.e. elles ne peuvent pas constituer un procès à elles seules. Le cas de *se mettre à* est d'ailleurs très clair, il s'agit d'une périphrase aspectuelle qui fait porter un aspect ingressif ou inceptif¹²⁰ sur le verbe véhiculant le procès. Dans tous ses emplois périphrastiques, *se mettre à* n'est donc jamais un verbe autonome mais simplement un auxiliaire aspectuel du verbe.

Malgré cela, si l'ingressif se manifestait en français et dans les autres langues par un seul auxiliaire ou un seul morphème¹²¹, nous n'aurions pas à nous poser la question de savoir comment s'organisent leurs traductions puisque nous n'aurions pas besoin de faire le choix du meilleur candidat. Or, le paradigme de traductions des périphrases est souvent bien fourni, comme nous allons le voir pour *se mettre à*.

Pour autant, nous ne passerons pas sous silence les emplois aspectuels de *se mettre à* devant un nom, par exemple *se mettre au basket/dessin/macramé, etc.* Selon nous, ces expressions sont le résultat de l'ellipse d'un verbe support (le plus souvent *faire*) devenu superflu pour la bonne compréhension de l'énoncé. D'autre part, nous ne

¹²⁰ Il existe des nuances subtiles entre les deux aspects selon l'endroit où on se place sur l'axe aspectuel. Cosériu (1979) place *se mettre à* en position d'ingressif, c'est-à-dire qui précède l'action (quand on se met à travailler, on ne travaille pas encore) alors que Laca (2004) le voit plutôt semble-t-il en inceptif, i.e. au début de l'action (quand on se met à travailler, on travaille déjà). Les deux visions semblent parfaitement défendables selon le contexte.

¹²¹ Cosériu (1979) et Laca (2004) ne proposent que la périphrase espagnole *ponerse a* comme traduction de *se mettre à*. Mais il est vrai que leurs articles respectifs traitent de la caractérisation de l'aspect dans les langues romanes et non pas du potentiel de traduction de chaque langue.

parlerons pas dans ce chapitre des inclusions de *se mettre à* dans certaines locutions figées comme *se mettre à table*, cette dernière étant d'ailleurs potentiellement ambiguë. Nous traitons ces dernières locutions comme des lexies à part entière qui apparaissent littéralement comme entrées dans nos bases, et non pas comme sous-entrées de *se mettre à* pas plus que nous ne traiterons cette dernière comme sous-entrée de *mettre*.

6.2 Travail préliminaire sur *se mettre à*

Se mettre à est une périphrase verbale remarquable parmi les plus productives en termes de polysémie.

Cette caractéristique se retrouve dans le paradigme de traductions établi à partir de nos algorithmes personnels. Dans les ensembles ci-dessous, nous n'avons pas pris en compte les « traductions indirectes » qui sont difficiles à répertorier et pour lesquelles l'algorithme que nous utilisons est beaucoup plus complexe que pour les deux autres paradigmes de traductèmes. Les « traductions indirectes » ne sont pas explicitement accessibles dans les dictionnaires, i.e. ce ne sont ni des « traductions accessibles » ni des « traductions cachées » (voir page suivante). Une traduction indirecte est la résultante d'un chemin d'une profondeur d'au moins 2 dans notre graphe lexical à l'exception des chemins qui reviennent au point de départ. Leur recensement passe obligatoirement par une validation de l'utilisateur.

D'autre part, nous n'avons que peu pris en compte les différences sémantiques subtiles qui existent entre les extensions verbales et nominales de la périphrase (par exemple *se mettre au travail* vs *se mettre à travailler*, les implications aspectuelles n'étant pas forcément les mêmes dans les deux cas).

6.2.1 Traductions accessibles

Les « traductions accessibles » sont les traductions proposées à l'entrée lexicale *se mettre à*, laquelle est classée dans le dictionnaire sous l'entrée *mettre*.

- Anglais : *start +ing ; start to ; set to ; set about ; get down ; get on with ; begin to ; take to ; take up*
- Espagnol : *romper a ; entrar en ; echarse a ; ponerse en ; ponerse ; ponerse a*

6.2.2 Traductions cachées

Les « traductions cachées » sont d'abord les entrées lexicales pour lesquelles *se mettre à* est proposé comme traduction dans la partie « bis » des dictionnaires bilingues. Elles servent à alimenter les LI. Nous plaçons également sous l'appellation

« traductions cachées » les termes qui sont suggérés comme équivalents des deux termes traduits sans qu'il soit nécessaire que le premier traduise ou soit traduit par l'un des deux autres. De façon plus synthétique, cela correspond à nos liens de synonymie SPS et SPTC définis au chapitre IV.

- Anglais : *address os to ; begin to ; begin ; break forth into ; break into ; burst into ; burst out ; get ; get at ; get down to ; come on ; tackle ; take to ; take up ; get going on/with ; fall ; fall about ; go about ; go at ; start ; start to ; set to ; set about*
- Espagnol : *apretar a ; arrancarse a ; dar por ; darse a ; largarse ; echar a ; echarse a ; apartar a ; ponerse en*

6.2.3 Bilan du recensement

6.2.3.1 Première vue

Les traductions que nous présentons ci-dessus ne concernent que les emplois “pleins” de la périphrase, i.e. hors de toute locution plus étendue et plus figée de *se mettre* et ses extensions (par exemple *se mettre à table, se mettre sur son trente et un*). Par surcroît, elles font toutes référence à l'emploi aspectuel de *se mettre à* dont la paraphrase approximative est *être dans un état proche de commencer quelque chose* ou *être dans la situation d'avoir déjà commencé ladite chose*, toujours ingressif ou inceptif.

Malgré ces restrictions, le bilan quantitatif de notre recensement effectué sur un seul dictionnaire bilingue à chaque fois est le suivant :

- pour l'anglais, vingt-cinq traductions au total, neuf accessibles, vingt-trois cachées et sept communes ;
- pour l'espagnol, treize traductions, six accessibles, neuf cachées et deux communes.

Le constat le plus flagrant, en dehors de la grande variété des traductions, est que les traductions communes sont largement minoritaires dans le recensement final, ce qui est le cas pour la totalité des verbes hautement productifs de notre base, comme l'ont déjà illustré les études faites au sujet des verbes *aborder, abandonner* et *traverser*.

En raison de la stabilité du sens de la périphrase en dehors de certaines locutions plus figées, nous n'avons pas voulu confronter les paradigmes de traductions dans un tableau à entrées multiples comme nous l'avons fait pour les verbes précédents. Il faut noter que pour ces derniers nous n'avons jamais employé deux fois

le même mode de représentation, beaucoup moins par coquetterie qu'à cause de la façon dont leurs sens sont dégroupés en fonction de leurs équivalents dans les langues cibles.

Partant de ce constat, il nous a paru évident qu'une représentation des données verbe par verbe était tout bonnement impossible et de toute façon sans intérêt, hors l'analyse linguistique qui va avec. C'est pourquoi nous allons en arriver incessamment à une réflexion globale sur une méthodologie qui nous permettra d'intégrer tous les verbes dans un espace commun avec des critères discriminants partagés par tous.

6.2.3.2 *se mettre* et ses autres emplois

Nous rappelons également l'existence de nombreuses autres occurrences de la périphrase *se mettre à* dans des locutions verbales plus ou moins figées. Bien que nous ne les fassions pas entrer dans notre recensement préliminaire, elles n'en sont pas ignorées pour autant ; la difficulté demeure de déterminer à partir de quel degré de figement une locution peut constituer une lexie à part entière¹²² dans nos champs lexicaux. Parmi les expressions figées bâties à partir de *se mettre*, nous relevons *se mettre à table*, *se mettre sur son trente et un*, *se mettre en quatre*, *se mettre à dos*, *se mettre en boule*, etc.

Faute de pouvoir découper sur le mode compositionnel bon nombre des traductions de ces locutions en langues cibles, la solution de les considérer comme lexies est inévitable bien qu'elle ne soit pas satisfaisante dans tous les cas. Par exemple *se mettre en colère* n'est pas précisément figée et peut avoir une lecture compositionnelle mais se traduira alternativement par une construction avec périphrase (par exemple en espagnol *montar en colera*) ou un verbe monolexical (*enojarse*, *encolerizarse*). En raison du caractère non découppable de ce dernier, il est certainement plus sage de les considérer comme lexies ou bien, comme nous l'envisageons réellement, d'en faire une double écriture en acceptant le flottement du degré de figement (deux entrées, une avec *se mettre en + objet*, une avec *se mettre en colère* littéralement).

6.2.3.3 Pré-exploitation des résultats

A partir de nos 38 traductèmes anglais et espagnols, il a été possible d'établir un tableau récapitulatif sur le modèle de celui que nous présentons dans ce chapitre pour le verbe *abandonner*. Toutefois, ce mode de représentation partielle ne rend pas clairement compte du « positionnement » de *se mettre à* en français, i.e. quelles sont ses relations synonymiques et parasynonymiques avec ses équivalents en langue source ? D'autre part, l'étude de ces derniers doit nous permettre, si elle s'intègre effectivement au champ envisagé, de faire apparaître un certain nombre de traductions indirectes que la représentation lexicographique ne fait pas ressortir.

¹²² Sur les lexies et le figement, voir notre réflexion au chapitre I

Aussi, à ce stade de notre étude, une réflexion plus profonde sur l'utilisabilité des LI est nécessaire. Nous ne pouvons nous contenter durablement d'un simple recensement des traductions cachées de chaque verbe dans le seul but d'enrichir quantitativement les bases lexicographiques. Ceci est d'autant plus vrai que chaque traduction d'un mot est porteuse d'informations qu'elle ne partage jamais intégralement (ou très rarement) avec les autres traductions du même mot. Les critères distinctifs, de type syntaxique, sémantique ou pragmatique pour ne citer que les plus pertinents, sont en outre relativement aisés à caractériser ; c'est pourquoi il serait préjudiciable de ne pas les faire apparaître dans une base de données digne de ce nom et apte à permettre l'enrichissement « automatique » des dictionnaires déjà disponibles. Au nom de ces réflexions émergentes, nous avons élaboré une méthodologie de représentation des données qui mène à des paradigmes lexicaux plus lisibles et plus exploitables que les LI primitifs. Le résultat obtenu est ce que nous appelons les Champs Sémantiques Multilingues Unifiés (CSMU), des champs protéiformes qui s'adaptent à tous les critères et à toutes les langues nécessaires.

Afin de compléter nos données initiales et de préparer nos nouvelles bases de données, nous utilisons la technique du « balayage » explicitée ci-après. Cette méthode est utilisable pour toute construction d'un sous-champ d'un réseau global. Celui-ci est implantable par l'intermédiaire de l'algorithme présenté au chapitre IV ; la technique du balayage est indépendante de ce dernier et ne prétend pas en remplacer l'usage. Elle se contente de l'annoncer.

6.3 Intégration par technique de « balayage »

L'intégration des données que nous appelons « technique de balayage » s'effectue de la façon suivante :

- Première étape : Chacune des traductions françaises des traductions accessibles et cachées de *se mettre à* est recensée. Un lien de synonymie plus ou moins direct¹²³ est établi entre les premières et la périphrase de référence, sur la base de notre algorithme principal.
- Deuxième étape : Chacune des traductions en langue(s) cible(s) des traductions françaises des traductions accessibles et cachées de *se mettre à* sont recensées et intégrées de la même façon que précédemment.
- Ainsi de suite... Il est évident qu'un tel travail risque fort de se révéler fastidieux s'il n'est pas correctement borné. Pour limiter la taille du

¹²³ Les traductions directes se voient affecter un lien de synonymie directe avec la lexie d'origine. Les cachées en auront un partiel.

champ, il peut se révéler nécessaire de prédéfinir un paradigme de synonymes du mot-noyau¹²⁴. Ce faisant, nous excluons du balayage tout terme en langue source qui ne ferait pas partie de ce paradigme initial ; et nous ne « rebalayons » bien évidemment pas à partir d'une unité lexicale qui a déjà été visée.

La première traduction anglaise de notre périphrase-noyau est *to begin*, laquelle sera par conséquent notre second verbe de référence¹²⁵ pour continuer la construction du CSMU autour de *se mettre à*.

Ensuite, ce sont les traductions propres de *to begin*, à savoir *commencer à* et *se mettre à*, qui sont prises en compte. La seconde ayant déjà été balayée (puisque c'est la lexie-noyau d'origine), nous n'en refaisons pas l'analyse :

- **1**¹²⁶*se mettre à + GVinf* → *to begin*
- **2***begin to + GV* → *commencer à, se mettre à*¹²⁷ (noyau) ;
- **3***commencer à + GVinf* → *begin to, start to, begin + ing, start + ing, set in* (impersonnel en français, substantivation du sujet en anglais), *get* (adjectivation de l'objet en anglais), [locutions : *commencer à en avoir assez, à bien faire*] ;
- **4***start to* → *commencer à + GVinf, se mettre à + GVinf, ne pas tarder à* (avec *soon* en anglais) ;
- **4***begin + ing* → *commencer (à, de), se mettre (à), entreprendre, entonner, déclencher, entamer, déboucher, prendre un nouveau, partir en, débiter dans....* ;
- **4***start + ing*
- etc.

Les informations syntaxiques, pragmatiques et lexicales que le dictionnaire propose sont présentes dans cet extrait et sont également intégrées dans nos champs lexicaux finaux, par l'intermédiaire de la base de données.

6.4 Bilan de l'intégration des périphrases à l'étude, et des lexies aux champs lexicaux de façon générale

L'avantage le plus saillant de ces champs sémantiques unifiés est à l'évidence de permettre un rapprochement lexico-conceptuel de langues différentes. La méthode soulève néanmoins de nombreuses questions pratiques et méthodologiques : en

¹²⁴ Comme nous le proposons pour *abandonner*.

¹²⁵ Le premier verbe de référence étant bien entendu *se mettre à*.

¹²⁶ le chiffre 1 indique que nous sommes à la première étape du balayage. Le verbe suivant, muni de l'indice 2, est par conséquent à la deuxième étape, et ainsi de suite.

¹²⁷ Déjà « balayé » et par conséquent ignoré pour la suite du processus.

premier lieu, quel crédit peut-on conférer à des œuvres lexicographiques humaines ? Si nous les jugeons infaillibles, les champs qui en découlent doivent l'être tout autant. Or, les erreurs et plus encore les lacunes qu'ils contiennent se répercutent fâcheusement au sein des graphes générés même si elles se rencontrent en nombre limité.

Dans une perspective d'automatisation, la mise en relation des lexies par balayage et relations synonymiques nous permet de diluer certaines associations « douteuses » et ainsi de pallier les insuffisances présumées des dictionnaires de langues. Toutefois, il n'est ni dans nos intentions ni dans nos prérogatives de critiquer ces derniers ; seul l'établissement d'une méthodologie originale pour la construction de nouveaux lexiques nous conduit à les analyser dans les moindres détails.

En ce qui concerne les périphrases verbales, et par extension toutes les locutions à sens identifiable, leur présence au sein des champs se révèle aussi utile qu'indispensable. La forme de nos lexiques leur confère une entrée autonome par rapport à leurs constituants monolexicaux, ce qui permet par exemple de clairement identifier *se mettre à* comme lexème, en le distinguant du verbe *mettre*. Nous ne sommes pas loin d'en proposer un traitement autonome dans les dictionnaires, de façon à aller plus loin que Rey et Chantreau qui persistent à les répertorier sous le terme monolexical « le plus représentatif » dans leur dictionnaire des locutions.

7. Prolongement des LI

7.1 La traduction intermédiaire : Paraphrase ou pas paraphrase ?

Au début de nos recherches, nous pensions que toute lexie devait être paraphrasée - ou formalisée¹²⁸ - afin d'être confrontée sous une forme non littérale aux candidats. De ce fait, la « traduction intermédiaire » devait être le premier travail indispensable sur le mot-source hors (ou parallèlement à) la reconnaissance des registres de langue. Nous partirons de ce principe dans les paragraphes qui suivent, avant de proposer une conclusion critique de cette méthode de transfert.

Pour susciter une traduction, un terme peut devoir être « réécrit ». Il faut donc accomplir un travail initial de *paraphrase*. Celui-ci n'est pas exclusivement déterminé par les entrées des dictionnaires monolingues mais en nécessite néanmoins l'emploi¹²⁹.

D'autre part, la paraphrase que nous préconisons ne doit pas être trop fine dès le départ de l'étude, par exemple il nous paraît inutile de considérer *toucher au rivage*

¹²⁸ Avec un formalisme qui en précise la signification. Voir à ce sujet la primitive proposée pour le verbe *abandonner* au milieu de ce chapitre.

¹²⁹ Pour aborder (français → italien), les sens résiduels « attaquer (un ennemi) » ou « commencer à traiter (une matière (intellectuelle)) » sont induits par les traductions proposées dans les dictionnaires bilingues mais sont ignorés - ou non explicités- par les monolingues que nous avons consultés (Larousse, Robert). Ces acceptions « nouvelles » sont bien présentes dans le TLF, lequel pourrait être outil lexicographique de référence s'il n'était aussi riche.

et *toucher le rivage* comme deux paraphrases-sources distinctes aussi bien parce que le sens profond est le même pour les deux constructions que parce que le mot-cible équivalent leur est également commun. Toutefois, comme il s'agit de deux constructions syntaxiques distinctes, il faudra nécessairement les implanter toutes deux explicitement lors de l'automatisation du lexique, ce qui motive une distinction claire entre paraphrases et critères discriminants dans la perspective de constituer la base de données.

En définitive, la paraphrase définitoire n'est pas indispensable dans l'optique de réaliser des graphes lexico-sémantiques car le format de la base de données permet d'intégrer un nombre de critères non préalablement délimité. D'autre part, nous pouvons créer autant de liens que nécessaire dans l'algorithme qui l'accompagne.

Au cœur de notre réflexion méthodologique, nous suggérons l'emploi des liens synonymiques entre les lexies de toutes les langues à implanter dans nos graphes. De ce fait, le meilleur synonyme ou la meilleure paraphrase d'un mot n'est peut-être rien d'autre que ses traductions (donc dans d'autres langues que lui-même) voire éventuellement les traductions de ses traductions à contexte équivalent (lien indirect de synonymie qui ramène à une confrontation entre mots d'une même langue).

7.2 Représentation des lexiques multilingues issus du LI. Projet de superposition de champs sémantiques en plusieurs langues.

7.2.1 Transition avant graphe

7.2.1.1 Aller au-delà des LI bidimensionnels

Certes, les Lexiques Inversés ont permis de donner une nouvelle vision des dictionnaires de langues. Toutefois, la représentation bi-dimensionnelle des correspondances interlexicales n'offrant pas de résultats satisfaisants au-delà du bilingue, nous avons réfléchi à une autre méthode de co-représentation des données qui permettrait en particulier d'implanter toutes les langues impliquées dans un même graphe.

7.2.1.2 Exploration d'un formalisme conceptuel

La première piste envisagée était d'élaborer des « cases lexico-conceptuelles » (Clc) dans un réseau sémantique préexistant, de type DI¹³⁰ ou Wordnet. Chacune de ces cases, rappelons-le, correspond à un sens déterminé du terme d'origine et contient toutes ses traductions les plus pertinentes dans les langues cibles. Cette méthode pose deux problèmes essentiels :

- Elle contraint une représentation initiale des connaissances fondée sur une langue donnée (le français en l'occurrence).
- Elle ne met pas en exergue l'inadéquation entre les systèmes lexico-conceptuels des différentes langues. On suppose en effet que la constitution des Clc a pour effet de figer le rapport interlangue dans une forme de pluriunivocité qui n'existe pas dans les faits.

Ces deux points ont donc conduit à la recherche d'une modélisation plus souple et peut-être plus floue du lexique multilingue. En partant du principe que chaque langue constitue un continuum instable et hétérogène et qu'il n'y a pas (répétons-le) adéquation avec les autres langues, il s'agissait de rendre ces disparités tout en respectant les spécificités de chacune.

En premier lieu, il nous a paru encore plus indispensable que judicieux de raisonner en terme de champs sémantiques et non plus de lexèmes, au nom du principe prévalant qu'il vaut mieux éviter d'utiliser préférentiellement une des langues impliquées comme point de départ de la représentation des connaissances. Ces champs sémantiques sont intégrés au réseau sémantique par l'intermédiaire de primitives dont nous donnons un aperçu un peu plus loin. Pour être clair, voici comment s'organiseraient ces champs :

- A partir des classes ontologiques, constitution de primitives conceptuelles (dans notre cas, ce seraient des prédicats primitifs formalisés) ;
- construction des champs sémantiques unilingues et superposition de ces derniers au sein du même graphe avec des marqueurs de langue (F pour français, E pour english, etc.). Ou encore construction d'un champ sémantique dans une langue donnée et ajout successif ou simultané des champs sémantiques associés dans les autres langues ;
- pour chaque mot intégré au champ, constitution d'une structure de traits que nous décrivons par ailleurs. Cette structure n'apparaîtra pas

¹³⁰ Dictionnaire Intégral de Dominique Dutoit, largement décrit au Chapitre I

sur le graphe mais pourra être directement activée depuis celui-ci (par hyper-lien). Dans un but de TAL, ces structures seront particulièrement importantes.

En principe, on ne devrait pas trouver plusieurs unités lexicales occupant la même case au sein du champ mais il peut y avoir des cas de parallélisme parfait faisant figure d'exception comme cela se rencontre assez souvent entre le français et l'italien. L'objectif est une représentation commune qui fait l'économie du pivot, ne met pas de langue en avant par rapport aux autres et met en exergue les décalages lexico-conceptuels interlangue.

Il reste maintenant à savoir comment construire ces fameux champs sémantiques. Pour cela, nous aurons besoin de constituer une méthodologie, un algorithme et de faire apparaître la démarche de l'implantation de données vers leur exploitation en TAL. Toute cette problématique est l'objet du chapitre IV.

7.2.2 Tableaux de données

Toutes les données présentes ci-dessous dans les fragments de notre base sont directement issues des dictionnaires, sans ajouts d'aucune sorte et moyennant certains aménagements pour les faire correspondre à nos catégories classifiantes.

La forme de ces bases illustre la question des synonymes que nous venons de soulever.

Il est à noter que *obj1* est l'objet ou le complément d'un verbe transitif et *obj2* est le second complément d'un verbe bitransitif, le second complément non obligatoire d'un verbe transitif ou le complément non obligatoire d'un verbe intransitif.

7.2.2.1 Essai sur aborder

Figure 6 : Fragment du tableau de données autour de *aborder*

<i>motsource</i>	<i>LS</i>	<i>motcible</i>	<i>LC</i>	<i>synLS</i>	<i>SynLC</i>	<i>obj1</i>	<i>Obj2</i>	<i>subj</i>	<i>constrLS</i>	<i>constrLC</i>	<i>registre</i>	<i>infoscomp ls</i>
aborder		abordar			atracar		(port, pays)		a/en obj2	a/en obj2		
aborder		atracar			abordar				a/en obj2	a/en obj2		
aborder		abordar				navire						
aborder		atacar		attaquer	acometer						fig	
aborder		acometer		attaquer	asaltar						fig	
aborder		asaltar		attaquer	atacar						fig	
aborder		atacar		attaquer	asaltar						fig	
aborder		acometer		attaquer	atacar						fig	
aborder		asaltar		attaquer	acometer						fig	

<i>motsource</i>	<i>LS</i>	<i>motcible</i>	<i>LC</i>	<i>synLS</i>	<i>SynLC</i>	<i>obj1</i>	<i>Obj2</i>	<i>subj</i>	<i>constrLS</i>	<i>constrLC</i>	<i>registre</i>	<i>infoscomps</i>
aborder		abordar				personne					fig	
aborder		abordar			tratar	(sujet) ¹³¹					fig	
aborder		tratar			tocar	(sujet)					fig	
aborder		tocar			abordar	(sujet)					fig	
aborder		abordar			tocar	(sujet)					fig	
aborder		tratar			abordar	(sujet)					fig	
aborder		tocar			tratar	(sujet)					fig	
aborder		enfocar		envisager								
aborder		emprender				ouvrage, lecture						
s'aborder		abordarse										
aborder		(acceder) ¹³²		(accéder)								
aborder		(acostar)		(accoster)								
aborder		(tener acceso)		(avoir accès)								
aborder		(unirse)		(joindre)								
aborder		(acercarse)		(s'approcher)								
aborder		(tocar)		(toucher)								

7.2.2.2 Essai sur abandonner

Figure 7 : Fragment du tableau de données autour de *abandonner*

<i>motsource</i>	<i>LS</i>	<i>motcible</i>	<i>LC</i>	<i>synLS</i>	<i>SynLC</i>	<i>obj1</i>	<i>obj2</i>	<i>subj</i>	<i>constrLS</i>	<i>constrLC</i>	<i>registre</i>	<i>infoscomps</i>
abandonner	f	abandonar	e	délaisser								
abandonner	f	dejar	e	laisser		(ouvrage)						
abandonner	f	descuidar	e	négliger		(devoirs)						
abandonner	f	descuidar	e		abandonar	(amis)						
abandonner	f	abandonar	e		descuidar	(amis)						
abandonner	f	soltar	e	lâcher		(rênes)						

¹³¹ La collocation objet « *un sujet* » est présente dans le syntagme exemple « *aborder un sujet* » qui accompagne les trois traductions *abordar*, *tratar* et *tocar*. Toutefois, elle n'est pas présentée de telle façon qu'on comprenne qu'il s'agit d'une restriction sur l'ensemble des objets possibles pour l'acception considérée. Dans ces conditions, nous le faisons apparaître entre parenthèses.

¹³² Les dictionnaires bilingues Larousse ont la particularité de proposer parfois des synonymes traduits de la même langue que la vedette en entrée. Ici *accéder* est présenté comme synonyme d'*aborder* et se traduit par *acceder* lorsque son acception se rapproche de celle(s) d'*aborder*. Ceci ne signifie pas que *acceder* est une traduction potentielle d'*aborder* sans l'exclure totalement. Cela ne veut pas non plus dire, et c'est très important, que *acceder* est **nécessairement** candidat à la traduction à l'entrée d'*accéder*. Nous avons choisi de ne pas ignorer ces couples synonyme-traduction dans nos bases de données, bien que nous ne puissions pas leur réserver le même sort que les traductions directement accessibles. Dans l'immédiat, nous les mettons entre parenthèses à l'intérieur de la base de données afin de les distinguer des autres types de données.

<i>motsource</i>	<i>LS</i>	<i>motcible</i>	<i>LC</i>	<i>synLS</i>	<i>SynLC</i>	<i>obj1</i>	<i>obj2</i>	<i>suj</i>	<i>constrLS</i>	<i>constrLC</i>	<i>registre</i>	<i>infoscomp ls</i>
abandonner	f	confiar	e	confier	dejar							
abandonner	f	dejar	e	confier	confiar							
abandonner	f	conceder	e	accorder		(detail)						
abandonner	f	desistir de	e		renunciar a	(prétentions)						
abandonner	f	renunciar a	e		desistir de							
abandonner	f	renunciar a	e		cesar	(lutte)						
abandonner	f	cesar	e		renunciar a	(lutte)						
abandonner	f	entregar	e	livrer		(biens)	(créancier)		a obj2	a obj2		
abandonner	f	entregar	e	remettre		(biens)	(créancier)		a obj2	a obj2		
abandonner	f	desahuciar	e			(malade)						
abandonner	f	abandonar	e						[intrans]			
s'abandonner	f	abandonarse	e									
s'abandonner	f	desanimarse	e	perdre courage								
s'abandonner	f	acorbardarse	e	prendre peur								
s'abandonner	f	descuidarse	e		dejarse							dans sa tenue
s'abandonner	f	dejarse	e		descuidarse							dans sa tenue
s'abandonner	f	entregarse	e	se livrer			(sommeil)		a obj2	a obj2		

Nous avons saisi des données presque complètes pour le verbe *abandonner* et ses synonymes et quasi-synonymes entre le français et l'anglais. Les bases qui en découlent seront présentées en totalité ou en partie, selon le volume, dans l'annexe qui sera jointe au document rédigé principal.

7.2.2.3 Essai sur se mettre (à)

Figure 8 : Fragment du tableau de données autour de *se mettre (à)*

<i>motsource</i>	<i>LS</i>	<i>motcible</i>	<i>LC</i>	<i>synLS</i>	<i>SynLC</i>	<i>obj1</i>	<i>obj2</i>	<i>suj</i>	<i>constrLS</i>	<i>constrLC</i>	<i>Registre</i>	<i>Infoscomp ls</i>
se mettre	f	go	e	se placer				objet				
se mettre à	f	turn	e			froid, chaud, orage						
se mettre à	f	go on	e			régime						
se mettre à	f	set to	e		get down to	travail						
se mettre à	f	set down to	e		set about one's	travail						
se mettre à	f	set about one's	e		set to	travail						
se mettre à	f	set to	e		set about one's	travail						
se mettre à	f	get down to	e		set to	travail						
se mettre à	f	set about one's	e		get down to	travail						
se mettre à	f	start	e		set about	traduction						
se mettre à	f	set about	e		start	traduction						

<i>motsource</i>	<i>LS</i>	<i>motcible</i>	<i>LC</i>	<i>synLS</i>	<i>SynLC</i>	<i>obj1</i>	<i>obj2</i>	<i>suj</i>	<i>constrLS</i>	<i>constrLC</i>	<i>Registre</i>	<i>Infoscomp</i> <i>ls</i>
se mettre à	f	take up	e			langue						
se mettre à	f	start	e		begin to	action			+ inf	+ inf		
se mettre à	f	begin to	e		start to	action			+ inf	+ inf		
se mettre à	f	start to	e		start	action			+ inf	+ inf		
se mettre à	f	start	e		start to	action						
se mettre à	f	begin to	e		start	action						
se mettre à	f	start to	e		begin to	action						
se mettre à	f	coming on to	e			pleuvoir						

7.2.3 Conclusion intermédiaire

Ces tableaux de données sont « opératoires », c'est-à-dire qu'ils vont directement permettre une implantation du graphe lexico-sémantique via l'algorithme approprié.

Au sens où nous l'entendons jusqu'à présent (précisé au Chapitre I), chaque ligne des bases correspond à une lexie, i.e. à un emploi précis d'un terme, et à son ou ses équivalent(s) en langue cible et éventuellement en langue source si le verbe proposé en traduction est muni d'un renvoi vers un autre verbe que celui représenté par la lexie. Dans l'immédiat et au risque de nous répéter, nous n'avons rien saisi d'autre que ce que les dictionnaires mentionnent explicitement dans leurs pages.

8. Transition avant étude sur corpus

8.1 Le statut des traductions dans le miroir

Avant de poursuivre l'étude du lexique par des considérations linguistiques et algorithmiques, il ne faut pas négliger de jeter un regard critique sur la structure et la validité du contenu des LI.

Tout d'abord, leur création n'améliore pas la lisibilité de la traduction d'une langue à une autre. Au contraire, l'afflux de nouveaux traductèmes en rangs désordonnés a plutôt pour effet d'alourdir les paradigmes sans réellement proposer de solution pour clarifier les choix et les critères qui permettent de les faire. S'ils demeurent importants dans la perspective d'études contrastives ciblées comme celles que nous avons menées dans ce chapitre, les LI seront remplacés par les CSMU dans une perspective de modélisation. D'un certain point de vue les bases de données menant aux CSMU sont la résultante de la fusion filtrée entre le LD et le LI tel que nous en caressons le projet, le tout présenté dans un format qui donne directement accès aux informations linguistiques propres à chaque lexie tout en respectant l'orientation de la traduction proposée par les dictionnaires. Autrement dit, le principe du miroir est quelque peu nuancé par la notion de priorité dans les traductèmes, le LI

ne saurait se substituer à un « second traducteur du LD ». D'autre part, la question de l'orientation du graphe est un faux problème préalable puisqu'un algorithme digne de ce nom peut aussi bien désorienter le graphe que le conserver en l'état initial.

8.2 Les limites de l'observation des dictionnaires

L'observation des dictionnaires est déjà en soi une étude sur corpus, et ce doublement : tout d'abord, nous admettons qu'un ouvrage lexicographique est un corpus, c'est-à-dire un ensemble d'éléments représentatifs de la langue et d'un paradigme particulier, en l'occurrence celui des lexies de plusieurs langues. D'autre part, les dictionnaires utilisent les corpus pour leur constitution.

Néanmoins, un certain nombre de contraintes pour la plupart liées au format de l'édition empêchent les ouvrages lexicographiques d'offrir une visibilité entièrement satisfaisante des phénomènes linguistiques qui préexistent à leur création. Il est également évident qu'un dictionnaire grand public se doit de refléter la langue académique, celle qui est préconisée par les « gardiens du Temple », et accessoirement les faits de langue qui sont présentés dans les dictionnaires monolingues du même éditeur. Il est peu probable qu'un dictionnaire sérieux se permette de répertorier minutieusement les fantaisies des auteurs littéraires en les présentant comme des faits de langue établis, notamment à destination des locuteurs non natifs. C'est pourquoi nous avons été amené à poursuivre nos réflexions par des analyses linguistiques débarrassées des contraintes liées au cadre lexicographique, en nous fondant exclusivement sur des données attestées extraites de la base littéraire FRANTEXT.

En particulier, nous avons relevé deux phénomènes linguistiques que les dictionnaires bilingues ne peuvent pas traiter de par leur nature :

- Tout d'abord, il est très difficile de sélectionner des traductions concurrentes pour une même acception en fonction de la nature sémantique des arguments du verbe, voire en fonction du co-texte un peu plus éloigné, ce dernier étant moins dépendant du verbe mais participant également à sa caractérisation sémantique et aspectuelle. Même dans le dictionnaire grand public le plus élaboré (i.e. offrant des informations relativement précises sur les arguments habituels pour chaque acception), le problème demeure car l'ouvrage ne propose que le choix d'un objet ou d'une classe d'objets sans entrer plus en détails dans sa nature sémantique. C'est pourquoi nous avons décidé de nous pencher sur le verbe *traverser*, en collaboration avec Laurent Mascherin qui s'intéressait à ce verbe au même moment, avec des objectifs légèrement différents. Le verbe *traverser* est fortement dépendant de son environnement co-textuel quant à sa caractérisation aspectuelle, ce qui le rend difficile à traduire si ce co-texte n'est pas précisément étudié.

- Ensuite, les dictionnaires bilingues ainsi que les monolingues destinés au grand public sont avant tout des ouvrages académiques qui se doivent de représenter la langue telle qu'elle est normativement établie à un moment du temps. Ceci exclut de pouvoir y trouver des sens de mots nouveaux ou des constructions syntaxiques nouvelles qui n'auraient pas été validées par les lexicographes comme faisant partie intégrante de la langue. Or, l'observation des œuvres littéraires fait clairement apparaître que les sens et constructions « inhabituels » ou « inattendus » sont loin d'être des phénomènes résiduels. À ce sujet, les auteurs littéraires font la preuve que la langue n'est pas seulement un système figé mais surtout un socle normatif à partir duquel de nombreuses fantaisies d'expression sont possibles. L'émergence de sens nouveaux pour les verbes eux-mêmes est parfois difficile à décrire ; nous l'avons notamment constaté avec certaines occurrences de *traverser* qui restent pour nous ininterprétables, et par voie de conséquence intraduisibles. En revanche, il est plus aisé de répertorier les constructions syntaxiques nouvelles (surtout si elles se répètent dans des œuvres littéraires différentes) de même que l'originalité dans la nature sémantique des arguments du verbe. Les verbes qui font partie du vaste paradigme des verbes de parole sont particulièrement malléables de ce point de vue et c'est ce qui nous a conduit à en proposer l'étude à partir de FRANTEXT dans le but avoué d'enrichir les bases de données d'origine lexicographique sans en modifier la structure.

Ces approches seront l'objet des développements suivants dont la finalité est d'observer la réalité du langage sans contraintes *a priori* afin de ramener la description de celui-ci aux intentions « réelles » des auteurs ou à tout le moins de s'en rapprocher le plus possible. À l'issue de nos analyses, nous espérons montrer comment l'observation de données attestées peut compléter ou enrichir nos bases de données sans en bouleverser la méthodologie dans le but de réunir toutes les informations linguistiques dans une même source d'informations.

9. Les verbes « de parole »¹³³ en contexte

Le nombre très élevé des verbes de parole en français ne nous a pas permis d'effectuer une synthèse complète des données lexicographiques les concernant. Toutefois, l'intérêt d'observer le paradigme dans son ensemble réside davantage dans les disparités et les similarités dans leurs structures argumentales sémantiques, et dans les constructions syntaxiques attenantes. Nous pouvons noter en particulier la multiplicité des types de compléments que l'on peut trouver derrière les verbes de parole. Une typologie de ces compléments devra prendre en compte de manière orthogonale leurs caractéristiques grammaticales aussi bien que leur contenu sémantique. À propos de ce dernier, nous remarquerons qu'un même objet sémantique peut être objet direct de certains verbes et objet indirect de certains autres.

9.1 Quelques mots sur la linguistique de corpus

Avant de nous lancer dans l'analyse proprement dite, nous allons proposer une petite réflexion au sujet du corpus et de son utilisation en linguistique.

Lorsqu'on parle de contexte dans le cadre de la linguistique, c'est généralement le co-texte qui vient immédiatement à l'esprit, c'est-à-dire le « texte proche » de l'occurrence que l'on veut décrire ou encore ce qui est dit ou écrit avant ou après cette occurrence. En réalité, le contexte est une notion beaucoup plus vaste qui englobe de nombreux types qu'on ne peut pas ignorer pour constituer une typologie mais qui ne sont pas tous perceptibles lorsqu'on n'a pas pleinement accès au contexte d'une production de langage. Pour essayer d'expliquer la notion, Georges Kleiber propose de l'illustrer par un certain nombre d'oppositions ou de « dimensions du contexte »¹³⁴ :

« (...) La meilleure manière, à notre avis, consiste à parcourir un certain nombre d'oppositions, de dimensions du contexte, reconnues ou non, plus ou moins fortement présentes dans la littérature, telles que (...) :

- 1) en contexte/hors contexte
- 2) contexte défini comme ensemble d'éléments qui influencent/contexte comme environnement
- 3) contexte inventaire/contexte comme cadre ou ensemble structuré

¹³³ Pour respecter la nouvelle orientation de recherche des membres du PDR de l'ILF auquel nous avons participé, nous devons nuancer cette appellation de « verbes de parole » en précisant qu'il s'agit avant tout de verbes impliquant nécessairement et explicitement de la production de langage humain dans des emplois bien déterminés. Cela en fait avant tout des « verbes de communication langagière » davantage que des « verbes de parole » au sens communément admis. Toutefois, l'annonce tardive de cette nouvelle dénomination et des réserves qui l'accompagnent nous ont conduit à ne pas bouleverser la structure de notre sous-partie qui traite de ces verbes, mais nous a contraint à en restreindre l'importance.

¹³⁴ Kleiber (1997). « Quand le contexte va, tout va et ... inversement », in *Co-texte et calcul du sens*, Claude Guimier éd., Presses Universitaires de Caen.

- 4) contexte : emploi rationnel (direct)/emploi absolu (indirect)
 - 5) contexte générique/contexte spécifique
 - 6) contexte linguistique (co-texte)/contexte extra-linguistique (ou situationnel)
 - 7) contexte immédiat (micro-contexte)/éloigné (macro-contexte)
 - 8) situation d'énonciation immédiate/situation plus large
 - 9) contexte gauche/contexte droite
 - 10) contexte statique/contexte dynamique (ou processuel)
 - 11) contexte analyste-observateur/contexte locuteur/interlocuteur
 - 12) contexte du locuteur/contexte de l'interlocuteur
 - 13) le contexte détermine le sens/le sens détermine le contexte
 - 14) contexte global ou texte (?)/contexte d'une unité (qui peut varier)
 - 15) le contexte pré-existe (est donné)/le contexte est construit, etc. »
- Kleiber (1997 :13)

Comme le suggère Kleiber dans ce relevé, la liste des oppositions et des dimensions peut être poursuivie longuement.

La première opposition est essentielle car elle ouvre le champ des possibles pour le contexte dès lors qu'il est admis que le *hors contexte* n'est rien d'autre que :

« (...) [l']analyse de l'élément en lui-même, sans influence d'autres éléments, le *contexte* dans cette perspective est conçu comme étant ce avec quoi se combine l'élément (en laissant un sens très large à la notion de *combinaison*) (...) »

Kleiber, op. cit., p.14

Dans cet ordre d'idées, l'analyse dite *hors contexte* devient un exercice de haute voltige sans spectateur dans la mesure où un élément n'est jamais complètement isolé, fût-il la partie d'un exemple construit, lui-même *hors contexte* dans un sens moins restreint. Ce paradoxe est également relevé par Kleiber étant entendu que le co-texte d'un exemple construit est déjà en soi un contexte pour l'élément visé.

À partir de ces considérations de départ et dès qu'on admet la séparation entre exemples construits et exemples attestés, l'opposition qui nous paraît la plus intéressante est la sixième de Kleiber : contexte linguistique (co-texte)/contexte extra-linguistique (ou situationnel). En la précisant, on parvient à expliquer en partie un bon nombre des autres oppositions qu'il propose. En premier lieu, il convient de ne pas confondre co-texte et contexte, le second englobant notionnellement le premier. Le co-texte d'une unité est toujours présent dans l'analyse linguistique, qu'il s'agisse d'étudier des exemples construits¹³⁵ ou des occurrences attestées. En revanche, le

¹³⁵ On peut se référer à diverses « polémiques amicales » (M. Charolles) et à certains exercices de style, comme dans Kleiber (1994) « Discours et stéréotypie : le contexte peut-il remettre d'aplomb une anaphore associative mal formée ? », in C. Schnedecker,

contexte extra-linguistique (que nous appellerons *situation d'énonciation* pour éviter de le confondre avec *contexte* au sens large) est nécessairement l'apanage d'une « situation réelle » de production langagière. Or, ce dernier point de vue ne va pas sans un certain paradoxe : si la situation d'énonciation est propre au réel, qu'en est-il des occurrences attestées issues de corpus dont on ignore avec le recul tout ce qui est du domaine extra-linguistique ? Nous n'avons pas pu résoudre ce problème lors de nos analyses linguistiques sur corpus littéraires, car l'auteur d'un roman est parfois le seul à pouvoir décrire précisément la situation d'énonciation associable à chaque passage de son ouvrage, si tant est qu'il eût l'intention de la caractériser.

Par conséquent, des travaux comme les nôtres ne peuvent pas prendre en compte de manière exhaustive des données qui nous échappent. Il ne nous reste généralement que le co-texte à étudier, et nos intuitions, nos connaissances et notre imagination pour placer celui-ci dans un cadre d'analyse digne de ce nom. De ce fait, la réponse à la question « le co-texte est-il un contexte restreint ? » est nécessairement positive, notre position d'analyste en herbe ne pouvant nous permettre d'inventer à la place des auteurs un contexte plus large dans lequel interviennent pourtant des facteurs essentiels pour la bonne compréhension globale des énoncés. En ce sens, le travail *hors contexte*, ou plus exactement sur exemples construits (éventuellement enrichis, comme nous le précisons en note), ne paraît plus aussi condamnable que ce qui se dit de plus en plus dans la communauté scientifique.

C'est dans la droite ligne de cette réflexion d'ensemble que nous estimons dangereuse la polémique qui consiste à stigmatiser les « bâtisseurs » de corpus *ad hoc*. Après tout, tous les exemples attestés ont été un jour construits par quelqu'un, certes rarement par un linguiste, et en un sens ont répondu également à un besoin particulier de l'auteur qui les a produits.

Même si le corpus est au centre de la polémique sur la linguistique contextuelle, il n'est pas facile d'expliquer clairement de quoi il s'agit. En tant que tel, un corpus est un ensemble de données homogène, c'est-à-dire qui concentre des textes ou des objets textuels de même nature. D'après cette définition initiale, on peut considérer qu'un ensemble d'énoncés construit par un linguiste est en soi déjà un corpus. C'est pourquoi nous devons en restreindre la définition pour pouvoir parler de linguistique de corpus sans nous contenter de l'opposer en tant qu'« objet impur » à la « grammaire universelle » élevée au rang de « système de la langue débarrassée de ses souillures que sont la culture, la société, l'humeur ou les pathologies du locuteur ».¹³⁶

M. Charolles, G. Kleiber et J. David (éd.), *L'anaphore associative. Aspects linguistiques, psycholinguistiques et automatiques*, Paris, Klincksieck.

À partir d'un seul exemple d'anaphore associative, Georges Kleiber montre qu'en étendant le contexte d'un exemple construit, on résout certaines incohérences qui préexistaient du fait de l'existence de l'occurrence *hors contexte* suffisamment étendu.

¹³⁶ Cette opposition est mise en relief par Damon Mayaffre dans sa conférence inaugurale au colloque JETOU 2005, *Rôle et place des corpus en linguistique* : « Rôle et place des corpus en linguistique : réflexions introductives ».

En linguistique de corpus proprement dite, le corpus est tout ensemble de données qui n'a pas été conçu dans l'optique explicite de devenir un observable linguistique. D'après Damon Mayaffre, la question est de savoir si le véritable corpus devenu un observable malgré lui est un « *observatoire* d'une théorie a priori » ou « un *observé dynamique* qui permet de décrire puis d'élaborer des modèles a posteriori. » Cette opposition épistémologique sert à Sinclair pour présenter ses deux linguistiques de corpus que nous allons décrire ci-après. Cette question de fond ne remet pas en question la nature propre du corpus qui demeure un statut d'observable duquel on fait ce qu'on veut.

Derrière ces deux interrogations, contradictoires en apparence, se cache toute la polémique actuellement vivace sur le sujet. En effet, avec l'émergence toute nouvelle des moyens informatiques de numérisation et de traitement des données textuelles, étudier un phénomène de langue sur corpus n'est plus un exercice fastidieux ni une impossibilité. Il reste néanmoins des domaines où la recherche d'occurrences demeure un travail de fourmis, comme c'est le cas pour les études en ancien français. Toutefois, la grande disparité et, disons-le tout net, la relative pauvreté des corpus du français peuvent remettre en question l'effort de systématisation d'une discipline que l'on nommerait « linguistique de corpus », faute de pouvoir la caractériser plus précisément.

Il ne s'agit pas pour nous d'entrer dans une polémique centrée sur des questions aussi circulaires que :

- Peut-on se passer des corpus en linguistique ?
- Comment faisait-on avant ?, sous-entendu avant l'ère de la numérisation ?

Avec les études que nous avons menées au sujet des verbes de parole et de traversée à partir de FRANTEXT, nous nous limiterons à proposer ou à corroborer des pistes de réflexion sur le rôle et la place que l'on peut accorder à un corpus textuel dans une étude linguistique, de même que nous essaierons d'en repérer et décrire les limites.

Selon John Sinclair, précurseur et pionnier de l'étude des collocations, auteur et initiateur du COBUILD¹³⁷, il existerait deux courants majeurs sous l'appellation de linguistique de corpus :

La « linguistique fondée sur le corpus » utilise le corpus pour valider ou vérifier des hypothèses antérieures à la consultation des données. Cité par Ooi (1998), Sinclair la décline comme suit :

Corpus-‘based’ linguistics

- a corpus is used to validate, check, and improve linguistic observations that have already been made; the corpus-based linguist does not feel ‘threatened’ by corpus data at all
- the linguist does not question received theoretical positions or well-established descriptive categories; instead, his position to language structure is already well-formed
- the corpus is used to help extend and improve linguistic description

An example of a relevant question: “Is *whom* still used in English, and if so how?”

Sinclair (1996)

L'expression la plus frappante dans la définition proposée par Sinclair est que le linguiste ne se sent pas du tout « menacé » par les données attestées dans le cadre de la *corpus-based linguistics*, ce qui semble induire qu'il existe des situations dans lesquelles le linguiste se sent menacé. On peut sentir dans cette description les limites entre une linguistique « confortable » mais dénuée de ressources extérieures et la linguistique dirigée par le corpus, laquelle présuppose que le linguiste n'a que peu d'a priori avant d'entamer ses recherches, comme nous allons le voir maintenant.

Également définie par Sinclair et directement comparée à la précédente, la linguistique « conduite » ou « dirigée » par le corpus n'a pas les mêmes orientations que son équivalente « fondée sur le corpus. » La caractéristique principale de la « corpus-driven linguistics » est qu'il n'y a pas d'hypothèse sur la langue qui soit antérieure à l'observation des données. Ce sont ces dernières qui font émerger les phénomènes que le linguiste doit décrire :

Corpus-‘driven’ linguistics

- a corpus is of prime importance in bringing out new ideas for the examination of data

¹³⁷ COBUILD : base de données des collocations pour l'anglais initiée par John Sinclair.

- the linguist believes that the kind of evidence emerging from corpora may be difficult to reconcile with established positions in the discipline, and he leaves open the possibility of the need for a radical change in linguistic theory in order to cope with the evidence
- evidence from the corpus is paramount, therefore the linguist makes as few assumptions as possible about the nature of the theoretical and descriptive categories

An example of a relevant question: “Is the distinction between grammar and lexis necessary?”
Sinclair (op.cit.)

En marge des deux orientations précitées, nous entrevoyons l’émergence potentielle d’un troisième courant, celui que nous appellerions la « corpus-associated linguistics ». Dans un esprit de réconciliation des théories, ce courant potentiel de la linguistique de corpus résulte de la collaboration entre des hypothèses ou des données pas forcément attestées et les occurrences issues de corpus, les unes et les autres se complétant sans s’annuler.

C’est de cette façon que nous voulons bâtir notre propre méthodologie de façon à réaliser des modèles protéiformes capables de s’adapter à toutes les sources de données possibles. Par extension, nous avons également voulu montrer qu’il n’y a pas forcément d’incompatibilité entre les sources monolingues et bilingues dès lors que la modélisation est supportée par un algorithme susceptible de prendre en compte tous les cas de figure.

9.2 Traitement des verbes de parole

Reprenons la définition que Jamrozik présente dans sa thèse (1992) :

‘La totalité du corpus envisagé compte 410 verbes qui, malgré l’appellation commune de verbes de parole, forment un ensemble sémantiquement et syntaxiquement hétérogène.

Par verbes de parole, ou verbes du dire, nous entendons les verbes qui dénotent une activité linguistique du sujet parlant, exercée en vue de communiquer un message au sens large du terme : transmission d’information, tentative d’influer sur le comportement de l’interlocuteur, demande d’information, expression des sentiments, etc. La communication ne se réduit pas ainsi à un simple échange de connaissances, mais englobe un cadre varié de relations interhumaines. Dans ce sens, notre analyse porte sur un champ plus limité par rapport à celui de B. Bojar qui a étudié tous les prédicats donnant « une information sur l’information » : nous ne nous occupons en effet que des cas où cette information est transmise au moyen de la parole, réduisant la classe des sujets verbaux à la catégorie des êtres humains.

La définition ci-présentée ne s'avère toutefois pas assez restrictive, ce qui a posé le problème d'un choix adéquat parmi les unités du corpus. En effet, elle permet d'inclure parmi les verbes de parole aussi bien les verba dicendi proprement dits (...) que les verbes au sémantisme plus large pouvant être considérés comme verbes de parole dans l'une de leurs acceptions (...)'

Jamrozik, 1992

Bien que donnant les informations essentielles pour notre nomenclature, Jamrozik offre une définition des verbes de parole en tant qu'actes de langage avant toute chose. Pour notre part, seule la transmission d'un message au moyen de la parole forme le sème commun à tous les verbes que nous regroupons sous l'appellation *verbes de parole*.

Par ailleurs, la restriction du troisième paragraphe entre les « verba dicendi proprement dits » au détriment des « verbes au sémantisme plus large » ne nous paraît pas fondée dans une perspective de linguistique contextuelle. Les premiers sont également des verbes polysémiques (*ça ne me dit rien, voilà un exemple parlant, etc.*) et ne sont « verbes de parole proprement dits » qu'en fonction de leur acception première¹³⁸. C'est pourquoi nous ne ferons pas de distinction initiale entre les deux catégories proposées par Jamrozik, et que nous ne traiterons que les acceptions et occurrences de chaque verbe qui dénotent une activité de parole **en contexte**. Partant de là, la typologie des verbes observés découle des occurrences d'un point de vue linguistique centré sur le corpus.

Une autre nomenclature des verbes de parole nous est proposée par Michel CHAROLLES et Béatrice LAMIROY dans les travaux préliminaires au Projet De Recherche ILF « transitivity » initié en 2002 et dirigé par Bernard COMBETTES¹³⁹, projet auquel nous avons contribué en proposant notamment les travaux que nous présentons ci-après. Pour ces auteurs, il s'agissait d'étudier le fonctionnement de la transitivity à partir d'une « champ sémantique relativement homogène », en l'occurrence les verbes de parole. En guise de définition, ils précisent :

¹³⁸ Nous admettons bien volontiers que les sens des verbes qui ne correspondent pas à un prédicat de parole sont minoritaires et ne correspondent qu'aux emplois métaphoriques de ces verbes. Notre propos est surtout de ne pas séparer les verbes de parole proprement dits de ceux qui ne sont verbes de parole que justement dans leurs emplois métaphoriques (comme c'est généralement le cas entre autres catégories pour les verbes qui caractérisent les cris d'animaux, *aboyer, bramer, glapir, etc.*).

¹³⁹ PDR qui incluait, outre Bernard Combettes, Michel Charolles, Béatrice Lamiroy et nous-même, la participation active de Pierre Le Goffic et Laure Sarda ainsi que la participation occasionnelle de Jacques François, Morgane Sénéchal et Guillaume Jacquy.

‘Les verbes de parole (...) incluent dans leur représentation conceptuelle profonde un objet effectué, à savoir les paroles produites. En effet, un verbe de parole est défini ici comme un verbe ne satisfaisant pas au test :

N0 V (E + (de + E) N1 + (de ce + E) Que P) *sans dire ou écrire un mot*

Prototypiquement, le sujet d’un verbe de parole est de type Nhum, c’est-à-dire agentif, alors que l’objet est au contraire inanimé. Par ailleurs, la production de paroles est en général – sauf exception, dans des cas comme *délirer*, p.ex. – une activité contrôlée par le sujet parlant, et téléique, puisqu’elle a la production langagière pour objet.’

Charolles & Lamiroy, 2002

En premier lieu, cette définition ne s’appuie pas sur la théorie des actes de langage. Peu importe l’intention du locuteur, ni même la présence ou non d’un interlocuteur, ce qui compte est ici l’activité elle-même, la production de paroles en tant qu’activité. Comme Jamrozik, ils précisent que le sujet doit être de type *NomHumain*, bien que le mentionnant avec les précautions « prototypiques » d’usage. Aucune séparation préliminaire n’est faite entre les « verba dicendi proprement dits » et les « verbes au sémantisme plus large. » Charolles et Lamiroy admettent sans le dire que tout verbe dont une acception satisfait à leur test syntaxico-sémantique peut figurer dans la nomenclature. Cette dernière a été élaborée essentiellement à partir des tables de Gross (1975). Leur nomenclature présente l’avantage de mettre tout de suite en avant une dichotomie verbes transitifs – verbes intransitifs, en vertu de l’analyse initialement souhaitée.

Dans sa version la plus générique, la classe des verbes de parole s’organise autour de deux verbes noyaux, *dire* et *parler*, dont les caractéristiques sont fort dissemblables : *dire* ne se conçoit pas en principe sans un objet qui est le contenu de la parole ; *parler* est habituellement perçu comme sémantiquement intransitif, potentiellement porteur d’arguments.

Bien que la dichotomie *dire/parler* soit la plus évidente pour caractériser le paradigme des verbes de parole, il existe d’autres verbes permettant un découpage plus fin. Manguin, François et alii (2004) proposent la sélection de six verbes « susceptibles d’exprimer une variété de modalités énonciatives¹⁴⁰ » : *appeler, conseiller, crier, demander, dire, parler*.

¹⁴⁰ Par « modalité énonciative », ces auteurs entendent « un type de relation entre un locuteur, un allocataire et/ou une déclaration, une injonction, une question, une requête ou simplement un propos de conversation. » Cette définition implique la prise en compte des actes de parole en plus de la simple distinction contenu / activité propre à la distinction entre *dire* et *parler*.

Partant des définitions précédentes, nous avons souhaité écrire la notre le plus précisément possible, mais en restant assez général afin de ne laisser échapper aucun verbe potentiellement *de parole* tout en excluant certains candidats douteux¹⁴¹.

Les principes d'établissement de la grille que nous proposons sont conçus pour une liste de verbes de parole assez différente de la nomenclature établie par Charolles & Lamiroy (2002), laquelle contient environ 600 verbes répartis entre deux catégories distinctes : intransitifs (150) et transitifs (+/- 450).

Afin de constituer une nouvelle nomenclature, nous avons choisi de fonder notre sélection sur le critère suivant : « *Pour un sujet sémantique Humain réalisé ou non, produire du langage humain*¹⁴² **décodable oral ou écrit** ».

Ce critère comporte le sème de production de langage humain, et celui-ci doit être explicite pour qu'un verbe soit retenu dans la liste. Comme Charolles & Lamiroy, nous considérons que les emplois ponctuellement « introducteurs de parole » de certains verbes comme *haleter*, *postillonner*, *éternuer*, etc. ne permettent pas à ceux-ci d'être désignés comme verbes de parole, d'autant que les corpus littéraires montrent que de nombreux prédicats sans rapport avec le langage peuvent introduire des paroles rapportées.

Dans certains cas, les verbes sélectionnés ne sont pas des verbes de parole en tant que tels et ne le deviennent que par ajout d'arguments adaptés (ex : *tenir* dans *tenir des propos...*). Pour ceux-là, le sujet profond doit demeurer [+Humain] et ils peuvent se retrouver classés dans la catégorie *Prédicats de parole composés* ou *Locutions verbales de parole* selon leur degré d'éloignement du critère de base. Dans la mesure du possible, nous intégrons à notre nomenclature des lexèmes monolexicaux mais les prédicats de parole composés n'en sont en rien exclus.

Pour qu'un verbe soit retenu dans la nomenclature, il faut et il suffit qu'une au moins de ses acceptions soit reconnue comme respectant le critère général, ceci afin de n'établir aucune hiérarchie *a priori* entre les prédicats de parole prototypiques et les autres.

Le critère argumental que nous mettons en avant, « *pour un sujet sémantique humain* », doit être augmenté des médias de la parole exprimés par métonymie¹⁴³ : *journaux*, *radio*, *télévision*, etc. Plutôt que « *sujet sémantique* », il faudrait ici parler d'« *origine de la production de parole* ». De même, *sujet humain non réalisé* est à comprendre au sens syntaxique ; en effet, il doit être toujours possible de rétablir

¹⁴¹ Ce qui supposait de poser des critères suffisamment clairs et complets pour que le doute n'ait plus prise sur notre nomenclature.

¹⁴² *Un sujet humain doit produire du langage humain* : cette redondance est nécessaire car un humain peut, par imitation ou pour toute autre raison, produire du langage animal ou des sons qui n'ont rien à voir avec du langage articulé.

¹⁴³ On rencontre quelques occurrences de ce type dans les textes littéraires, « le haut-parleur qui hurle des ordres » ou « la radio qui braille des nouvelles » pour les cas les plus simples.

sémantiquement le sujet-agent du verbe de parole, soit explicitement dans la phrase (formes passives à complément d'agent¹⁴⁴), soit implicitement (formes passives sans complément d'agent¹⁴⁵), soit extrinsèquement à la phrase (sujet anté-réalisé¹⁴⁶, etc.)

Le critère principal « **produire du langage humain décodable oral ou écrit** » est peut-être trop restrictif dans la mesure où certains verbes exprimant l'absence de parole fonctionnent comme des verbes de parole produite. Par ailleurs, la forme négative de la plupart de ces verbes permet de créer un prédicat verbal de parole : *taire/ne pas taire, cacher/ne pas cacher, dissimuler/ne pas dissimuler*. Aussi, si *ne pas dire* est analysé au même titre que *dire*, il apparaît pertinent d'étudier *taire* au même niveau que *ne pas taire*¹⁴⁷. Concernant notre corpus, le parti pris initial était d'y faire figurer les formes négatives de ces verbes qui fonctionnent « à l'envers », mais il est évidemment difficile d'en occulter la forme affirmative dans l'analyse.

De tous ces commentaires, nous extrayons la définition finale que nous proposons et qui a été adoptée dans les grandes lignes par nos collègues :

« Pour un sujet sémantique +Humain réalisé, exprimé par métonymie ou récupérable comme agent d'une forme passive, un verbe est potentiellement un verbe de parole si une au moins de ses acceptions se rapporte explicitement à la **production de langage humain décodable, oral ou écrit**, y compris pour une forme négative dudit verbe et dans le cas où le verbe sert de support à un objet conforme à un contenu de parole, hors paroles rapportées. »

Les verbes utilisés en incise ou comme introducteurs de paroles rapportées, et n'étant pas dans leur définition même des verbes de parole selon les critères retenus, ne figurent pas dans la nomenclature finale.

Exemple tiré de Frantext :

(...)papier pour y inscrire au crayon - on nous le fournirait - le numéro de chaque wagon destiné à l'O..T. - Les numéros sont là, vous voyez ? Nos yeux absents, abrutis. Herr Lindt **toussota** : - C'est un peu... imprévu pour vous, n'est-ce pas ? Mais c'est la guerre, la guerre ! (...)
Schreiber Boris /Un silence d'environ une demi-heure/1996

À moins d'une indication contraire dans le dictionnaire, *toussoter* n'entre pas dans la nomenclature en tant que verbe de parole puisqu'il ne satisfait pas la condition sémantique initiale. De nombreux autres verbes, comme *éternuer*, *se moucher* peuvent évoquer la parole par analogie et introduire de la parole sans surprendre le lecteur. Toutefois, avoir initié du langage dans le corpus n'est pas un critère suffisant pour être

¹⁴⁴ *Il a été dit par le président qu'il ferait beau demain.*

¹⁴⁵ *Il a été dit qu'il ferait beau demain.*

¹⁴⁶ Dans certains cas de discours direct ou indirect libre notamment.

¹⁴⁷ C'était du moins l'avis unanime des membres du PDR sur la question.

considéré comme un verbe de parole¹⁴⁸, d'autant que tous les verbes nous paraissent avoir le potentiel suffisant pour jouer ce rôle, avec un minimum d'imagination :

« *Bonjour*, ferma-t-il la porte. »

(suggéré lors d'une réunion du groupe ILF « transitivité », 2004)

Étant entendu que les nomenclatures existantes ne présentent pas de manière exhaustive tous les verbes de parole respectant nos critères, étant donné par ailleurs qu'il n'existe pas de moyen automatisé de les répertorier lexicographiquement ou par le corpus, nous avons choisi en dernier ressort de faire une recherche systématique dans le dictionnaire en lisant un à un tous les articles. Bien que particulièrement coûteux en temps et en énergie, ce travail de fourmi est probablement le seul à même de nous permettre d'aboutir à l'exhaustivité dans la construction de nos bases, à l'image de la construction des Lexiques Inversés présentée au début de ce chapitre.

Au final, nous avons extrait 869 verbes de parole dont la liste est présentée ci-dessous. Il est à noter qu'une fois encore la sélection que nous avons opérée s'appuie exclusivement sur les données proposées par le dictionnaire. Ce faisant, nous excluons éventuellement certains verbes rares et nous en acceptons certainement dont les emplois en tant que prédicats de parole sont désuets ou un peu forcés.

Conformément à la manière dont nous l'avons constituée, cette nomenclature se présente initialement sous la forme de plusieurs sous-listes : verbes intransitifs, verbes avec objets directs syntagmes nominaux, verbes avec complétives¹⁴⁹, verbes et compléments d'objet indirect syntagmes prépositionnels¹⁵⁰. Chaque verbe peut apparaître dans plusieurs des sous-listes mais n'est présent qu'une fois dans la nomenclature récapitulative ci-dessus.

Nous faisons remarquer que la présence de prépositions et de complémenteurs dans la liste finale correspond à la façon dont le dictionnaire répertorie le prédicat en tant que se rapportant à la parole : par exemple, *tourmenter* n'a pas d'acception de verbe servant à dénoter la production de parole, celle-ci n'est récupérable que derrière la sous-entrée *tourmenter de*.

Le point de départ à notre analyse de la classe des verbes de parole repose sur une intuition préalable. En effet, Jamrozik (1992) et Charolles & Lamiroy (2002) établissent des typologies de verbes fondées sur des propriétés intrinsèques, et en déduisent une séparation *de facto* entre les verbes transitifs et intransitifs.

¹⁴⁸ Cela n'empêche pas une telle occurrence d'être considérée comme un prédicat de parole et le lexème verbal comme un verbe introducteur de paroles rapportées. Néanmoins, cela n'octroie pas à ces verbes un sens habituel de « verbes de parole » susceptible d'apparaître tel quel dans un dictionnaire.

¹⁴⁹ Avec précision du complémenteur.

¹⁵⁰ Avec précision de la préposition.

Pour notre part nous avons imaginé, hors de toute observation d'exemples attestés, qu'il n'y avait pas une distinction aussi tranchée entre la classe des transitifs et celle des intransitifs. En poussant plus loin l'hypothèse, nous pensions que la plupart des verbes devaient entrer potentiellement dans les deux catégories et que seuls quelques-uns, les plus irréductibles, résisteraient au phénomène d'appartenance potentielle aux deux classes selon les emplois.

Afin de vérifier cette hypothèse, nous avons choisi une nouvelle fois d'utiliser la base littéraire FRANTEXT avec un objectif principal double :

- montrer que l'articulation transitif versus intransitif n'est pas le meilleur moyen de catégoriser les verbes de parole ;
- mettre en exergue l'ensemble des traits syntaxiques et sémantiques propres à chaque verbe en contexte.

La prééminence éventuelle de la forme sur la sens, et vice-versa, est un thème récurrent dans notre recherche. Montrer que l'un prime sur l'autre ne fait pas partie de nos objectifs, ce qui n'empêche pas cette question de se poser avec acuité dès lors qu'il s'agit de relever et de décrire tous les phénomènes mis en œuvre dans la caractérisation d'une occurrence lexicale.

Tout comme Jamrozik, nous considérons qu'il n'y a pas de relation hiérarchique entre les deux et que forme et sens constituent un tout infrangible caractéristique de ce qu'est le langage humain.

Dans le cadre d'une analyse syntaxique puis syntaxico-sémantique des verbes de parole en français, nous arrivons assez facilement à un découpage en seulement une dizaine de constructions possibles, en effectuant des regroupements le cas échéant. La liste probablement exhaustive des constructions observées dans le corpus est la suivante, exemples à l'appui :

Schématiquement, les verbes de parole sont susceptibles de sous-catégoriser trois grands types d'arguments : *à qui ou quoi les paroles sont adressées, quel est le contenu ou le thème des paroles produites et comment ou avec quels matériaux les paroles sont proférées*. Cette « trichotomie » nous a conduits, sur notre proposition personnelle et avec l'aval critique du PDR, à définir les trois classes suivantes :

- *interlocuteur*, le destinataire de la parole ;
- *contenu*, comme son nom l'indique se rapporte au contenu des paroles, de quoi on parle, quel est le thème du discours ;

- *métalangage*, qui donne une information sur la forme mais pas directement sur le contenu des paroles produites.

A ces trois classes, il faut sans doute ajouter *l'origine du discours*, le sujet sémantique évoqué précédemment, lequel est aussi un argument des verbes de parole et peut également être réalisé sous la forme d'un complément dans les constructions passives, d'où la création d'une quatrième classe :

- *locuteur*, le destinataire, l'émetteur du discours.

La grille d'analyse syntaxico-sémantique et les résultats de l'analyse effectuée dans ses cases sont présentés en annexe.

Réalisée sous tableur Excel, la grille d'analyse est constituée de quatre sous-tableaux décrits séparément :

- le tableau syntaxique pour la nature du sujet, des compléments et éventuellement des circonstants, ainsi que des informations plus spécifiques à propos des pronoms ou des prépositions le cas échéant ;
- le tableau sémantique des objets conformément à ce que nous avons décrit plus haut ;
- le tableau sémantique du verbe qui se rapporte aux informations non syntaxiques portées par le prédicat verbal indépendamment de ses arguments¹⁵¹ ;
- le tableau étymologie¹⁵².

Afin de répondre à une remarque qui nous a été faite plusieurs fois, nous précisons que cette grille d'analyse n'a pas été réalisée dans l'ignorance des tables de Maurice Gross, et n'ont pas davantage comme objectif de les supplanter, ni même de les concurrencer. Le fait d'avoir voulu élaborer nos propres outils résulte de trois paramètres acceptés par l'ensemble des membres du projet ILF dont nous donnons la liste par ailleurs : d'une part, il est plus aisé pour nous de coder les informations liées aux verbes étudiés dans un format qui nous est personnel ; deuxièmement, nos critères syntaxiques sont adaptés à la structure argumentale des verbes de notre nomenclature, comprenant le discours indirect, les paroles rapportées, les didascalies, etc., et excluent *de facto* les formes non rencontrées ; enfin, notre grille propose également le codage

¹⁵¹ Cette partie a été développée par Michel Charolles et Béatrice Lamiroy dans les travaux préliminaires du PDR.

¹⁵² Réalisé par Bernard Combettes.

de critères sémantiques (nature des objets), pragmatiques et étymologiques que les tables de Gross ne présentent pas.

Par surcroît, la forme des grilles et leur support progiciel permettent d'envisager tout ajout et toute modification jugés utiles lors de l'analyse.

Le nombre très élevé des verbes de parole en français ne nous a pas permis d'effectuer une synthèse complète des données lexicographiques les concernant¹⁵³. Toutefois, l'intérêt d'observer le paradigme dans son ensemble réside davantage dans les disparités et les similarités de leurs structures argumentales sémantiques, et dans les constructions syntaxiques attenantes.

A l'opposé, le cas d'un verbe comme *traverser* (cf. Haton & Mascherin 2005) est beaucoup plus facile à envisager dans la mesure où ce verbe ne possède qu'une seule structure argumentale profonde, Sujet-Verbe-Objet direct. Conséquemment, l'analyse que nous avons faite de *traverser* était essentiellement sémantique, les données syntaxiques n'intervenant que pour la description éventuelle d'occurrences complexes, par exemple les sujets de métaphores sous la forme de complétives infinitives ou de subordonnées.

La pré-analyse du corpus a confirmé notre hypothèse de départ, à savoir que la plupart des verbes de parole sont susceptibles d'être utilisés dans des formes transitives. Toutefois, les différences de fréquence dans l'apparition de ces formes sont assez nettes et méritent une attention particulière.

A titre de premier exemple, nous citerons le cas du verbe *aboyer*, strictement intransitif pour ses emplois habituels (cri du chien) et également classifié comme verbe de parole intransitif par Charolles et Lamiroy. Les occurrences d'*aboyer* comme verbe de parole dans les textes littéraires sont très majoritairement transitives, avec des constructions syntaxiques très diverses :

Le chauffeur venait d'opter pour le grand escalier quand une concierge aboya :
L'escalier de service, c'est pour les chiens?
Sabatier Robert/Trois sucettes à la menthe 1972

Ils ne réagissaient pas du tout, elle tira le premier, colla une petite bourrade au
second, les enfourna dans le jardin. Et vous allez voir ce que vous allez voir!
aboya-t-elle,
elle regarda vers la route.
Rochefort C. /Encore heureux qu'on va vers l'été 1975

¹⁵³ Notons toutefois que nous avons dépouillé les données d'un nombre important de verbes, en particulier pour ceux qui ont des contenus sémantiques explicites comme les verbes de cris d'animaux. Nous disposons ainsi d'une grille syntaxico-sémantique remplie pour quelques milliers d'occurrences. Bien que les conclusions que nous présentons ici soient très générales, elles sont issues de l'observation d'un nombre important de données analysées ou vérifiées par nos soins.

C'est le numéro à côté... Nettement moins briquée, la courette... Nous revoilà
entre murs et débris de la répression de base...
Un képi méprisant m'aboie le chemin.
Degaudenzi Jean-Louis /Zone 1987

Il appuya sur l'interphone, aboya à une des secrétaires de préparer la lettre de
licenciement de la "mal-baisée de l'entrée, me souviens pas de son nom"
Vergne A. /L'Innocence du boucher 1984

Il y a encore quelques années, le sens d'aboyer en tant que verbe de parole était presque totalement ignoré des dictionnaires. Depuis peu, cette acception est apparue (encore un peu plus tard dans les ouvrages bilingues), et depuis l'an dernier au moins deux dictionnaires font état d'un transitif indirect *aboyer contre/après* fonctionnant pour les chiens comme pour les humains. En ce domaine, on voit bien que l'usage précède, et de loin, la grammaire et le lexique normatifs.

Une telle richesse syntaxique se rencontre pour la plupart des verbes de cris d'animaux qui deviennent métaphoriquement des prédicats de parole humaine, lesquels sont abondamment représentés dans la base littéraire.

De notre point de vue, cela est probablement dû au fait qu'il s'agit de verbes qui n'étaient pas prévus pour cela à leur apparition, ce qui laisse une certaine liberté dans leurs emplois ultérieurs dès lors que la métaphore est lexicalisée. On voit bien pour *aboyer* que les contraintes habituelles sur les verbes semblables à *dire* d'une part et ceux semblables à *parler* d'autre part ne s'appliquent pas. Selon les cas, *aboyer* peut être assimilé à l'un ou l'autre de ces deux verbes dominant le paradigme. Il est également possible de le rapprocher des autres verbes de référence proposés par Manguin et François (2004), *appeler*, *conseiller*, *crier* et *demander* avec une préférence marquée pour le troisième, *crier*, ce qui peut s'expliquer par les liens proposés par les dictionnaires de synonymes qui ont servi de base aux travaux du CRISCO sur la synonymie lexicale : tous les verbes se rapportant aux cris d'animaux (*hurler*, *beugler*, *brailler*, *braire*, *bramer*, *criailler*, *glapir*, *piailler*, *rugir*) se retrouvent comme synonymes de *crier* et d'aucun autre des cinq verbes de référence. Notons qu'étonnamment *aboyer* n'est pas présent dans cette liste.

Le phénomène que nous relevons pour *aboyer* vaut pour presque tous les verbes de la même famille dans FRANTEXT. Cela est également le cas pour les verbes dits « d'élocution » ou ceux qui d'une manière générale se rapportent explicitement au contenu de la parole dans le lexème verbal. Cette classe a pour particularité de véhiculer ses propres « circonstants affectant la parole ». Dans ses emplois en tant que verbe de parole, *aboyer* renvoie ainsi à *dire en aboyant* aussi bien qu'à *parler en aboyant*, voire à *appeler/ ?conseiller*¹⁵⁴/*demander en aboyant* ou à *crier comme en aboyant*.

¹⁵⁴ Contrairement aux autres verbes de référence choisis par l'équipe du CRISCO, *conseiller* semble assez mal fonctionner pour gloser les emplois du verbe *aboyer*, sans

Il ne faut pas occulter qu'un des écueils à une étude globale des verbes de parole est la sous-représentation de certains d'entre eux dans FRANTEXT.

On n'y dénombre aucune occurrence pour de nombreux verbes comme *criticailler*, *écrivailier*, *joualiser*, *pindariser*, *réciproquer*, etc., ce qui rend impossible leur classification autrement qu'en suivant les données des dictionnaires. Par ailleurs, de nombreux autres verbes de parole sont peu représentés (moins de dix occurrences), soit parce qu'il s'agit de verbes rares, soit parce qu'ils ne sont pas souvent employés avec l'acception de production de parole. Pour ceux-là également, il n'est pas évident de tirer des conclusions linguistiques pertinentes.

La conclusion qui s'impose à nous à la lecture des résultats obtenus est que la plupart des verbes de parole bien représentés dans la base FRANTEXT peuvent être manipulés indifféremment selon les deux schèmes de la dichotomie précitée. Seuls quelques verbes irréductibles (comme *bavarder*) ne peuvent être utilisés que dans l'une des deux formes. Conséquemment, le dépouillement du corpus nous permet de dégager une nouvelle répartition des verbes selon de nouveaux critères moins restrictifs qu'une simple opposition activité/contenu, l'activité de parole étant propre au verbe *parler* et le contenu se rapportant à l'argument second de *dire*.

Toutefois, bien qu'il soit particulièrement riche en occurrences originales et apparemment pertinentes, nous estimons que le corpus n'induit pas de jugement péremptoire sur leur caractère valide ou invalide au sens linguistique. En revanche, il constitue à l'évidence un excellent indicateur de productivité lexico-sémantique, en cela que les auteurs - et les locuteurs d'une langue en général - peuvent employer nombre de mots dans des acceptions et des constructions intelligibles que les grammaires traditionnelles et la lexicographie ne prévoient ou ne mentionnent pas, repoussant ainsi les limites du possible et de l'impossible langagiers (cf. Bronckart 2000, Rousseau 2000). De fait, un corpus actualisé, comme l'est FRANTEXT, est un bon indice de l'évolution du lexique, d'autant plus que les dictionnaires et les grammaires évoluent et s'enrichissent à son contact (il suffit d'observer la réactualisation permanente des acceptions dans le Petit Robert pour s'en convaincre).

L'analyse des verbes de parole fait ressortir essentiellement de nouvelles constructions syntaxiques¹⁵⁵. Dans ce cas, il n'y a pas lieu de modifier les structures dans l'algorithme que nous envisageons et allons incessamment décrire ; il suffit d'enrichir la base de données pour les prendre en compte.

Le traitement de ces nouvelles données trouve sa particularité dans le fait qu'aucune traduction n'est sélectionnée ni aucun synonyme en langue source

doute à cause de l'impossibilité de trouver une connotation négative à *conseiller* et a contrario une connotation visiblement positive à *aboyer*.

¹⁵⁵ Les constructions syntaxiques dites nouvelles sont directement perceptibles. En revanche, il n'est pas aisé de mettre en exergue des sens nouveaux pour l'ensemble des verbes.

directement perceptible puisque nous travaillons sur une base littéraire du français sans renvoi à d'autres langues et sans paraphrase ou glose accessible. De ce fait, les structures qui en découlent sont implantées dans le graphe sans lien initial de synonymie. L'intérêt d'avoir des structures isolées réside dans la possibilité d'enrichir le graphe de manière collaborative ou indirectement par le jeu du miroir lexicographique. Nous pouvons également tenter de l'enrichir nous-même par le biais de notre seule intelligence, avec les risques que ce la comporte.

Prenons le cas du verbe *bramer*, majoritairement transitif dans le corpus FRANTEXT (en tant que verbe de parole) mais strictement intransitif d'après nos dictionnaires bilingues. D'après ces derniers, les traductions anglaises de ce verbe sont *to bawl* (*brailler* comme synonyme LS¹⁵⁶) et *to wail* (*se lamenter* comme synonyme LC¹⁵⁷).

Par enrichissement collaboratif, il est envisageable de proposer à un utilisateur de nos bases toutes les traductions de *bramer* et leurs équivalents dans le miroir. Un locuteur du français n'aura aucun mal à déterminer que *brailler* est potentiellement transitif direct alors que *se lamenter* l'est beaucoup plus difficilement. *To bawl* apparaît dans ce cas comme la meilleure traduction pour une occurrence telle que :

« (...) elle resurgit toujours, intacte, après avoir bramé pendant trois jours que la vie lui est insupportable (...) »
(Dormann Geneviève / La Petite main /1993).

Par le jeu du miroir, nous remarquons que *to bawl* (la première traduction) est potentiellement transitif, ainsi que *brailler*. Par ailleurs, *se lamenter* et *to wail* sont tous les deux intransitifs, ce qui nous conforte dans l'idée que les premiers sont les plus appropriés pour la traduction d'une occurrence telle que ci-dessus.

Les constructions et les sens qui n'apparaissent pas dans les dictionnaires qui nous servent à construire la base de données peuvent être incluses de la manière suivante :

Figure 9 : Fragment de CSMU pour les verbes de parole à partir de données attestées

mot-source	mot-cible	syn LS	syn LC	colloc objet	colloc sujet	constr LS	constr LC	registre	infos complémentaires
Faboyer					Humain	Obj-discours indirect			Production de parole
Faboyer					Humain	Obj-discours rapporté.			Production de parole

¹⁵⁶ Synonyme en langue source, dont nous donnons la définition par ailleurs ainsi que dans le glossaire des acronymes en fin de document.

¹⁵⁷ Synonyme en langue cible, dont nous donnons également la définition par ailleurs ainsi que dans le glossaire des acronymes en fin de document.

						incise			
Faboyer				[contenu de parole]	Humain				Production de parole
Faboyer				Humain	Humain	À obj			Production de parole
Faboyer				[contenu de parole]	Humain	De Inf			Production de parole
Faboyer				[contenu de parole]	Humain	Que Prop			Production de parole
Fbramer	Abawl	Brailier		[contenu de parole]	Humain	Que Prop	<i>That Prop / to vb</i>		Production de parole
Fbramer	Awail	Se lamenter			Humain				Production de parole

Rien n'exclut a priori de laisser vierges les cases de traductions dans la mesure où la base de données est destinée à demeurer ouverte à tout enrichissement ultérieur. Avec un peu d'audace, nous aurions peut-être pu remplir le champ *mot-cible* en jouant sur les variations possibles du verbe anglais *to bark*. Cependant, nous estimons que cela n'est pas notre rôle dans le cadre de ce sujet. À ce sujet, nous pouvons poser la question des mémoires de traduction qui existent pour de nombreuses œuvres littéraires traduites ainsi que les ressources des aligneurs multilingues qui contiennent des données de nombreuses sources qui ne respectent pas à la lettre les grammaires et les dictionnaires. Ce que nous saisissons dans nos bases est un ensemble de « mémoires de faits de langue » pour les exemples attestés et des « mémoires de correspondance interlangue » pour les données lexicographiques. L'apport des aligneurs et des « mémoires de traduction » à l'ensemble permettra d'enrichir les premières et éventuellement de valider les secondes.

Pour ce qui concerne *aboyer* en tant que verbe de parole transitif, il existe des traductions répertoriées dans des concordanciers bilingues qui sont généralement *to snail at* (littéralement : *montrer les dents devant/à/contre*), ce qui était imprévisible à partir de nos connaissances de la langue ou des données des dictionnaires. De ce fait, nous pensons que l'utilité de conserver la mémoire des faits de langue non traduits est grande, et l'intérêt de travailler ultérieurement avec des aligneurs/concordanciers de textes le sera tout autant.

CHAPITRE III

Étude autour du verbe *traverser*. Compléments et complémentarité des données contextuelles et lexicographiques.

1. La version lexicographique de *traverser*

1.1 Des acceptions de *traverser*

1.1.1 *Traverser* ressemble à *aborder*

Comme pour *aborder* au chapitre II, nous remarquons la présence dans les significations de *traverser* de deux acceptions concurrentes en tant que verbe de mouvement (avancée ou percée dynamique), ainsi qu'une acception métaphorique principale motivée par les précédentes (traversée d'une période temporelle).

Cependant, *traverser* admet des emplois statiques plus récents, lesquels sont incompatibles avec les emplois d'*aborder* qui sont exclusivement dynamiques ou apparentés, y compris dans les sens métaphoriques si l'on admet une certaine abstraction dans la conceptualisation du mouvement. Par ailleurs, *traverser* sert également pour des occurrences de parcours temporels, ce qu'encore une fois *aborder* admet plus difficilement même s'il semble tout à fait possible d'*aborder une crise* ou *une période* d'un type particulier, ce qui induit une avancée du temps (avec aspect ingressif) au même titre que les traversées.

En outre, l'intérêt d'étudier *traverser* réside dans les nuances aspectuelles que les différents emplois du verbe font apparaître, et dans l'influence que ces nuances ont sur le choix de la bonne traduction en contexte. Les premiers verbes que nous avons présentés auparavant n'ont pas la même subtilité dans la caractérisation de l'aspect, bien que leurs emplois n'en soient pas dénués pour autant.

1.1.2 Histoire du verbe

1.1.2.1 Évolution des acceptions du verbe

Données issues une nouvelle fois du TLF :

Étymol. et Hist. **A. 1.** Fin ^x s. fig. « percer de part en part » (*Passion*, éd. d'Arco Silvio Avalle, 338); **2.** 1580 « se frayer un passage au travers » (GARNIER, *Antigone*, éd. W. Foerster, III, p. 58); 1839 fig. « pénétrer subitement au plus profond de l'être » (BALZAC, *Béatrix*, p. 201); **3.** 1718 part. passé « avoir les vêtements mouillés jusqu'à la peau » (*Ac.*). **B. 1.** *Ca* 1100 *traverser un pont* (*Roland*, éd. J. Bédier, 2690); **2.** *ca* 1140 « parcourir un espace d'un bout à l'autre » (*Pélerinage Charlemagne*, éd. G. Favati, 101); **3.** 2^e moit. du ^{xiii} s. « franchir dans le sens transversal quelque chose qui s'étend en longueur » (*Gaufrey*, 117 ds T.-L.); **4. a)** 1580 (de choses sans mouvement) « s'étendre, s'allonger au travers de » (B. PALISSY, *Disc. admir.*, p. 180); **b)** 1583 (de choses mobiles) « se déplacer, aller à travers » (GARNIER, *Les Juifves*, éd. W. Foerster, III, p. 163); **5.** 1808 « se trouver dans une période caractérisée par un certain état » (DELILLE, *Trois règnes*, III ds LITTRÉ); 1839 *traverser les âges* (LAMARTINE,

Recueils poétiques, I, XXV (éd. Garnier, 1925) ds ROB. 1985). **C. 1.** 1174-76 fig. « faire obstacle, s'opposer à, empêcher » (GUERNES DE PONT-SAINT-EMAXENCE, *St Thomas*, éd. E. Walberg, 2363); **2.** 1687 « mettre en travers en faisant obstacle » (BOSSUET, *Louis de Bourbon* ds LITTRÉ); **3.** 1694 *traverser l'ancre* (CORNEILLE); **4.** 1913 alpin. (*Écho des Alpes*, n° 12, déc., p. 530 ds QUEM. DDL t. 27). **D.** Empl. pronom. **1.** ca 1160 « aller au travers de » (*Moniage Guillaume*, 2084 ds T.-L.); **2.** 1680 (en parlant d'un cheval) « jeter la croupe d'un côté, la tête d'un autre » (RICH.). Du lat. pop. *traversare*, lat. *transversare* « remuer en travers », de *tranversus*, v. *travers*.

1.1.2.2 Historique des emplois des prépositions contenant *travers*

Les historiques des locutions prépositionnelles intégrant le substantif *travers* sont également instructifs, notamment parce qu'un prédicat verbal de type « être en travers de », « passer à travers », « aller au travers de », etc., sont des paraphrases synonymes presque parfaits de *traverser*.

Prononc. et Orth.: [tʁa vɛʁ]. Att. ds Ac. dep. 1694. **Étymol. et Hist. A. 1.** *En travers*¹⁵⁸ **a)** ca 1100 *en traver* « par un chemin de traverse » (*Roland*, éd. J. Bédier, 3239); **b)** 1160-74 *en travers* « de part en part » (WACE, *Rou*, éd. A. J. Holden, III, 8273); 1885 fig. *ça me reste en travers* (ZOLA, *Germinal*, p. 1231); **c)** 1691 mar. « se dit de la position d'un bâtiment dont la direction de la quille est perpendiculaire à la direction du vent ou du courant » (OZANAM, p. 278); 1771 mar. *mettre en travers* (BOUGAINVILLE, *Voy.*, II, p. 315 ds LITTRÉ); 1876 *s'échouer en travers* (*Lar.* 19^e); **d)** 1829 *en travers de* « de manière à contrecarrer quelque chose » (DUMAS père, *Henri III*, I, p. 1); 1834 fig. *se mettre en travers de* (BALZAC, *Œuvres div.*, t. 2, p. 669); 1861 *en travers du chemin* (de qqn) (AUGIER, *Effrontés*, pp. 373-374); **e)** 1847 « dans une position transversale par rapport à l'axe de l'objet considéré » (MÉRIMÉE, *A. Guillot*, p. 90); **2.** *à travers* **a)** 1160-74 *a travers* (WACE, *Rou*, éd. A. J. Holden, 1531); **b)** *id. travers les chans* (ID., *ibid.*, III, 4467); 1606 *à travers les champs* (NICOT);

¹⁵⁸ Les soulignements ont été ajoutés par nous a posteriori sur les données du TLF afin de faire apparaître dans le texte les occurrences pertinentes pour notre propos.

(...)

3. de travers a) *ca* 1155 « en travers de » (WACE, *Brut*, éd. I. Arnold, 5310); *ca* 1200 « d'une extrémité à l'autre dans le sens de la largeur » (*Godefroy de Bouillon*, 245 ds T.-L.); 1588 « dans le mauvais sens, pas en place »

(...)

4. au travers de *ca* 1210 « d'une extrémité à l'autre dans le sens de la largeur » (*Folque de Candie*, éd. O. Schultz-Gora, 10477); 1561 « au milieu de, parmi » (J. GREVIN, *L'Olympe*, éd. L. Pinvert, p. 281); 1667 « par l'intermédiaire de quelque chose » (RACINE, *Britannicus*, II, 2); (*L'Équipe*, 26 sept. ds PETIOT 1982);

(...)

b) 1694 « défaut de ce qui n'a pas une ligne normale, de ce qui est dévié » (Ac.); 1803 « défaut de l'acier dans le canon d'un fusil » (BOISTE). Du lat. *traversus*, var. de *transversus* « oblique, transversal »; fig. « contrariant »; part. passé adj. de *transvertere* « tourner vers, à travers », « transformer », « détourner ».

1.1.3 Dégrouperment des acceptions

Les acceptions de *traverser* dans un dictionnaire actuel du français¹⁵⁹ sont les suivantes :

- 1a- **passer, pénétrer de part en part, à travers** (*un corps, un milieu interposé*)
- 1b- *fig* : **passer, pénétrer de part en part, à travers**
- 2- **se frayer un passage à travers** (*des personnes rassemblées*)
- 3- **parcourir d'une extrémité, d'un bord à l'autre** (*un espace*)
- 4a- **couper, aller d'un bord à l'autre** (*une voie de communication*)
- 4b- *absolu* : **couper, aller d'un bord à l'autre**
- 4c- *en parlant de choses mobiles* : **couper, aller d'un bord à l'autre**
- 5- *choses sans mouvement* : **être, s'étendre au travers de**
- 6a- **aller d'un bout à l'autre de** (*un espace de temps*)
- 6b- **dépasser** (*un état durable*)
- 7a- **se présenter à** (*l'esprit*)
- 7b- **passer par** (*la tête*)
- 8- *équit* : **mettre de travers**
- 9- *vx* : **se mettre en travers de, s'opposer à**

¹⁵⁹ Il s'agit ici du Nouveau Petit Robert. Le TLF n'est toujours pas utilisé pour le dégroupement pour les raisons énoncées au sujet du verbe *aborder*.

1.2 Des traductions de *traverser*

1.2.1 Les traductions peu marquées¹⁶⁰ sémantiquement

La traduction en anglais à la fois la plus générique et la moins connotée est *to cross*. Ce verbe ne s'emploie toutefois que pour les parcours statiques ou dynamiques sur des surfaces planes ou assimilées. Il ne convient pas pour traduire le verbe *traverser* lorsque celui-ci fait référence à des transpercées ou à des traversées difficiles sur des terrains vus comme des entités en trois dimensions. L'analyse linguistique fine que nous proposons dans la suite du chapitre mettra l'accent sur cet aspect (au double sens du terme du mot) des choses.

Pour la plupart des autres acceptions de *traverser*, le dictionnaire propose *to go through* comme candidat principal. Bien que commun à un nombre important de sens de *traverser*, *to go through* est ostensiblement plus marqué que *to cross* de par la présence de la préposition *through*, laquelle est une extension sémantiquement pleine du lexème verbal. Nous détaillerons les rôles sémantique et aspectuel de cette préposition et d'autres au cours de ce chapitre, et de manière encore plus fouillée dans la partie entièrement consacrée à l'étude de *traverser* en contexte langagier.

1.2.2 Les traductions marquées¹⁶¹

Par l'observation du dictionnaire, les traductions plus spécifiques peuvent être séparées en deux catégories principales :

- les verbes monolexicaux dont le morphème principal est sémantiquement corrélé au lieu où s'effectue la traversée ;
- les verbes à particule, contenant un verbe de déplacement (dans la plupart des cas) et une préposition qui véhicule un ou plusieurs aspect(s) particulier(s), en référence à la configuration spatiale du terrain traversé ou en fonction de l'accomplissement ou non - au sens aspectuel – de la traversée.

¹⁶⁰ Par « traduction peu marquée », il faut comprendre « un terme en langue cible qui ne porte pas de trait intrinsèque, sémantique ou aspectuel qui prolonge le sens habituel du verbe. » (Voir à ce sujet Traits Sémantiques Intrinsèques et de Prolongement, Chapitre I).

¹⁶¹ Par opposition aux « traductions non marquées », nous appelons « traduction marquée » tout terme en langue cible qui porte dans sa structure prédicative des traits sémantiques qui prolongent l'acception du verbe qu'il traduit (TSI ou TSP). Il en est ainsi par exemple de *to sail* qui peut traduire *traverser* en précisant explicitement le véhicule utilisé pour la traversée, à savoir un navire, et un support géographique, en l'occurrence une étendue d'eau.

En termes de productivité, les verbes à particule fusionnés sont particulièrement remarquables. Même si le dictionnaire ne le mentionne pas toujours explicitement, il est fort probable que les prépositions *through* et *across*, et dans une moindre mesure *over*, peuvent servir de catalyseur pour la production lexicale de traductions de *traverser* à partir de verbes qui n'étaient pas nécessairement dévolus à cette fin.

1.3 Comparaison LI / LD en anglais

Le lexique droit des traductions de *traverser* en anglais est relativement pauvre, eu égard à la productivité étonnante des prépositions sus-citées. Cela étant, quelques exemples présupposent une productivité importante des supports verbaux de mouvement, de déplacement et d'action de façon générale.

En revanche, on retrouve dans le LI un nombre plus élevé de traductions grâce notamment à la grande productivité des prépositions de mouvement *across* et *through* et également de *over* qui n'est pas du tout représentée dans le LD, certainement par la faute d'un sémantisme aspectuel trop appuyé. Comme nous venons d'en faire l'hypothèse, beaucoup de verbes n'ayant pas d'acception proche de celles de *traverser* peuvent être traduits par ce dernier par l'adjonction de la préposition appropriée.

Figure 10 : tableau de correspondance des traductions anglaises de *traverser*

Dico droit	Indice T1	Indice T2	Dico inversé	Sens proche	Prolongement du sens
amble through		2	amble through		D'un pas tranquille
barge through		2	barge through		En bousculant
be across	5		be across		
beetle through		3-4a	beetle through		En vitesse
belt across		3-4a	belt across		À toutes jambs, à toute blinde
come across		= cross T2	come across	= cross	
come through	1a		come through		
cross	3-4a-4c-5-7	3-4	cross		
cross over		indéfini ¹⁶²	cross over	cross	
cross under	3-4a		cross under		par en-dessous
cut		5	cut	cross, intersect	
cut across	5		cut across		
cut right across	5		cut right across		
ford	3-4a		ford		À gué
get across		indéfini	get across		<i>lit</i>
get over		indéfini	get over	cross	
get through		indéfini	get through		
go across		indéfini	go across	cross	
go right through	1a		go right through		de part en part
go through	1a-6b-6a	6b	go through	suffer, endure	
have across	5		have across		
lie over		5	lie over		
live through	6a		live through		
make one's way through	2		make one's way through	passer à travers	
negotiate		indéfini	negotiate	cross	
occur to	7	7	occur to	come to mind	
pass through	6b	1-2-3-4	pass through		
run across	5	2-3-4-5	run across		en courant (LI)

¹⁶² La mention *indéfini* signifie que le dictionnaire ne précise pas pour quels emplois de *traverser* ce dernier est la traduction du verbe concerné. C'est notamment le cas pour les verbes qui sont assimilés à *to cross* par les rédacteurs mais cela n'ajoute pas vraiment de précisions sur les acceptions qu'ils traduisent réellement.

run right across	5		run right across		
sail		3	sail		en bateau
shoot through	1a-1b		shoot through		une douleur / balle (sujet)
stride across	3	2-3-4	stride across		à grands pas
swim across		3	swim across		à la nage
travel down	6a		travel down		
travel through		3	travel through		
undergo	6b		undergo		
wade across	3-4a	4a	wade across		à gué
walk across		indéfini	walk across		
zigzag through		indéfini	zigzag through		en zigzaguant

Comme c'était le cas pour *abandonner* et contrairement aux représentations du verbe *aborder*, ce tableau ne rend pas compte des différences liées à la nature des arguments, surtout en ce qui concerne les compléments d'objet. Par exemple, le dégroupement des acceptions dans le Petit Robert ne permet pas de montrer la différence de répartition des traductèmes entre *to cross* et *to go through* en rapport avec les acceptions 3 et 4 (a, b et c) alors que ces deux verbes apparaissent visiblement en distribution complémentaire dans le Robert & Collins. Le rôle majeur joué par les collocations dans le choix de la traduction n'est cependant pas suffisant pour représenter nos données de la même façon que pour *abandonner* étant donné la polysémie très marquée de *traverser* par rapport au premier. C'est notamment ce dernier argument que nous mettons en avant pour justifier l'étude linguistique particulière que nous avons menée avec Laurent Mascherin sur *traverser*. De façon générale, ce tableau (pourtant non exhaustif du côté du LI) donne une idée assez effrayante de la complexité du verbe *traverser* quant à sa traduction eu égard aux possibilités offertes par le dictionnaire. Bien que ce verbe ne fasse pas partie des plus simples à caractériser, il est très représentatif d'un phénomène généralisé à l'ensemble des dictionnaires.

1.3.1.1 Essai sur traverser

Figure 11 : Fragment du tableau de données autour de *traverser*

<i>motsource</i>	<i>LS</i>	<i>motcible</i>	<i>LC</i>	<i>synLS</i>	<i>SynLC</i>	<i>obj1</i>	<i>obj2</i>	<i>suj</i>	<i>constrLS</i>	<i>constrLC</i>	<i>Registre</i>	<i>infoscomps</i>
traverser	f	atravesar	e		cruzar	pays, rue						
traverser	f	cruzar	e		atravesar	pays, rue						
traverser	f	traspasar	e		calar	(manteau)						
traverser	f	calar	e		traspasar	(manteau)						

<i>motsource</i>	<i>LS</i>	<i>motcible</i>	<i>LC</i>	<i>synLS</i>	<i>SynLC</i>	<i>obj1</i>	<i>obj2</i>	<i>suj</i>	<i>constrLS</i>	<i>constrLC</i>	<i>Registre</i>	<i>infoscompls</i>
traverser	f	atravesar	e		pasar	(temps)						
traverser	f	pasar	e		atravesar	(temps)						
traverser	f	pasar por	e			„l'esprit“ ,la cabeza'						
traverser	f	cruzar por	e			„l'esprit“, la imaginacion'						
traverser	f	atravesar	e									« de part en part » = « de parte a parte » ¹⁶³
traverser	f	cross	e			rue, pont, montagnes, mer		personne, véhicule				
traverser	f	go through	e			ville, forêt, tunnel		personne, véhicule				
traverser	f	swim across	e			rivière						à la nage
traverser	f	take a ferry across	e		cross by ferry	rivière						en bac
traverser	f	cross by ferry	e		take a ferry across	rivière						en bac
traverser	f	ford	e		wade across	rivière						à gué
traverser	f	wade across	e		ford	rivière						à gué
traverser	f	stride across	e			pièce						à grands pas
traverser	f	cross	e		run across			pont, route				
traverser	f	run across	e		cross			pont, route				
traverser	f	cross under	e					tunnel				

2. Vers une analyse plus approfondie de *traverser* en contexte

À partir de notre pré-analyse du verbe *traverser*, il est évident qu'une étude strictement lexicographique, même étayée par l'intuition, se révèle finalement insuffisante lorsqu'il s'agit de distinguer des valeurs aspectuelles qui sont exprimées au-delà du prédicat verbal. Or, ces valeurs jouent un rôle non négligeable dans la sélection des vedettes appropriées dès lors qu'on entend les traduire.

Par ailleurs, les dictionnaires et les typologies classent les verbes en fonction de plusieurs critères bornés : leurs acceptions (contextualisées ou non), les collocations habituellement associées à ces acceptions et leur syntaxe prototypique, à défaut d'être exhaustive. Sur ce point également, l'absence de prise en compte systématique du contexte empêche d'avoir une lecture complète du potentiel sémantique et structurel des mots de la langue. Toutefois, nous constatons que l'évolution des dictionnaires

¹⁶³ Il s'agit ici de la traduction de « traverser de part en part » par « atravesar de parte a parte ». La question est de savoir si « traverser de part en part » peut être une lexie ou bien si « de part en part » n'est rien de plus qu'une information complémentaire. Cette dernière vision est plus cohérente avec le sens de l'expression qui est bien conforme au sens originel de « traverser », ainsi qu'avec son faible degré de figement.

contemporains tend vers une meilleure prise en compte de ces facteurs, et ce doublement :

- Tout d'abord, les phrases exemples sont plus nombreuses et parfois attestées.
- Ensuite, les dictionnaires que nous consultons font entrer davantage les constructions syntaxiques et les acceptions nouvelles et émergentes, certes avec parcimonie, mais avec une cohérence qui s'approche davantage de l'évolution littéraire observée dans nos données textuelles.

Afin de compléter les données et leur ordonnancement que nous présentons au début de ce chapitre, nous élargissons nos travaux à l'exploration de la base littéraire FRANTEXT¹⁶⁴, avec un objectif clair : compléter et enrichir les données lexicographiques tout en trouvant les moyens à mettre en œuvre pour que les données nouvelles s'intègrent aux structures que nous proposons. Il n'est pas question ici de remettre en cause le travail des lexicographes ni de le faire à leur place mais simplement d'associer au format du dictionnaire des analyses linguistiques sur données attestées permettant d'avoir un regard critique et objectif sur certains phénomènes linguistiques.

Nous discuterons un peu dans ce chapitre de la validité ou de l'utilité supposée d'une analyse à partir des corpus en ne perdant pas de vue que les caractéristiques de ces derniers varient grandement d'une base à l'autre. Ces questions ont été traitées dans le chapitre II et seront partiellement reprises au chapitre conclusif qui dresse le bilan de nos recherches et des outils que nous avons utilisés ou créés pour les mener. Avant tout, quelques remarques s'imposent au sujet de la base littéraire FRANTEXT :

- Il s'agit en premier lieu d'une base historique, c'est-à-dire qu'elle comprend des textes littéraires depuis l'origine de la langue française jusqu'à nos jours. Sa finalité est avant tout de servir l'ensemble de la communauté linguistique et littéraire, qu'il s'agisse des diachroniciens ou des synchroniciens, des exégètes ou des grammairiens.
- Ensuite, la base est paramétrable à l'envi. Il est ainsi possible et facile de choisir la période sur laquelle nous souhaitons travailler ainsi que le type de textes à explorer. Pour notre part, nous avons sélectionné les textes de toutes origines postérieurs à 1960, conformément à la

¹⁶⁴ Pour une réflexion globale sur la place des corpus en linguistique, voir en fin de chapitre II.

décision prise dans le cadre du projet ILF « transitivité » dirigé par Bernard Combettes et auquel nous avons participé activement.

FRANTEXT est une base de données textuelles initialement créée pour étayer le travail des lexicographes chargés de rédiger les entrées du TLF. On estime que la base a nourri près de 90 % des exemples du dictionnaire fait à l'INALF entre 1963 et 1993. Sa finalité première était par conséquent d'alimenter un ouvrage lexicographique qui se devait d'être le reflet le plus fidèle du français littéraire des 19^{ème} et 20^{ème} siècles. À l'origine, la base ne contenait que des œuvres ou des textes des deux siècles précités. Puis, petit à petit, elle a été enrichie d'œuvres plus anciennes (pour les travaux en langue ancienne réalisés à l'ATILF) ou plus récentes en fonction des besoins ponctuels des chercheurs.

FRANTEXT contient à peu près 80 % d'œuvres romanesques et 20 % de textes scientifiques, techniques voire philosophiques. A ce jour, elle comporte 3747 textes en français du 16^{ème} au 21^{ème}, 218 textes des 14^{ème} et 15^{ème} ainsi que 22 textes médiévaux. Cette somme d'œuvres numérisées représente 1,5 milliard de caractères et environ 218 millions d'occurrences de mots.

Les textes présents dans la base sont tous intégralement numérisés et balisés à partir de leur édition originale. Il est possible d'interroger la base sur n'importe quel mot fléchi (pour la version intégrale) ou sur un lemme (pour la version catégorisée). Grâce au programme STELLA mis au point par Jacques Dendien à l'INALF, de nombreuses possibilités de paramétrage sont offertes, en particulier le choix de la période visée, celui des thèmes littéraires, la possibilité de consulter des co-textes plus ou moins étendus, etc. Il est également possible de créer automatiquement des listes ou d'écrire des grammaires formelles pour faciliter la recherche.

Enfin, la base est enrichie en permanence. Les chiffres que nous avons présentés sont fort susceptibles d'avoir augmenté depuis leur saisie car le travail de numérisation s'effectue ponctuellement et continuellement, à la demande des chercheurs et des circonstances (textes pour l'agrégation ou pour des colloques).

Le hasard des participations a fait que nous avons étudié les verbes en corpus dans deux contextes tout à fait différents. Tout d'abord, nous présentons la variété aspectuelle du verbe *traverser* et de ses traductions probables à partir de données littéraires. Par ailleurs, nous avons traité dans le chapitre précédent le comportement des prédicats verbaux en contexte à partir d'une nomenclature beaucoup plus large : celle des verbes de production de parole. À la fin de cette troisième partie, nous proposerons une synthèse des données recueillies et analysées dans le but de les confronter et, idéalement, de les faire fusionner avec les données lexicographiques.

3. Pourquoi compléter les données par le corpus

3.1 Une question qui divise ou qui anime la communauté

Depuis quelques années la linguistique s'empêtre dans un débat méthodologique autour de la question des faits de langue associés aux théories ou aux outils. Il a été notamment reproché aux linguistes sans corpus de construire des exemples ad hoc en fonction de leurs besoins personnels d'un point de vue théorique au mépris de la réalité du langage (cf. Kleiber 1997).

L'enjeu central de cette polémique est de parvenir à répondre à la question : peut-on finaliser des idées ou des théories linguistiques sans recourir à des occurrences attestées du phénomène que l'on traite ? En marge de ce débat, rappelons les tenants du travail lexicologique de Mel'çuk, Polguère et alii, lesquels ont fondé leurs représentations sur des données intuitives et le revendiquent.

Outre que de nombreux congrès et colloques fleurissent un peu partout ces dernières années sur la question, nous trouvons également un grand nombre d'ouvrages reproduisant la discussion à l'échelle de la communauté. Dans Ooi (1998), nous trouvons la définition quasi-originelle de la linguistique de corpus selon Sinclair (1996), lequel propose une caractérisation de cette linguistique émergente par la distinction entre l'étude dirigée par le corpus et la linguistique fondée sur le corpus. Une troisième voie est également envisageable, comme nous l'avons suggéré au chapitre II en ce qui concerne la linguistique appliquée au corpus. À la suite de cette définition presque pionnière, des ouvrages plus récents ont mis en évidence la querelle sous-jacente qui oppose d'irréductibles partisans du « tout-attesté » à de non moins irréductibles partisans de la construction de données, ces derniers étant généralement plus discrets dans la polémique.

3.2 Le plus beau corpus du monde ne peut donner que ce qu'il a¹⁶⁵

Durant notre participation¹⁶⁶ au colloque JETOU 2005 consacré au corpus, son rôle et son utilisation, nous avons insisté sur la complémentarité des modes d'observation des données exploitables :

« (...) Toutefois, bien qu'il soit particulièrement riche en occurrences originales et apparemment pertinentes, nous estimons que le corpus n'induit pas de jugement péremptoire sur leur caractère valide/invalidé au sens linguistique. En revanche, il constitue à l'évidence un excellent indicateur de productivité lexico-sémantique, en cela que les auteurs - et les locuteurs d'une langue en général - peuvent employer nombre de lexies dans des acceptions et des constructions intelligibles que les grammaires traditionnelles et la lexicographie ne prévoient ou ne mentionnent pas. De fait, un corpus actualisé, comme l'est FRANTEXT, est un bon indice de l'évolution du lexique, d'autant plus que les dictionnaires et les grammaires évoluent et s'enrichissent à son contact (il suffit d'observer la réactualisation permanente des acceptions dans le Petit Robert pour s'en convaincre).

En définitive, toute notre étude nous fait tendre vers une même conclusion : les corpus, les grammaires, les dictionnaires et l'intuition sont tous des outils d'analyse pertinents et indissociables ; et il nous paraît peu judicieux de prétendre porter un regard exhaustif et péremptoire sur la langue sans en observer tous les paramètres, qu'ils relèvent de la norme ou de l'usage, l'un influençant l'autre et inversement dans l'avancée du temps. »

Haton, 2005

Ce point de vue délibérément candide, mais ne l'étant évidemment pas au fond, a été bien accueilli lors de cette présentation, ce qui a renforcé notre détermination à pratiquer une « linguistique mixte ». La discussion finale n'a pas apporté davantage d'éclairage sur la validité de la linguistique de corpus mais ce qui nous a frappé est l'absence de « consensus mou » qui se dégage du débat.

En définitive, nous nous plaçons volontiers dans une position intermédiaire à la dualité proposée par Sinclair : nous sommes pour une linguistique appuyée par le

¹⁶⁵ On retrouve semblable épigraphe, « la plus belle fille du monde ne peut donner que ce qu'elle a » dans l'article de Corbin, Corbin, Tutin & Aliquot (1996) qui commentent et critiquent les perspectives d'informatisation du TLF. Selon ces auteurs, « (...) le TLF informatisé ne pourra finalement répondre aux requêtes des linguistes qu'à la mesure de ce que ceux-ci auront été capables de lui donner en pâture. » Le vœu final de Corbin & alii est que « ce produit puisse évoluer de telle sorte que les versions ultérieures soient susceptibles d'intégrer des connaissances nouvelles qu'il aura peut-être contribué à faire acquérir. » La question se pose nécessairement de la même façon pour tous les corpus.

¹⁶⁶ « Transitivité, emplois prototypiques et créativité langagière : comment un corpus littéraire réorganise la typologie des verbes de parole en français ». Article paru dans les actes des journées JETOU 2005 sur la linguistique de corpus.

corpus, lequel permet de vérifier des hypothèses et des résultats intuitifs ou issus d'autres catégories d'occurrences (construites ou lexicographiques), et aussi pour un prolongement motivé par la remise en cause des hypothèses initiales à partir de l'observation des données attestées. C'est à ce point qu'il faut être particulièrement vigilant : la remise en cause des théories d'origine ne doit pas occulter le caractère spécifique du corpus observé. En effet, un corpus littéraire ne peut avoir d'autre prérogative que d'illustrer la créativité des auteurs, et ainsi de servir d'indicateur du potentiel de la langue, comme nous l'avons dit par ailleurs.

Cette prudence de circonstance ne résout toutefois pas le problème : ce n'est pas parce que le corpus n'est rien d'autre que ce qu'il est qu'il ne doit pas servir de référence. La question peut en fait être retournée : qu'est-ce qui fait qu'un ensemble de données construites ne constitue pas un corpus acceptable ?

Il est donc très clair que la question de l'opposition entre les données construites et attestées est désormais centrale en linguistique. Nous ne saurions adhérer littéralement au jugement de Grevisse et Goosse (1993), qui considèrent qu'on ne doit pas confier les rênes de la grammaire aux auteurs littéraires qui ne sont pas des locuteurs très sérieux ni très fiables de leur point de vue ; sans aller non plus vers un groupement « tout-corpus à tout prix » qui semble jauger avec méfiance les linguistes qui ont encore l'outrecuidance de se construire des exemples sans se préoccuper de l'usage, nous nous poserons la simple question : Peut-on considérer qu'un linguiste est moins bon locuteur de sa propre langue qu'un écrivain, qu'un journaliste ou que quelque autre auteur (selon le corpus utilisé) ? La réponse est non, bien entendu. Cependant, il peut être tentant pour un linguiste d'adapter ses exemples aux phénomènes qu'il souhaite faire émerger. C'est d'ailleurs à partir de ce phénomène d'adéquation (inconsciente ?) que la polémique naît. D'autre part, la multiplicité des auteurs semble garantir qu'il n'y a justement pas de jugement porté sur la langue qui ne provienne d'une seule source scripturale.

Pour cette dernière raison, nous avons décidé d'inclure le plus petit nombre possible d'occurrences construites dans l'étude. On trouvera tout de même quelques exemples de ces dernières pour illustrer des phénomènes très particuliers liés aux verbes que nous analysons, ce qui n'empêche pas la circonspection vis-à-vis d'eux puisque nous avons probablement fait comme tout le monde : nos exemples construits sont sans nul doute *ad hoc*.

4. Traverser dans le corpus

4.1 Introduction à la traversée

Comme nous l'avons montré pour les seules données lexicographiques, le verbe *traverser* en français est l'objet de nombreuses traductions en anglais. Dans cette partie, nous chercherons à déterminer dans quelles mesures les traductions de ce verbe correspondent aux dimensions temporelles et spatiales qu'il exprime ; plus

précisément encore, nous chercherons à illustrer par le corpus quelles sont les données co-textuelles qui contraignent sémantiquement le prédicat verbal, ainsi que les traductions associées à chaque occurrence.

Avant toute chose, l'étude est ouverte par l'écriture d'une typologie ontologique des différents emplois du verbe issue de données attestées dans les textes littéraires. Cette ontologie ne saurait être exhaustive dans la mesure où sa constitution illustre avant tout la grande variabilité des « parcours » concrets ou non que le sujet *traversant* peut effectuer *au sein* ou *au travers* de l'objet *traversé*.

Ceci effectué, nous articulerons cette étude autour de trois axes. Le premier concerne la représentation en deçà du verbe *traverser* des concepts de mouvement et de bornage. Le second ouvre l'étude vers une représentation et une modélisation de la structure spatio-temporelle de ce verbe en fonction de facteurs intrinsèques et extrinsèques. Le troisième concerne le rapport à l'anglais des différents sens de *traverser* en contexte grâce au relevé des critères déterminatifs de l'aspect que nous aurons effectué.

Le mouvement est le concept de base du verbe *traverser* dans la plupart de ses emplois non métaphoriques. Il permet d'effectuer une première distinction entre les emplois qui induisent une traversée dynamique et les états au sens aspectuel. Par ailleurs, les emplois non spatiaux plus ou moins métaphoriques peuvent être ramenés à des emplois statifs et être modélisés comme tels au sein d'un espace n-dimensionnel. En effet, à de rares exceptions près (que nous ne négligerons pas), le verbe *traverser* se retrouve au contact d'au moins un argument concret ou considérable comme tel, en position sujet ou complément d'objet, ce que nous préciserons ultérieurement à propos des emplois temporels et des métaphores de la pensée.

Corrélativement, il est impossible de faire l'étude des occurrences de *traverser* sans prendre en compte systématiquement la notion de bornage. Par définition, une traversée, ou une activité en général, ne peut être perçue comme accomplie que si les bornes de l'événement coïncident avec les bornes de l'action. Dans le cas précis du verbe *traverser*, les dimensions de l'objet et son aspect borné sont des données à prendre nécessairement en compte pour ce genre d'analyse.

En inspectant la structure aspectuelle et sémantique du prédicat verbal centrée sur *traverser*, il apparaît que la première ne peut être complètement saturée qu'en fonction de données qui dépassent le simple cadre argumental du verbe. Outre que ce point de vue est admis, sinon défendu, par un nombre toujours croissant de linguistes, nous entendons montrer comment fonctionnent tous les facteurs qui participent de la construction du sens, indépendamment les uns des autres aussi bien qu'interdépendamment. Les éléments (intrinsèques) que nous avons isolés dans le chapitre précédent à partir de l'observation des dictionnaires se trouvent complétés par des informations extrinsèques qui dépassent parfois le cadre de la phrase.

À l'instar de ce qu'on peut observer dans les modèles développés par Denis & Muller (2004), toutes les occurrences du verbe semblent pouvoir être modélisées au sein d'un espace à quatre dimensions, sans distinction formelle entre le Temps et l'Espace. La saturation des traits aspectuels et sémantiques évoqués plus haut concourt à la construction des modèles à la hauteur de l'influence de chaque élément du discours impacté.

En dernier lieu, les variations sémantico-aspectuelles mises en relief par le contexte sont mises en parallèle avec l'important paradigme de traductions du verbe *traverser* en anglais. Nous ne reviendrons pas sur les raisons du choix de ce verbe particulier pour la présente étude. Nous pouvons toutefois présumer avec raison que les prédicats verbaux qui s'y apparentent peuvent subir la même analyse systématique avec les mêmes effets sur le transfert interlangue. Plus largement, les verbes français induisant du mouvement sont susceptibles d'être associés à un nombre de traductions anglaises particulièrement élevé et ne constituant pas un paradigme fermé grâce à des mécanismes de création lexicale que nous allons expliquer.

4.2 Diagrammes ontologiques des traversées à partir de la base littéraire

4.2.1 Présentation

Les diagrammes que nous proposons, ainsi qu'une grande partie des travaux de ce chapitre, sont inspirés par les recherches que nous avons effectuées en étroite collaboration avec Laurent MASCHERIN [Haton & Mascherin 2005]¹⁶⁷ ; presque rien de ce que nous présentons dans les pages qui suivent n'a été à l'origine le fruit d'un seul cerveau.

Les caractéristiques essentielles des schémas à venir sont que le sujet qui *traverse* est représenté par une flèche qui précise le sens de la traversée ainsi que les bornes de début et de fin de l'action. Dans certains cas, cette flèche n'a pas de tête s'il n'y a pas de mouvement (dans le cas des transpercées statiques en particulier). D'autre part l'objet traversé est dessiné sous la forme d'une surface grisée lorsqu'il s'agit d'un parcours, d'une épaisseur grisée quand c'est une transpercée.

¹⁶⁷ Travaux qui ont fait l'objet d'une communication orale au congrès STLL05 qui s'est tenu à Rijeka (Croatie) en avril 2005 sur le thème « Space and Time in language and literature ». Article à paraître, Sébastien Haton & Laurent Mascherin : « A cross-languages study about Space and Time : the French verb *traverser* and its equivalents in English », Zadar Publications.

4.2.2 Les traversées accomplies

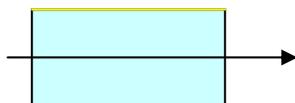
Les traversées accomplies listées ici partagent la caractéristique de l'accomplissement au sens aspectuel, i.e. le sujet traversant effectue un parcours qui part de l'extérieur de l'objet impacté vers l'autre côté de ce même objet.

4.2.2.1 Traversée en largeur, prototype de la traversée

La traversée en largeur avec valeur d'accomplissement est la plus étudiée dans les études linguistiques sur les verbes de mouvement. Dans l'exemple ci-dessous, rencontré maintes fois dans FRANTEXT, on suppose que la traversée s'opère d'un point qui est d'un côté (mais pas à l'intérieur) de l'objet traversé et s'achève de l'autre côté de cet objet. En l'absence de co-texte perturbateur de la traversée (interruption explicite), ce phénomène d'accomplissement est constant pour les traversées en largeur.

(1) = **Ils traversèrent la rue**
*TR*¹⁶⁸ : *They crossed the street*

Figure 12 : traversée dans la largeur



Nous pouvons noter que cette représentation correspond à l'objet linguistique le plus prototypique en matière de description. C'est également la seule qui admet l'emploi absolu de *traverser* dans un co-texte peu développé. « *Ils ont traversé* » est immédiatement interprété comme un emploi avec élision de l'objet de type *rue*, *route* ou *toute voie pouvant être arpentée dans sa largeur pour passer d'un de ses bords à l'autre*.

4.2.2.2 Traversée en longueur, emploi « déviant »

Rencontré une seule fois, cette interprétation n'est pas mentionnée dans les dictionnaires et s'oppose aux sèmes de base du verbe puisqu'il n'y a aucun moyen de repérer les bornes de la traversée autrement qu'en imaginant qu'une route a un début et une fin, et qu'elle est par conséquent un objet sécable dans sa longueur. Nous avons trouvé cette occurrence sur Internet en cherchant des exemples de « *traverser pendant* (une durée déterminée) ».

¹⁶⁸ TR = Traduction en anglais la plus probable

En l'occurrence, il s'agit probablement d'un glissement de type métonymique avec une ellipse involontaire : c'est la route qui *traverse* les pays et les voyageurs la *parcourent* ; conséquemment ces derniers *traversent* les pays par la route. Nous la présentons ici car elle s'oppose à tous les autres emplois du verbe autant qu'elle constitue, dans son statut d'anomalie linguistique, une opposition intéressante à la précédente représentation « dans la largeur ».

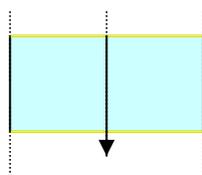
Si on respecte le sens habituel du verbe *traverser*, le parcours dénoté par cette phrase devrait être interprété comme une succession de zigzags d'un bord à l'autre de la route, avec avancée vers « Quito », dernière ville jointe par la « route ancestrale baptisée le chemin de l'inca » selon le co-texte :

(2) Ils **ont traversé**, pendant neuf mois, **à vélo et à pied**, la route ancestrale **baptisée «le chemin de l'inca»**, qui relie Santiago du Chili à Quito en Équateur.

(http://www.figarotudiant.com/associations_focus/20041127.ETU0009.html)

TR : *They biked and walked across the road "the lane of the INCA"*

Figure 13 : traversée dans la longueur



4.2.2.3 Traversée « angulaire »

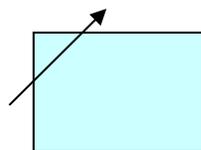
La traversée se résume dans cet exemple à un passage furtif à l'intérieur d'un espace ramené à un angle. Cet exemple illustre la capacité de *traverser* à désigner un parcours de faible ampleur au sein d'un espace visé comme *traversable*. Il montre également que la *traversée* est une activité dénotant bien d'autres choses que ce que suggère son acception prototypique.

Comme la traversée dans la largeur, la traversée dite angulaire a une valeur d'accomplissement puisque l'espace visé est entièrement parcouru d'un bord à l'autre. La différence réside dans la partie de l'objet qui est traversée, laquelle partie est d'ailleurs précisée dans le co-texte suivant au moyen de la périphrase « une partie de ». L'accomplissement de la traversée est attesté par l'intermédiaire de la proposition coordonnée « puis montèrent au premier étage de la caserne ».

(3) Ils **traversèrent** une **partie** de la cour, **puis montèrent au premier étage de la caserne**. (CLAVEL Bernard /Le Cœur des vivants/1964)

TR : They crossed or walked across one part of the yard and then walked/went upstairs to the first floor (from the outside to the outside by a corner)

Figure 14 : traversée angulaire



De l'extérieur vers l'extérieur par un coin

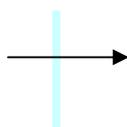
4.2.2.4 Traversée de part en part avec mouvement, transpercée 1

La première des transpercées est dynamique. L'occurrence du verbe décrit le mouvement du sujet qui *traverse* l'objet de part en part.

(4) = Le général Espagne reçoit une balle qui traverse la cuirasse.

TR : A bullet went/shot through General Espagne's breast-plate
(transpiercing)

Figure 15 : transpercée dynamique



Transpercée dynamique

4.2.3 Traversées partielles

Par le jeu de la malléabilité du langage, un nombre important des *traversées* langagières ne sont pas accomplies et renvoient à des parcours tronqués, c'est-à-dire qui ne saturent pas le sème prototypique « d'un côté à l'autre de l'objet ».

4.2.3.1 De l'extérieur vers l'intérieur

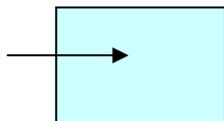
Dans l'occurrence suivante, « Véronique » n'était pas « dans le terrain vague » au début de l'action. La phrase « Elle ouvrit la portière...et se glissa à l'intérieur » signifie que la traversée s'interrompt à l'intérieur du terrain vague où sont garées les voitures. Comme pour l'occurrence de transpercée dynamique précédente, c'est le contexte qui nous permet de déterminer le type de parcours effectué, précisé dans une moindre mesure par l'emploi de l'imparfait de l'indicatif.

(5) Véronique traversait le terrain vague où étaient garées les voitures. Elle ouvrit la portière de notre vieille Porsche achetée d'occasion

l'an passé, et se glissa à l'intérieur. (AVENTIN Christine /Le Cœur en poche/1988)

TR : *Véronique walked across/through the waste ground where the cars were parked (to get into a car parked inside the waste ground)*

Figure 16 : traversée de l'extérieur vers l'intérieur



4.2.3.2 De l'intérieur vers l'extérieur

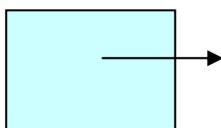
Dans l'exemple suivant, la maison est située à l'intérieur du village, ce dernier étant l'objet traversé. Il est probable une fois encore que la capacité du langage à subir des ellipses en permanence permet d'occulter le fait que le sujet ne « traverse *que la partie du village située entre la maison et la campagne* ».

Cette conviction n'altère pas néanmoins la réalité de l'occurrence observée. Il faut noter une fois de plus l'importance du co-texte, notamment ici le co-texte gauche qui précède la phrase et dans lequel on apprend que le personnage est dans sa maison située à l'intérieur du village qu'il va *traverser*.

(6) **Il sort brusquement (de la maison [left co-text]¹⁶⁹), traverse le village, et va déambuler dans la campagne où les paysans rient de lui.**(CABANIS José /Les Jeux de la nuit/1964)

TR : *He is going out home, then walks over/accross/through the village to wander into the open country (from the inside to the outside of the village)*

Figure 17 : traversée de l'intérieur vers l'extérieur



¹⁶⁹ Même si cela n'apparaît pas dans cet extrait, il est bien précisé dans les phrases qui précèdent celle-ci que la maison est à l'intérieur du village.

4.2.3.3 De l'intérieur vers l'intérieur, parcours « rectiligne »

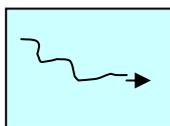
Le parcours représenté ci-dessous n'est pas précisément rectiligne mais plutôt motivé par un but. Le sujet traverse pour atteindre un emplacement prédéterminé. Nous admettons volontiers que le même énoncé puisse se rapporter à une situation où le sujet ignore encore où va s'arrêter son parcours au moment où la traversée commence. Nous pouvons noter que la traversée ne commence que lorsque le sujet est à l'intérieur de la pièce, et par conséquent qu'elle suit l'entrée.

Dans le co-texte, le circonstant « en zigzaguant » permet de désigner un type de déplacement particulier. Dans la traduction vers l'anglais, il autorise la sélection du verbe à particule *to zigzag through/across*.

(7) Il **traverse de biais** la salle où l'on dînait, il n'atteignit qu'**en zigzaguant le grand fauteuil de paille** qui faisait face à celui de la Comtesse, **au bout de la table**. (CHANDERNAGOR Françoise /L'Enfant des Lumières/ 1995)

TR : *He zigzagged through/across or cut across the dining room (to reach a place inside the room)*

Figure 18 : traversée de l'intérieur vers l'intérieur avec objectif



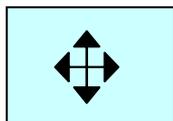
4.2.3.4 De l'intérieur vers l'intérieur, parcours « anarchique »

Cette occurrence du verbe *traverser* peut être paraphrasée par « traversait en tous sens » pour en caractériser la juste acception.

(8) Je lui en ai donné une et il est resté tranquillement appuyé sur ma portière pour la fumer et **me parler de son fils de huit mois qui traversait le salon à quatre pattes** (DJIAN Philippe /37°2 le matin/1985)

TR : *He talked me about his son who is/was going about and around into the living room on all fours (in all directions)*

Figure 19 : traversée en intérieur sans objectif



Intérieur-Intérieur sans objectif, dans toutes les directions

4.2.4 Traversées statiques

4.2.4.1 Transpercée statique

Contrairement à la transcercée dynamique décrite plus haut, la percée statique induit une absence de mouvement du sujet traversant. Il en résulte que la traversée est dans ce cas un état.

(9) Les vieux pilotes de Chine et d'Arabie regardaient ainsi s'orienter de lui-même, au niveau du bol d'eau, **l'oiseau peint et flottant sur son index de liège traversé d'une aiguille aimantée.** ... Rien là d'inerte ni de passif. Dans cette fixité du vol qui n'est que laconisme, l'activité demeure combustion. (SAINT-JOHN PERSE/Oiseaux/1962)

TR : (...) a magnetic needle going right through its cork pointer (de part en part) (...).

Figure 20 : transcercée statique



4.2.4.2 Traversée statique sans percée

Le prototype de cette traversée est « le pont qui traverse la rivière ». Ici, il n'y a pas non plus de parcours, le point A « de départ » et le point B « d'arrivée » sont hiérarchiquement équivalents en l'absence de mouvement et de direction se rapportant au sujet.

Dans ce cas, la traversée a comme caractéristique première de s'effectuer d'un seul tenant sans direction particulière. Dans l'exemple suivant, il est possible d'interpréter « la route » comme ayant une direction, dans la mesure où le co-texte mentionne « à la sortie de la ville », ce qui suppose un mouvement orienté. Cependant, cette orientation motivée par le début de phrase ne rend pas compte du caractère inanimé ni surtout de sa réalité intrinsèque : la route traverse la rivière dans les deux sens. Cet exemple est peut-être interprétable par ellipse : ce serait les voyageurs qui traversent en réalité. Dans cette optique, le fait que le pont traverse dans les deux sens est sémantiquement dominé et neutralisé par l'orientation du déplacement des voyageurs.

Pour cet exemple, nous avons choisi la représentation littérale de la situation en ignorant le parcours physique des personnes qui se déplacent sur la route qui traverse la rivière.

(10) **À la sortie de la ville, la route traverse une petite rivière sur un pont.** (Cavanna François /Les Russkoffs/1979)

TR : At the end of the town, the road crosses a small river on a bridge

Figure 21 : traversée statique en largeur



4.2.4.3 Une traversée « semi-statique » ou « semi-dynamique » ?

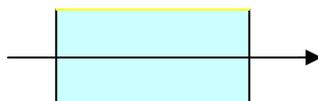
Le cas de la rivière qui traverse un territoire implique du mouvement davantage que l'exemple précédent où une route ne génère du mouvement que par le truchement de ceux qui l'empruntent.

Un cours d'eau coule toujours dans un certain sens, de sa source vers son embouchure, et possède en outre des bornes généralement identifiables dans sa longueur comme dans sa largeur. Malgré cela, les cours d'eau suivent un mouvement immuable et constant¹⁷⁰ et leur traversée n'est pas forcément liée à la direction du courant d'un point de vue conceptuel. « Une rivière qui traverse un pays » peut très bien être considérée comme une étendue d'eau longiligne dont on ignore le déplacement du flot pour ne considérer que sa réalité géographique.

(11) (...) la porte entrouverte de l'un des bungalows. **Le chantier : une rivière le traverse, un pont inachevé** ; au loin, un lac. Les appels de la garde : bruits de langue, de gorge, choc de fer sur du fer, de fer sur du bois, petit cris, hoquets (...) (KOLTÈS Bernard-Marie /Combat de nègre et de chiens suivi des Carnets/1983)

TR : The building site: crossed by a river, an unfinished bridge; (...)

Figure 22 : traversée semi-dynamique



¹⁷⁰ Si l'on excepte les cas de détournement, de tarissement ou de barrage par l'humain ou par des circonstances naturelles (sécheresse et crues par exemple).

4.2.5 Traversées ininterprétables dans l'espace de l'objet

4.2.5.1 Traversées temporelles

A moins de considérer le temps comme une surface, nous ne voulons pas pour l'instant intégrer les traversées temporelles dans notre série de diagrammes ontologiques, lesquels ne concernent donc que les traversées spatiales, dynamiques et statiques.

Toutefois, nous proposons en section 3.4 une représentation des emplois temporels dans des modèles équivalents aux emplois spatiaux. En l'occurrence, il s'agit d'une projection de tous les emplois du verbe dans un espace à quatre dimensions, une seule dimension étant nécessaire pour représenter la plupart des emplois temporels. Nous verrons néanmoins que certains de ces derniers peuvent être conceptuellement associés à des traversées spatiales, lorsque l'objet visé n'est plus explicitement un intervalle temporel mais un événement.

4.2.5.2 Métaphores de la pensée

Contrairement aux emplois temporels, qui impliquent des sujets concrets dans une période de temps donnée, les métaphores de la pensée mettent en jeu des facteurs abstraits. Les sujets sont des idées, des pensées ou toute autre manifestation de l'esprit qu'il n'est pas possible de matérialiser (sauf sous forme de dessin ou de graphie pour les images mentales ou les idées verbalisées) ; l'objet est systématiquement le cerveau ou un de ses avatars langagiers (esprit, pensée, etc.).

De ce fait, aucun de nos diagrammes n'est susceptible de représenter fidèlement la « réalité » de ce type d'occurrences.

4.3 Mouvement et bornage dans les traversées

4.3.1 Pour commencer avec l'aspect

Commençons par établir les distinctions de base : en physique fondamentale, le temps est unidimensionnel, tandis que l'espace est tridimensionnel. Cette nuance essentielle va motiver certaines traductions de traverser en anglais, d'autant que les occurrences de *traverser* sont soumises à la double influence du Temps et de L'Espace dans la saturation de ses traits aspectuels

Au-delà de cette opposition, les dimensions temporelles et spatiales impliquent deux types de variables complémentaires qui sont le mouvement et le bornage.

Comme le soulignent Muller & Sarda (1998:133), le verbe *traverser* peut être paraphrasé par **passer d'un côté à l'autre côté de x**, et dans ce cadre il implique donc une dimension directionnelle et une dimension de localisation. Par ailleurs, cette paraphrase en l'état présuppose un accomplissement de la traversée en tant que

franchissement d'une ligne virtuelle (ou d'un espace) qui sépare un point A d'un point B, tous deux situés de part et d'autre de l'objet traversé.

Le sens aspectuel du verbe traverser n'a jamais été classifié dans le cadre d'une typologie des procès. Lorsqu'il apparaît comme exemple prototypique d'une classe aspectuelle, il y entre toujours sous la forme prédicative *traverser la rue* ou en emploi absolu avec la même acception stéréotypique (Vendler 1967¹⁷¹ ; Récanati & Récanati 1999) et dans ce cadre, il est considéré comme un verbe d'accomplissement. Au niveau lexical, Victorri (1997), Sarda et Muller (1998:144, figure 6) le considèrent également comme un verbe télique et transitionnel, donc d'accomplissement, en considérant que le sujet qui traverse dépasse l'objet traversé. Ces visions sont évidemment conditionnées par une étude globale des verbes de mouvement ou par une volonté de bâtir une typologie générale fondée sur l'usage habituel ou les traits génériques plutôt que sur le « sable » de l'analyse systématique de chaque verbe selon ses emplois potentiels. En contexte, chaque verbe déploie son éventail d'emplois possibles, voire impossibles du point de vue normatif (cf. L'impossible dans la langue, Rousseau 2000). De ce point de vue, nous ne pouvons pas ignorer les emplois « anormaux » ou « déviants », volontairement ignorés ou sous-évoqués dans les études globales. Le point de vue le plus ouvertement hostile à la classification des emplois inhabituels est celui de Vendler dans son essai de 1967 sur la typologie des classes de procès. Il y présente entre autres une occurrence du verbe *to cross* qu'il appelle *oddtity*, c'est-à-dire « bizarrerie » ou « anomalie » en français :

'For those who like oddities: *it took the battalion twenty minutes to cross the border; They are crossing the border.* Such are the borderline cases I mean to ignore at this stage'
(Vendler, 1967)

La consultation des données lexicographiques nous permet donc de dégager quatre classes sémantiques pour ce verbe, lesquelles renvoient à de multiples valeurs aspectuelles. Ces différentes dimensions aspectuelles et spatiales sont liées intrinsèquement à la structure prédicative (voir ce qui se passe dans le TLF par exemple) :

- parcours spatial (dynamique, statif ou en emploi absolu) ;
- transpercée (dynamique ou statif) ;
- parcours temporel (statif au sens spatial et dynamique au sens temporel) ;

¹⁷¹ Vendler ne présente pas le prédicat verbal *traverser la rue* mais son équivalent en anglais *to cross the street*. À l'issue de la démonstration, ses conclusions sont équivalentes.

- métaphore de la pensée (emploi figuré associé au sens de *transpercée*).

Toutefois, les arguments du verbe ne suffisent pas à déterminer de façon systématique comment ces traits aspectuels sont effectivement saturés. Dans les prochains paragraphes, nous montrerons comment le prédicat au sens large et le contexte jouent un rôle dans la reconnaissance des traits aspectuels qui ne dépendent pas du contexte non langagier. En premier lieu, nous distinguerons les arguments en mouvement, ou soumis à la traversée, des objets intrinsèquement bornés qui la subissent.

4.3.2 *Ce qui bouge*

4.3.2.1 L'objet en mouvement

Nous admettons que *l'objet en mouvement* est non pertinent dans l'analyse. Par là il faut entendre qu'il n'existe pas à notre connaissance d'occurrence de traverser pour lequel l'objet-site est en mouvement avec un effet pertinent sur la saturation des traits aspectuels de l'énoncé dans son ensemble. Cela n'exclut pas la possibilité d'un énoncé tel que le suivant que nous avons construit :

(10) *Nous traversons avec peine le pont du bateau* (NDLR : /bateau/
qui traversait lui-même l'océan)

Le circonstant *avec peine* est un indicateur de la difficulté qu'il y a à se déplacer sur un objet en mouvement tel que le *pont d'un bateau*. Faute de données corpusculaires en nombre suffisant, nous n'intégrerons ce type d'occurrence qu'avec prudence et parcimonie.

4.3.2.2 Le sujet en mouvement

De fait, puisque l'objet en mouvement n'est pas observable, la pertinence du mouvement est reportée sur le sujet et éventuellement sur l'observateur puisque c'est ce dernier qui constate le mouvement et le transmet en tant que locuteur. Nous verrons que, contrairement à Reichenbach (1980), nous n'intégrons pas l'observateur dans la modélisation du prédicat, avant tout parce que l'observation subjective est l'élément le plus difficile à isoler et à caractériser dans un texte écrit par quelqu'un d'autre que soi (cf. la question de la polysémie absolue¹⁷²).

¹⁷² Notion que nous présentons au Chapitre I.

Afin de ne pas mélanger toutes les acceptions du verbe au sein d'un même modèle de fonctionnement, il est nécessaire de bien faire la différence entre les emplois de *traverser* :

- Pour les parcours spatiaux avec déplacement dans un espace ou sur une surface, le mouvement du sujet est manifeste. Il n'y a pas de parcours spatial sans mouvement du sujet sémantique, formes passives comprises.
- Les transpercées se partagent en deux classes, l'une inclut les occurrences où la percée s'opère avec un sujet en mouvement qui « passe d'un côté à l'autre de la cible », l'autre contient les emplois statifs pour lesquels la traversée est un fait, une observation d'un état dans une interaction immobile mais pas immuable. Une transition entre les deux est possible si l'objet qui transperce est en mouvement avant le contact avec la cible puis reste fiché dedans.
- Les emplois temporels seront discutés ultérieurement. Dans un premier temps, nous considérons qu'il ne s'agit plus d'un sens métaphorique dérivé du spatial, mais bien d'un sens propre lexicalisé dans la langue (cf. Todorov, 1967). Or, la structure argumentale de *traverser* suppose l'existence d'un sujet qui effectue réellement une traversée (même statiquement) sur un objet qui n'est pas censé bouger.
- Les emplois que nous considérons comme réellement métaphoriques ne comportent les notions de mouvement et de bornage que d'un point de vue conceptuel, et il ne s'agit d'ailleurs là que de notre propre vision de ces acceptions. Les métaphores *stricto sensu* se heurtent à l'analyse n-dimensionnelle que nous proposons car il n'est pas possible de repérer le mouvement réel du sujet ni les bornes de l'objet en tant qu'espace traversé.

4.3.2.3 Prise en compte globale du mouvement dans l'analyse

Bien qu'étant un des traits sémantiques les plus saillants dans la structure prédicative prototypique de *traverser*, le mouvement est manifestement un trait qui ne concerne pas l'ensemble des acceptions du verbe. Parmi les nombreux traits que nous essayons de mettre en exergue, celui-ci ne concerne réellement qu'une frange minoritaire des acceptions nomenclaturées précédemment.

Toutefois, la mise en perspective homogène de l'action de *traverser* en contexte par des modèles graphiques n'est pas à remettre en cause, puisque le mouvement n'est qu'un des éléments prévus par nos modèles sans être la condition de leur constitution.

4.3.3 Ce qui est borné

4.3.3.1 Les bornes du sujet sémantique

Comme tout objet identifiable, le sujet est borné en soi mais son bornage n'est pas initialement pertinent dans l'analyse aspectuelle du prédicat. En effet, dans les emplois les plus fréquents du verbe *traverser*, le sujet est ramené à la traversée elle-même. C'est donc le parcours et non la chose traversant qui est pris en compte pour la reconnaissance de la structure aspectuelle finale du prédicat.

Certains emplois du verbe nécessitent cependant la juste reconnaissance des bornes du sujet : c'est le cas notamment des transpercées « de part en part » qui supposent que les extrémités du sujet traversant sont de part et d'autre de l'objet traversé. Ceci pose par ailleurs un problème de modélisation, comme nous allons le voir dans le sous-chapitre correspondant.

4.3.3.2 Les bornes de l'objet sémantique

A l'inverse du mouvement éventuel, l'objet porte les marques de bornage saillantes dans le procès. En effet, en reprenant la paraphrase de Muller & Sarda, nous admettons dans un premier temps que *traverser* est un verbe dit télélique quand il y a *passage d'un côté à l'autre de X*. Il est notable qu'aucune mention de l'objet traversant n'est faite dans cette paraphrase, alors que l'objet traversé est intrinsèquement borné dans la prédication puisque *le passage d'un côté à l'autre* implique la présence d'au moins deux côtés (cf. ontologie et modèles), et subséquemment de bornes spatiales.

De même, il n'est pas difficile d'étendre, disons plutôt d'aménager cette paraphrase pour d'autres emplois moins accomplis : *passage d'un point A à un point B avec focalisation prédicative sur l'objet Y, sans présomption d'inclusion ou d'exclusion des points A et B à l'espace circonscrit par Y*. De ce point de vue, Y présente des bornes identifiables et autorise la mise en perspective du type de traversée effectué en fonction de cette fameuse inclusion ou non de A et B. C'est à partir de ce phénomène d'inclusion / exclusion que les marques aspectuelles de la traversée sont identifiables.

Il en ressort que le bornage saillant est une valeur propre à l'objet dans l'étude de la structure aspectuelle du prédicat.

4.3.3.3 Les limites de l'action et de l'événement

Les bornes de l'action ne sont pas forcément dépendantes du mouvement et elles ne sont pas directement corrélées aux bornes de l'objet traversé. Dans notre nomenclature, il s'agit des bornes spatiales de l'événement à l'intérieur et au-delà de l'espace traversé. En d'autres termes, quel que soit le type de traversée dont il s'agit et

quelle que soit sa forme, les limites de l'action correspondent aux bornes de début et de fin de la traversée, i.e. à l'intersection du bornage spatial et du bornage temporel.

Nous rappelons que l'objet est toujours délimité en ce qui concerne les emplois spatiaux de *traverser*. Par ailleurs au niveau aspectuel, l'événement, c'est-à-dire le parcours effectué, est toujours contraint. En d'autres termes, une traversée possède des bornes temporelles qui correspondent à la délimitation de l'intervalle durant laquelle l'action se déroule.

Les limites de l'événement conditionnent les limites de l'action précédemment décrites : l'action commence à l'instant t_0 qui est la borne initiale de l'événement et s'achève à l'instant t_1 qui est la borne finale de l'événement. Ceci implique une lecture de la nature de l'action dénotée par telle ou telle traversée en fonction de sa durée.

Pour bien distinguer l'action de l'événement, nous pouvons dire que la première répond à la question « où le parcours commence et où s'arrête-t-il ? » en montrant la relation entre la réalité spatiale du parcours (incluant la spatialité de l'objet qui traverse) et celle de l'espace parcouru ; quant au second, il répond à l'interrogation suivante : « Quand la traversée commence et s'achève-t-elle ? », ce qui fait qu'il a trait à la structure temporelle de l'action. Les bornes de l'action et de l'événement sont intriquées dans la réalité du monde et du langage et nous le mettrons en avant dans nos modèles.

4.3.3.4 Résumé du bornage

Les dimensions aspectuelles du changement et de l'aboutissement (télicité) ne correspondent pas toujours à la dimension spatiale de l'objet traversé (nous reviendrons sur ces questions dans la partie suivante). Comme nous pouvons le remarquer à partir des représentations ontologiques de *traverser*, il est possible de parler de traversée dès lors qu'il y a parcours au sein d'un objet ayant la particularité de posséder des bornes externes identifiables.

Nous constatons rapidement que la paraphrase de Sarda est partiellement vraie et non généralisable, puisque parmi tous nos diagrammes ontologiques seuls deux représentent explicitement le **passage d'un côté à l'autre côté de x**. Ainsi, si toute traversée dynamique implique en principe le passage d'un point A vers un point B, ce mouvement n'est pas nécessairement structuré par l'espace traversé et ses bornes¹⁷³.

¹⁷³ Cela revient à dire qu'une traversée dynamique est toujours un déplacement d'un point vers un autre, mais que la connaissance des bornes de l'espace traversé n'est pas nécessairement suffisant pour la caractériser.

4.3.4 La dimension aspectuelle des situations

4.3.4.1 Quelques définitions

La dimension aspectuelle des situations peut être définie comme l'action intrinsèque au prédicat, c'est-à-dire en dehors de la représentation du temps en tant qu'appartenant aux ensembles *passé-présent-futur* et *antérieur-simultané-postérieur*. Au niveau sémantique, plusieurs caractéristiques permettent de définir quel est le type d'action exprimé.

Cette dimension aspectuelle est partiellement dépendante de l'interprétation du co-texte extra-prédicatif, en particulier pour les occurrences d'un verbe comme *traverser*, pour lequel la recherche des bornes de l'action et de l'événement sont tributaires des actions et événements antérieurs et postérieurs, à savoir les actions qui précèdent et qui permettent de déterminer le point de départ, et les actions qui suivent, lesquelles donnent l'éclairage sur le point d'arrivée.

Partant de ce postulat, il nous faut effectuer deux distinctions :

- une distinction entre les propriétés aspectuelles propres au lexème verbal et les propriétés aspectuelles prédicatives et extra-prédicatives ;
- et une autre entre les classes aspectuelles et les traits aspectuels.

4.3.4.2 Sens aspectuel du verbe : propriétés intrinsèques et potentiel

Tout verbe, en tant qu'élément central de l'expression d'un procès, possède un sens aspectuel, soit de façon prototypique, soit en terme de potentialité. Le contexte qu'il soit prédicatif ou phrastique peut entraîner une modification du sens aspectuel primaire ou simple du verbe. Cette modification de la valeur aspectuelle du verbe en fonction du contexte peut-être analysée en terme de combinaisons de valeurs aspectuelles ou selon une analyse prédicative ou phrastique.

Ce qui nous intéresse tout particulièrement dans notre analyse globale du verbe *traverser* est le décalage perceptible entre l'aspect « prévisible » porté par le verbe et sa réalisation en contexte. En tout état de cause, il apparaît que les verbes sont lexicalement marqués comme porteurs de traits spécifiques¹⁷⁴ mais que ces traits sont soit confirmés, soit modifiés soit complètement neutralisés par le contexte énonciatif. D'où il ressort qu'une théorie préliminaire visant à dresser une typologie lexicale hors contexte peut être facilement remise en question dans le cadre d'une analyse des prédicats en contexte.

Ceci conforte notre hypothèse que les traits aspectuels ne sont pas seulement inhérents au lexème verbal mais dépendent essentiellement du cadre énonciatif au sens

¹⁷⁴ Ce que confirment les classifications aspectuelles des auteurs que nous avons consultés.

large. De par sa forte polysémie, *traverser* pose initialement un problème de typologie par le prototype telle que nous le retrouvons dans les études existantes. Cette vision des choses n'exclut pas, évidemment, que le lexème verbal conserve son propre potentiel aspectuel activable en dehors de tout contexte langagier.

4.3.4.3 Typologie aspectuelle, classes et traits distinctifs

Comme nous l'annoncions au paragraphe précédent, la structure temporelle du verbe peut-être représentée sous la forme d'intervalles et de bornage. Afin de limiter dans un premier temps la description de la dimension aspectuelle de *traverser*, et par extension de tout autre verbe, nous choisissons de nous appuyer sur les quatre classes aspectuelles de Vendler, s'établissant à partir des propriétés sémantiques des verbes et prédicats, et matérialisés sous la forme de traits sémantiques :

- verbes d'accomplissement ;
- verbes d'achèvement ;
- verbes d'état ;
- verbes d'activité.

Dans notre étude, nous distinguons systématiquement les traits sémantiques inhérents au sens aspectuel du lexème verbal des traits sémantiques extrinsèques.

Parmi tous les traits sémantiques et temporels, quatre suffisent pour distinguer le sens aspectuel des verbes au sein de cette typologie à quatre classes. Ces traits sont soit activés (+) soit non (-) :

- changement : le procès implique ou non un changement entre sa phase antérieure et sa phase ultérieure ;
- momentané : le procès est vu sans durée ;
- téléique : le procès a besoin d'atteindre un point culminant pour être dit réalisé ;
- dynamique : le procès est conçu comme se déroulant dans le temps.

Hors contexte d'énonciation, ces traits nous permettent d'organiser les types de procès comme suit :

Figure 23 : typologie aspectuelle

	Changement	Momentané	Télique	Dynamique
Achèvement (partir ¹⁷⁵ , trouver,...)	+	+	Pas d'intervalle temporel donc pas de télicité	+
Accomplissement (traverser, ...)	+	-	+	+
Activité (courir, danser,...)	-	-	-	+
État (désirer, aimer,...)	-	-	-	-

Les verbes associés à chaque classe sont choisis en fonction de leur emploi le plus courant. L'un de nos objectifs majeurs demeure de montrer qu'il n'est pas pertinent pour notre étude d'opérer une telle classification préalable verbe par verbe car le co-texte et la situation d'énonciation modifient profondément les valeurs aspectuelles « attendues », motivant une réorganisation des classes aspectuelles emploi par emploi. De plus, les traductions d'un verbe comme *traverser* sont tributaires de la saturation des *aspectèmes*¹⁷⁶.

La limitation des traits aux quatre précités nous semble obligatoire pour permettre une distinction claire des éléments exprimant l'aspect et de leur combinatoire.

4.3.4.4 Remise en cause de la typologie des verbes, bornage et télicité

Vet (1980)¹⁷⁷ et François (1989)¹⁷⁸ font intervenir un trait prédicatif qui est l'agentivité et qui concerne l'intervention d'un agent (agent animé) au niveau du procès. Nous pouvons nous demander si la distinction entre procès agentif et non-agentif permet de construire un nouveau type de procès ou bien si elle ne fait que rendre compte d'un changement de type de procès. Car, pour reprendre un argument de Karolak (2003), en soi, l'agent n'exprime pas du temps, c'est l'agent en mouvement qui exprime du temps.

¹⁷⁵ Les verbes présentés ne sont en aucun cas circonscrits à une seule classe aspectuelle. Il s'agit seulement de leurs propriétés prototypiques.

¹⁷⁶ L'aspectème est la plus petite unité porteuse d'aspect.

¹⁷⁷ **Vet C.**, 1980. *Temps, aspects et adverbess de temps en français contemporain*, Genève : Droz.

¹⁷⁸ **François J.**, 1988. *Changement, causation, action. Trois catégories sémantiques fondamentales du lexique verbal français-allemand*. Librairie Droz.

A nos yeux, la problématique la plus intéressante pour analyser ce verbe consiste dans l'opposition entre le bornage et la télélicité. Pour ce faire, nous nous appuyons sur l'analyse de Kozłowska (1998) qui considère que la télélicité est inhérente au lexème verbal et que le bornage est extrinsèque, c'est-à-dire dépendant des valeurs contextuelles. Pour Kozłowska, la notion de borne est à distinguer de la notion de télélicité :

'La télélicité est alors une propriété inhérente (intrinsèque) d'une éventualité : une éventualité est par défaut télélique ou non télélique. Par contre, le bornage n'est pas une propriété inhérente (c'est une propriété extrinsèque).'
Kozłowska (1998 : 220)

En reprenant cette distinction pour notre compte, nous montrerons comment le co-texte aide à fixer les bornes de l'événement, mais aussi comment la reconnaissance de ces bornes mène à la détermination de la présence ou non de télélicité liée au procès en contexte. C'est ce que nous évoquons quelques pages plus tôt à propos de l'événement : « la recherche des bornes de l'action et de l'événement sont tributaires des actions et événements antérieurs et postérieurs ».

La difficulté est que ces deux types de bornages peuvent être rassemblés sous la même définition posée par Martin : Un procès est borné s'il porte en lui-même une limite en deçà de laquelle il ne peut être dit réalisé et au-delà de laquelle il ne peut être que recommencé et non pas prolongé. (Martin 1988 : 5)

Nous émettons l'hypothèse que pour être télélique le verbe *traverser* doit remplir deux conditions :

- Tous ses emplois doivent impliquer une borne finale, ce qui n'est pas forcément le cas en contexte.
- Si un emploi ne possède pas ce trait il faut démontrer que le contexte empêche la réalisation de cette borne, qui reste visée.

A contrario, si nous pensons que le verbe est non-télélique, il faut montrer que dans toutes ses occurrences, c'est le contexte discursif qui va poser la borne finale. Autrement dit, le verbe *traverser* n'est peut-être pas intrinsèquement télélique, ce qui induit que c'est le cadre discursif tout entier qui contient et détermine des fragments téléliques et d'autres qui ne le sont pas.

4.3.4.5 Ce qui détermine le sémantisme du verbe

Une des questions principales est de savoir si *traverser* est déterminé par l'objet¹⁷⁹ qui est traversé, par le parcours effectué par l'objet traversé ou bien par le contexte situationnel encadrant le procès de la traversée. L'observation des emplois du verbe nous montre que le sens de *traverser* en contexte est fonction de tous les éléments précités, mais que la saturation de ses traits aspectuels dépend du co-texte, voire de l'absence de co-texte. En effet, les exemples construits dont nous présenterons quelques spécimens dans la partie « modélisation » donnent facilement lieu à plusieurs interprétations, chacune d'entre elles suggérant un ensemble de traits aspectuels distincts des autres.

Il est notable que l'absence de co-texte induit des propriétés particulières pour le verbe *traverser* comme pour n'importe quel autre verbe. En l'occurrence, un énoncé sans contexte langagier supplémentaire tel que « nous avons traversé la place » ne peut s'interpréter que comme un acte accompli sous peine d'être ininterprétable, faute d'informations. C'est certainement pour cette raison que *traverser* entre dans les typologies en tant que verbe télique.

4.3.5 Valeurs aspectuelles de traverser et bornage

4.3.5.1 Présentation des tests de recherche des bornes

Pour la recherche des bornes, nous aurons d'abord recours à deux tests standard :

- le test de compatibilité avec *en/pendant* suivi d'un syntagme exprimant une durée, où *en* valide les prédicats téliques et *pendant* valide les prédicats non-téliques. Si les deux tests sont possibles (voir Renaud 2001), il faut étendre le test au contexte élargi ;
- le test de l'implication¹⁸⁰ afin de déterminer si le procès est compris entre deux changements (Vet 1980 : 61). Dans l'exemple du verbe *traverser*, le changement correspond à un changement spatial, *traverser*

¹⁷⁹ Nous considérons à travers le mot objet les individus animés / non-animés et les objets animés / non-animés. Le terme est à prendre ici au sens syntaxique.

¹⁸⁰ Le test de l'implication selon Vet fonctionne comme la recherche d'un changement entre un état antérieur au prédicat et un état ultérieur. Autrement dit, il y a télicité si avant P(x) il n'y avait pas x et si après il y a x. Dans le cas du verbe *traverser* formalisé par Traverser(x) où x est l'espace traversé, le test d'implication est valide si avant l'emploi du verbe il n'y avait pas traversée de x et si après la traversée est accomplie. Or, on peut « traverser quelque chose » sans qu'il y ait nécessairement accomplissement de la traversée toute entière (simple activité ou traversée partielle).

quelque chose pour revenir au point de départ n'impliquant pas de changement.

La combinaison des deux tests nous permet de déterminer la classe aspectuelle de chaque emploi du verbe :

- Les verbes d'état se caractérisent par l'échec aux deux tests, ils n'ont pas de durée et n'impliquent pas de changement.
- Les verbes d'achèvement ne répondent pas au test temporel par *en/pendant* car ils induisent une absence de durée. Toutefois, le changement est nécessaire pour que le « climax » propre à l'achèvement soit atteint.
- Les verbes d'activité et d'accomplissement impliquent tous les deux une durée, une reconnaissance des bornes de l'événement. Les deux classes se distinguent par le test de changement. Une activité se caractérise par l'absence de changement entre ses bornes antérieures et postérieures, tandis que l'accomplissement est nécessairement associé à un changement.

4.3.5.2 Les emplois statifs : recherche des bornes

Reprenons l'extrait des Oiseaux de Saint-John Perse :

« Les vieux pilotes de Chine et d'Arabie regardaient ainsi s'orienter de lui-même, au niveau du bol d'eau, l'oiseau peint et flottant sur son index de liège traversé d'une aiguille aimantée. ... Rien là d'inerte ni de passif. Dans cette fixité du vol qui n'est que laconisme, l'activité demeure combustion. »
SAINT-JOHN PERSE /Oiseaux/1962

Test temporel :

- * L'aiguille aimantée traverse l'oiseau peint et flottant en une heure.
- ? L'aiguille aimantée traverse l'oiseau peint et flottant pendant une heure.

À partir de ce test, nous pouvons dire que les emplois statifs de *traverser* ont plusieurs caractéristiques :

- Les tests de l'implication et du changement ne fonctionnent pas, ni le test *en/pendant*.
- L'agent est non-animé, non agentif. Ainsi, le trait de Vet et François semble être pertinent ici, dans le sens où il permet de guider l'interprétation du sens aspectuel du verbe. Toutefois, il s'agit avant tout d'une conséquence de l'acception particulière du verbe, la transpercée statique ne s'effectue pas normalement, et par définition, avec un sujet agentif animé.

Parmi les emplois non-agentifs, il existe une autre catégorie plus ambiguë, il s'agit du temps qui fait avancer l'événement sans déplacement apparent de l'agent. Nous détaillerons cette catégorie dans la partie contrastive de ce chapitre.

4.3.5.3 Les emplois ponctuels : recherche des bornes

Les emplois dits ponctuels se retrouvent principalement dans la valeur métaphorique des emplois de *traverser*. En effet, l'espace traversé étant indélimité et indélimitable, seul le test du changement permet de définir ses propriétés sémantiques et il ne semble y avoir aucune phase préparatoire. De plus, les adverbes juxtaposés au prédicat sont des indices de l'immédiateté de la situation. Le changement est clair puisque ces verbes impliquent qu'avant la traversée « l'idée » n'était pas là est qu'après elle est là.

Il y a des moments, oui, où vous me faites peur... et votre type, là, l'homme à la fleur, ça ne **vous a jamais traversé le crâne qu' il avait peur de vous**, lui aussi?

Rivoyre Christine de /Les Sultans/1964

C'est alors que **son esprit fut brusquement traversé par une idée**, une image, qui ne l'avait pas encore effleurée, une phrase cent fois entendue - il y avait tant de gibets ! -, une plaisanterie affreuse sur le sexe des pendus (...)

Chandernagor Françoise /L'Enfant des Lumières/1995

Et c'est ce jour-là que, pour la première fois, **cette pensée exaltante me traversa l'esprit** : " Et si l'on pouvait exprimer cette langue par écrit ?"

Makine Andreï /Le Testament français/1995

Les emplois de *traverser* pouvant être glosés par *transpercer* peuvent également être des verbes d'achèvement et peuvent avoir le double sens achèvement / statif.

Le malheureux mannequin qui traversa une fenêtre du Montana pour s'écraser sur le pavé n'instaura jamais une école du désarroi.
Blondin Antoine /Ma vie entre les lignes/1982

4.3.5.4 Les emplois dynamiques : recherche des bornes

Les emplois dynamiques de *traverser* possèdent une caractéristique essentielle, le sujet qui traverse doit être un animé, objet ou individu. Plus exactement, l'agent doit être **en mouvement**, ce qui ne le contraint pas nécessairement à se mouvoir de son plein gré, et conséquemment n'implique pas qu'il soit intrinsèquement animé.

Observons deux exemples construits du verbe *traverser* et l'application des tests de Vet :

(1) *X traverse la route.*

(2) *X traverse un pays.*

(1a) *X traverse la route en quelques secondes* (implique que x est d'un côté de la rue et présuppose que x était de l'autre côté de la rue).

(1b) *X traverse la route pendant quelques secondes* (présuppose éventuellement que x était d'un côté de la rue et n'implique pas que x a nécessairement traversé entièrement).

(2a) *X traverse un pays en quelques jours* (implique que x est d'un côté du pays et présuppose que x était de l'autre côté du pays).

(2b) *X traverse un pays pendant quelques jours* (pas d'implication, pas de présupposition).

Hors contexte, le test *en / pendant* ne fonctionne pas clairement ; *traverser* peut être perçu comme un verbe d'accomplissement ou d'activité, seul le contexte élargi permet de rendre compte du bornage, même s'il s'agit ici d'un exemple déviant dont nous avons déjà discuté dans la présentation des diagrammes ontologiques (3) :

(3) **Ils ont traversé**, pendant neuf mois, **à vélo et à pied, la route** ancestrale **baptisée «le chemin de l'inca»**, qui relie Santiago du Chili à Quito en Équateur.

(http://www.figaroetudiant.com/associations_focus/20041127.ETU0009.html)

TR : *They biked and walked across the road "the lane of the INCA"*

Nous noterons que les deux sens de traverser avec *en* et *pendant* demandent deux orientations spatiales différentes, celles-ci correspondant à la distinction entre la

traversée « en largeur » - qui est le prototype de la traversée - et les autres traversées moins linéaires. En effet, (1a) se situe préférentiellement dans un espace unidimensionnel et (1b) dans un espace bidimensionnel. Il en est de même pour (2a) et (2b).

Les lectures n-dimensionnelles de l'espace ne sont pas fixées à l'avance, elles dépendent du contexte énonciatif comme nous l'avons vu avec les exemples ci-dessus. Mais l'objet traversé implique sans doute intuitivement un sens prototypique qui est (1a) pour l'exemple 1 et (2a) pour l'exemple 2. Ainsi dans (1a), la lecture (prototypique) est l'accomplissement ; nous pouvons considérer que la valeur intrinsèquement télique du prédicat est fonction du franchissement des bornes externes de l'objet. Quant à (2b) la lecture (toujours prototypique) est l'activité, *traverser* implique un parcours dans l'objet traversé.

En conclusion à ces observations générales, nous dégagons deux emplois dynamiques identifiables : celui pour lequel l'intervalle temporel dépend de l'intervalle spatial vu comme unidimensionnel, dans ce cas l'événement pour être dit réalisé doit absolument atteindre les bornes externes de la cible (lectures privilégiées : *traverser une route, rue, rivière*) ; ou bien, l'intervalle temporel fixe des limites au sein desquelles se situe la traversée (*maison, pays, village, ville, etc.*), le verbe *traverser* dans ce cas est limité par des bornes internes de la cible. La traversée n'a pas besoin d'atteindre un point culminant pour être effectuée.

L'agent qui traverse est un animé. Y a-t-il une distinction au niveau du contrôle de l'agent sur le procès, ce qui reviendrait à distinguer objet d'individu ? L'objet traversé doit être spatialement étendu et limité ; l'agent doit dépasser les limites de cet espace.

A travers le corpus nous pouvons remarquer que le trait de changement n'est pas limité à l'espace traversé, on peut traverser une ville depuis une maison se situant dans la ville, il y aura changement puisque X va d'un point à un autre et parce que les points de départ et d'arrivée ne sont pas confondus, sauf si le co-texte le mentionne (traverser pour revenir au point de départ).

4.3.5.5 Quelle classe pour traverser ?

D'après les auteurs et en vertu des caractéristiques de ses emplois absolus, le verbe *traverser* en tant que tel est caractéristique de la classe des verbes d'accomplissement, mais uniquement dans un sens prototypique où son emploi implique un changement d'espace et l'atteinte d'un point culminant. En effet, en contexte, la structure prédicative du verbe *traverser* est d'abord soumise à deux variables X et Y, tel que *X traverse Y*. La valeur de ces deux variables influe considérablement sur le degré de télicité du verbe et parfois même sur les autres variables. À côté de cela, le contexte élargi a un rôle prépondérant dans les valeurs aspectuelles et le bornage de ce verbe.

Le bornage temporel intrinsèque correspond au bornage spatial uniquement dans le cas où l'objet serait considéré comme une frontière entre deux lieux, sinon seul le contexte et donc le bornage extrinsèque permet de déterminer les bornes temporelles initiales et finales de la traversée. En fait, nous pouvons supposer que cet objet est vu comme unidimensionnel, comme le Temps.

Les dimensions spatio-temporelles du verbe *traverser* sont profondément intriquées à l'instar de la relation entre l'action et l'événement, et nous pouvons conclure à une double valeur aspectuelle de ce verbe : accomplissement et activité, les dimensions aspectuelles d'achèvement et d'état étant des valeurs secondaires mais pas du tout impossibles.

4.4 Modélisation

4.4.1 Préparation des modèles

La présente partie sera consacrée à l'explicitation de modèles qui font ressortir la plupart des traits aspectuels, y compris et avant tout le mouvement et le bornage, pour toutes les occurrences de *traverser*. Nous avons créé ces modèles pour l'analyse de données issues d'une base littéraire, en l'occurrence FRANTEXT. Toutefois, la modélisation que nous allons décrire permet de représenter toutes les occurrences de *traverser*, qu'elles soient des exemples construits ou non, car l'un de nos objectifs principaux est de démontrer que *traverser* est un verbe dont les emplois peuvent entrer dans les quatre classes de Vendler dont il a été question précédemment : accomplissement, activité, état et achèvement, et ce selon les variables co-textuelles qui entourent le verbe.

Tout comme Einstein (1986, traduction) et Muller & Denis (2004), nous admettons qu'il n'y a pas de distinction réelle entre Temps et Espace et qu'ils peuvent être réunis au même niveau d'un modèle quadridimensionnel. D'un point de vue physico-mathématique rapporté au langage naturel, nous pensons qu'il existe un Univers à quatre dimensions inséparables, et que le Temps en tant qu'entité unidimensionnelle n'est pas à séparer de l'Espace tridimensionnel. La distinction essentielle que nous faisons se situe dans la représentation graphique sur deux axes : l'axe temporel et les trois dimensions spatiales « aplaties » sur l'axe spatial.

Grâce à l'étude systématique du corpus, nous avons imaginé une représentation orthogonale des sens de *traverser*. En effet, la plupart des emplois du verbe *traverser* s'inscrivent dans des cadres spatiaux et temporels identifiables ; les recherches des bornes temporelles et spatiales s'effectuent indifféremment dans le co-texte et sont intriquées, comme nous l'avons déjà dit.

En fait, les emplois spatiaux et temporels du verbe semblent conceptuellement proches, bien que leurs paradigmes de traductions respectifs ne soient pas toujours comparables. Ainsi, nous partons de l'hypothèse que Temps et Espace doivent

appartenir à un même champ, et conséquemment apparaître dans un même repère mathématique orthonormé, lequel doit préserver la nature dimensionnelle de chacun. En outre, toutes les occurrences décrites par le truchement de ce modèle le seront systématiquement en fonction du temps et de l'espace, même pour les occurrences qui n'en nécessitent pas la représentation.

4.4.2 Caractéristiques des modèles

4.4.2.1 Forme mathématique

Pour des raisons essentiellement pratiques de lisibilité, nos modèles ne comportent donc que deux axes, l'un spatial et l'autre temporel. Les trois dimensions spatiales sont projetées sur un même axe, nonobstant la possibilité de transformer l'axe spatial en une succession de surfaces ou de volumes selon les cas.

Notre modélisation de *traverser* se fait dans un repère orthonormé. Ses caractéristiques sont les suivantes :

- Le repère orthonormé est en apparence bidimensionnel (2D), il manque donc deux dimensions spatiales. Nous ne perdons pas de vue que l'objet *traversé* peut être perçu comme une ligne (1D), une surface (2D) ou un volume (3D), et que tous les trois existent à tout instant du temps, laquelle est la quatrième dimension. D'un autre côté, nous avons également établi que la détermination des dimensions apparentes du sujet *traversant* n'est pas pertinente dans la recherche des traits aspectuels, et donc sans effet sur notre modèle. Par conséquent, beaucoup d'emplois du verbe nécessiteront une représentation de l'objet en quatre dimensions, mais cela n'est pas vraiment un problème, comme nous allons le montrer, tant que le sujet ne nécessite pas d'être représenté lui aussi en perspective.
- Le temps est en abscisse et l'espace en ordonnée pour tous nos modèles.
- Nous l'avons déjà dit, le contexte (présumé ou non) pose les bornes et notre représentation en nécessite six : 2 bornes spatiales internes pour fixer les limites de l'objet (B_o1 et B_o2), deux bornes spatiales externes pour fixer les limites de la traversée par rapport à l'objet (B_t1 et B_t2) et deux bornes temporelles du début de l'évènement jusqu'à son terme (E_1 et E_2).
- Comme vous pourrez le voir sur nos modèles, le temps est partagé en lignes verticales noires, chaque ligne représentant un instant ou plus

précisément la position du sujet et de l'objet dans l'espace à un instant du temps. Nous n'avons pas creusé plus avant la notion « d'instant du temps » car cela nous amènerait à considérer que tout laps de temps contient un nombre infini d'instant, et cela rendrait immédiatement caduc notre mode de représentation. Aussi, nous nous contenterons d'admettre ce précepte d'infinitude tout en adoptant une vision plus souple dans nos dessins.

- Dans la mesure du possible et du lisible, l'objet *traversé* est figuré par une ligne (1D, typiquement une rivière ou une frontière), une surface (2D, comme un pays plat) ou un volume (3D, pas de prototype) répété(s) à chaque instant du temps.
- Le sujet *qui traverse* est présenté comme une succession de points figurant la position du sujet par rapport à l'objet suivant l'avancée du temps. Afin de présenter des modèles homogènes pour tous les emplois décrits, cette succession de points ne représente que **le point d'impact ou de contact du sujet traversant sur l'objet traversé**. C'est pourquoi certains modèles ne figurent pas exactement la nature de la *traversée* dès lors que celle-ci est déterminée par la nature et la position spatiale du sujet (comme les transpercées, dont nous dirons quelques mots ultérieurement). Les points sont prolongés par une flèche lorsque la traversée s'effectue dans une certaine direction, donc avec un mouvement. Pour les emplois statiques, théoriquement sans direction, la tête de la flèche n'est pas présente, la traversée n'étant pas orientée.

4.4.2.2 Ce que montrent les modèles

Avec seulement trois phrases prototypiques contenant *traverser*, nous avons pu montrer par la modélisation son appartenance potentielle aux quatre classes aspectuelles majeures.

Bien entendu, nos modèles sont particulièrement utiles et efficaces pour la prise en compte de facteurs temporels extrinsèques au prédicat, voire à la structure argumentale ; c'est pourquoi nous l'employons notamment pour l'analyse de corpus, comme nous l'indiquions au préalable.

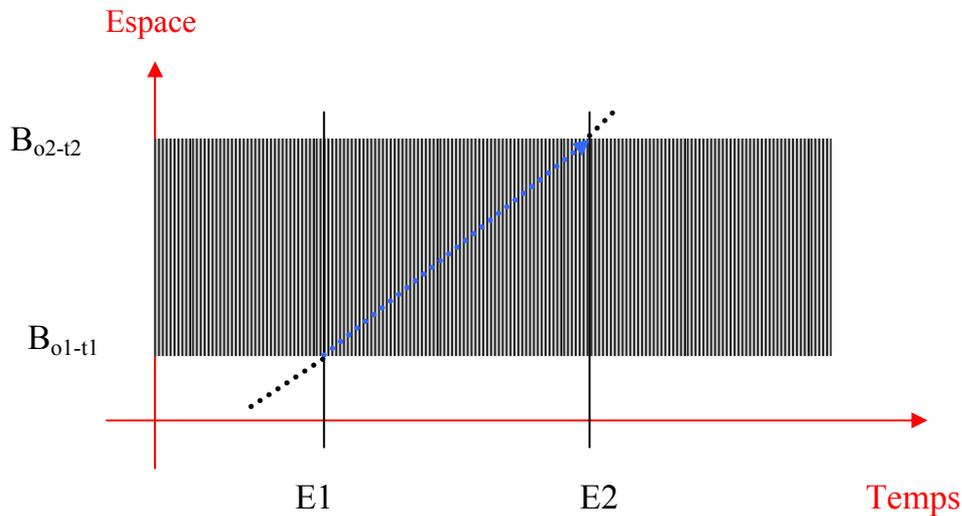
4.4.3 Les modèles spatiaux

4.4.3.1 Premier modèle : Emploi prototypique, « *traverser la rue* »

« Au milieu de la place, Robert se retourna. Le père Girard était debout sur le pas de sa porte. Robert prit la grand-rue et continua jusqu'à la route de

l'Arbresle. Là, il **traversa** et alla s'asseoir sur la murette qui borde le terrain vague. »
Clavel Bernard /Malataverne/1960

Figure 24 : Modèle 1 pour traverser



Ceci est la représentation de parcours dans un espace vu comme étant unidimensionnel. L'objet traversé n'est pas la surface qui couvre le repère mais bien chaque ligne verticale noire indiquant sa position à un instant donné, position qui ne change quasiment jamais¹⁸¹. Le sujet bouge également dans le temps sur cette surface, comme le montre la flèche de points successifs. Cette dernière est rectiligne, ce qui signifie que la traversée en tant que mouvement est ininterrompue et qu'elle s'effectue en ligne droite. Sauf indication contraire dans le co-texte de certaines occurrences (cf. figure 17 dans les diagrammes ontologiques, où le parcours s'effectue en zigzags), nous partons du principe que les traversées prototypiques sont toutes rectilignes et ininterrompues.

Les bornes B_{o1} et B_{o2} de l'objet, i.e. les limites de la rue dans sa largeur, sont confondues avec les bornes de l'action B_{t1} et B_{t2} , lesquelles fixent les limites de la traversée aussi bien spatialement que temporellement. En d'autres termes, au-delà des bornes B_{tx} , il n'y a pas encore et/ou plus, en principe, de traversée. Les bornes temporelles de l'événement E_1 et E_2 fixent les limites de la traversée dans le temps. Par définition, les bornes de l'action et celles de l'événement sont orthogonalement concomitantes.

Malgré ce que nous venons de dire, une des difficultés qui apparaît au travers de ce schéma est que les bornes de l'action se confondent avec les bornes de l'objet. Or,

¹⁸¹ Un objet traversé est statique selon notre définition. Plus exactement, nous considérons que le fait qu'il soit ou non en mouvement n'influe pas sur la détermination des bornes de la traversée sauf cas exceptionnels, peu ou jamais rencontrés dans la littérature.

nous admettons que *traverser la rue* n'est pas une activité circonscrite à la chaussée lexicalisée par *rue*, mais doit plutôt représenter le *passage d'un côté à l'autre* de cette *rue*. Autrement dit, les bornes véritables de la traversée dans ce cas sont celles de l'espace qui borde la chaussée, généralement les trottoirs ou les bas-côtés (cf. aussi les prolongements co-textuels de type « jusqu'à la boutique en face... »). Par convention, nous considérons la rue comme un ensemble à bornes identifiables en admettant pour l'instant que ce qui borde la rue fait partie de la rue lorsqu'on la considère comme l'objet d'une traversée.

Pour cet emploi dit prototypique, il y a :

- +télicité, car les bornes de l'évènement se confondent avec les bornes de l'action ;
- +changement car la situation après traversée est différente de ce qu'elle était avant qu'elle fût entreprise (test d'implication) ;
- +mouvement puisque le sujet bouge dans l'espace.

Il s'agit donc d'un ACCOMPLISSEMENT.

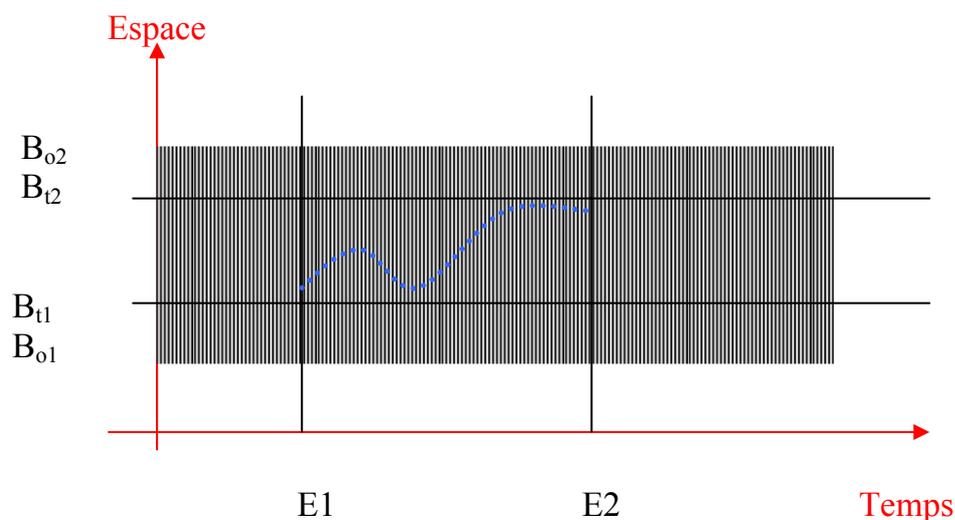
4.4.3.2 Second modèle : Emploi habituel, « *traverser un territoire*¹⁸² »

« Il a eu aussi une passion pour une certaine Josette (...) qu'il a emmenée **traverser** la France sur sa chère moto. Du coup, au cours d'une discussion sur les prochaines vacances, vous lui proposez un rallye à deux sur une Mitsubishi 900. »

Buron Nicole de /"Chéri, tu m'écoutes ? : alors répète ce que je viens de dire..." /1998

Figure 25 : Modèle 2 pour traverser

¹⁸² Par « territoire », nous entendons l'ensemble des espaces sur lesquels on peut se déplacer et qui peuvent être assimilés à des surfaces. Une ville peut faire partie de cet ensemble mais n'en est pas le prototype car elle peut aussi bien être perçue comme un objet à trois dimensions (jungle urbaine, labyrinthe, ensemble de résidences) qu'à deux dimensions (réseau de rues, chemins et routes)



L'objet est ici vu comme bidimensionnel, comme une surface. C'est la représentation d'un parcours dans un espace à deux dimensions. Dans l'occurrence ci-dessus, le sujet traversant est *la moto*, bien qu'il soit exprimé en termes Humain-animé (*Josette et vous SUR la moto*) ; *la France* est perceptible en tant qu'ensemble de routes sur lesquelles *Josette et vous* se déplacent par *moto* interposée.

A la place d'une flèche, un ruban de points aurait montré ici que le parcours peut s'effectuer dans n'importe quelle partie du pays traversé ; ce qui importe est qu'il n'y ait pas de rupture (i.e. de « téléportation ») dans la traversée du sujet. En effet, il n'est pas envisageable que le parcours passe de Guéret à Landivisiau d'un instant du temps au suivant¹⁸³.

Pour cet exemple, on notera :

- +changement car la situation après traversée est différente de ce qu'elle était avant qu'elle fût entreprise ;
- +mouvement car le sujet se déplace, en considérant une nouvelle fois que le véritable sujet en mouvement est la moto ;
- pas de télélicité car les bornes de l'action sont internes à celles de l'objet dans les limites de l'évènement.

Il s'agit donc d'une ACTIVITÉ.

¹⁸³ Une nouvelle fois, ce principe est valable seulement si le co-texte n'apporte pas de contradiction. *Josette et vous* peuvent très bien avoir interrompu leur traversée à moto en un point donné et avoir voyagé par d'autres moyens, moto à l'arrêt.

4.4.3.3 Second modèle bis : Emploi habituel, « traverser une ville »

(...) un observatoire d'où il pourrait surveiller les opérations sans se mettre en péril. Protégé au milieu d'une escorte, silencieux, l'Empereur s'éloigna au petit trot en **traversant** Essling, ensuite un grand bois touffu qui reliait ce village au Danube.

Rambaud Patrick /La Bataille/1997

L'objet est également vu ici comme bidimensionnel comme une surface ou éventuellement tridimensionnel s'il est perçu en tant qu'obstacle sur un parcours plus vaste que la simple traversée. Il s'agit de la représentation d'un parcours dans un espace à deux dimensions, avec franchissement total de l'espace visé. Dans l'occurrence ci-dessus, le sujet traversant est *l'Empereur et son escorte*, exprimé en termes Humain-animé ; *Essling* est perceptible en tant que « entité citadine » dans laquelle *l'Empereur et son escorte* se déplacent avec comme objectif de la dépasser, comme l'atteste le co-texte droit.

Contrairement à l'exemple précédent, la flèche de points dans le modèle doit être rétablie dans tous les cas en lieu et place d'un ruban, parce que le parcours s'effectue d'un point A extérieur à la ville jusqu'à un point B également extérieur et situé de l'autre côté de la ville.

Pour cet exemple, on notera :

- +changement car la situation après traversée est différente de ce qu'elle était avant qu'elle fût entreprise ;
- +mouvement car le sujet se déplace ;
- +télicité car les bornes de l'action sont confondues à celles de l'objet, conformément à l'orientation de l'objet que nous avons choisie dans le modèle, et dans les limites de l'évènement ;
- -momentané car le procès est vu avec durée : il a fallu « un certain temps » au cortège pour *traverser la ville*.

Il s'agit donc là d'un ACCOMPLISSEMENT.

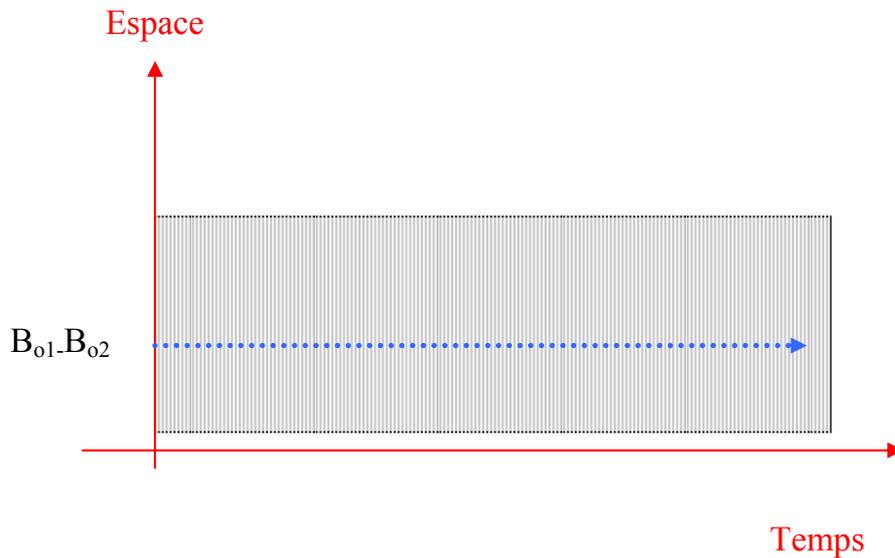
4.4.3.4 Troisième modèle : transpercée dynamique

(4) = Le général Espagne reçoit une balle qui traverse la cuirasse.

TR : A bullet went/shot through General Espagne's breast-plate

(transpiercing)

Figure 26 : Modèle 3 pour traverser



Dans ce dessin, nous n'avons pas représenté la situation avant et après la transpercée. Nous ne pouvons pas ignorer dans ce cas que la traversée fait partie d'une action plus vaste qui englobe le point de départ de la balle (lorsqu'elle sort du canon du fusil), son parcours dans l'atmosphère avant d'atteindre *la cuirasse*, la traversée de la cuirasse et son point d'arrivée qui est la poitrine du Général.

Les caractéristiques de ce modèle sont :

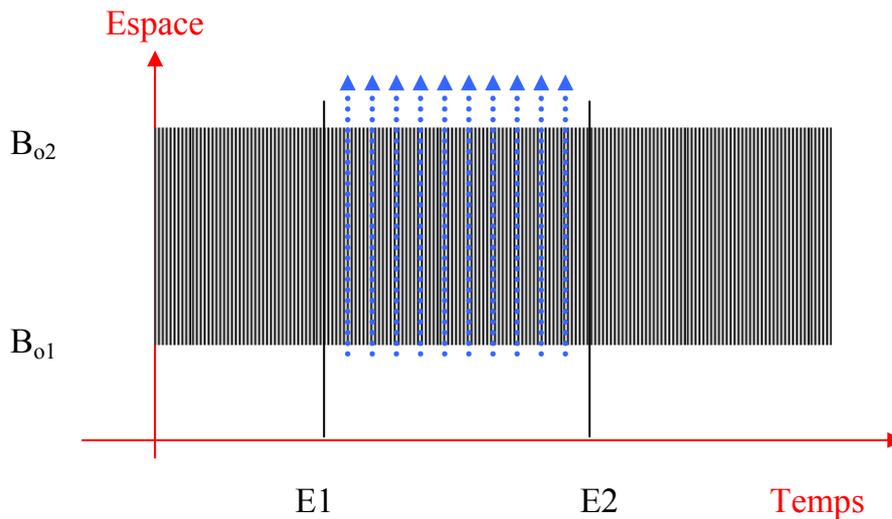
- +changement car la situation après traversée est différente de ce qu'elle était avant qu'elle fût entreprise ;
- +mouvement car le sujet se déplace ;
- +télicité car les bornes de l'action sont confondues à celles de l'objet, conformément à l'orientation de l'objet que nous avons choisie dans le modèle, et dans les limites de l'évènement ;
- +momentané car le procès est vu comme immédiat.

La transpercée de la cuirasse est ainsi considérée comme un climax, montrant qu'il s'agit d'un ACHÈVEMENT.

4.4.3.5 Transpercée statique

La flèche traverse la porte et RESTE en travers de la porte. La flèche de points successifs figurant le sujet est parallèle à l'axe temporel et perpendiculaire à l'axe spatial, ce qui indique qu'il ne bouge pas. Les bornes de l'action ne sont pas repérables et celles de l'évènement demandent à être précisées.

Figure 27 : Modèle 4 pour traverser



Nous relevons :

- -changement car la situation est décrite à un certain moment du temps, sans mention de l'état antérieur (en l'occurrence) ni éventuellement postérieur ;
- -mouvement car le sujet est immobile ;
- - télicité car les bornes de l'évènement ne sont pas repérables ;
- - momentané car le procès est vu sans durée définie.

Il n'y a donc ni changement, ni mouvement ni télicité. Il s'agit là d'un ÉTAT. Il est amusant de noter que la transpercée dynamique et son pendant statique s'opposent pour tous les critères de distinction que nous avons utilisés, bien qu'elles ne diffèrent dans leur définition que par le trait de mouvement (celui-ci impliquant tous les autres critères).

4.4.4 Les emplois métaphoriques ou apparentés

4.4.4.1 Traversées temporelles : parcours dans une période de temps

« Les pyramides ont traversé les siècles »

TR : Pyramids travelled down (or went through) the centuries

Figure 28 : Modèle 5 pour traverser



Les modèles pour les sens temporels de *traverser* sont très proches des *transpercées dynamiques*. En effet, la flèche de points est orientée et toujours perpendiculaire à l'axe spatial. En revanche, il est non seulement parallèle à l'axe temporel mais il fusionne avec totalement, à l'intérieur des bornes de l'évènement, lesquelles ne sont pas dessinées ici car nous ignorons où les placer hors contexte. L'objet traversé n'a donc pas d'autre substance que le Temps lui-même (0D spatial).

Pour cet emploi, nous relevons :

- +/-changement car si le sujet est toujours le même aux deux bornes de l'évènement, l'usure peut l'avoir modifié. D'autre part, il n'est pas exclu qu'un objet à deux instants du Temps ne soit pas nécessairement le même au sens physique.
- -mouvement car le sujet est immobile, ou à tout le moins n'induit pas de mouvement repérable qui influencerait sur la nature de la traversée ;

- +/-télicité car les bornes de l'événement sont volontiers extensibles (ce qui a traversé les siècles antérieurs peut peut-être continuer à traverser les prochains) ;
- -momentané car le procès s'inscrit intrinsèquement dans une certaine durée.

A première vue, il peut s'agir d'un ETAT si l'on neutralise les traits de changement et de télicité, et nous nous contenterons de cette conclusion intermédiaire à ce stade de l'analyse en attendant de l'enrichir incessamment en confrontant les emplois temporels et certaines transpercées.

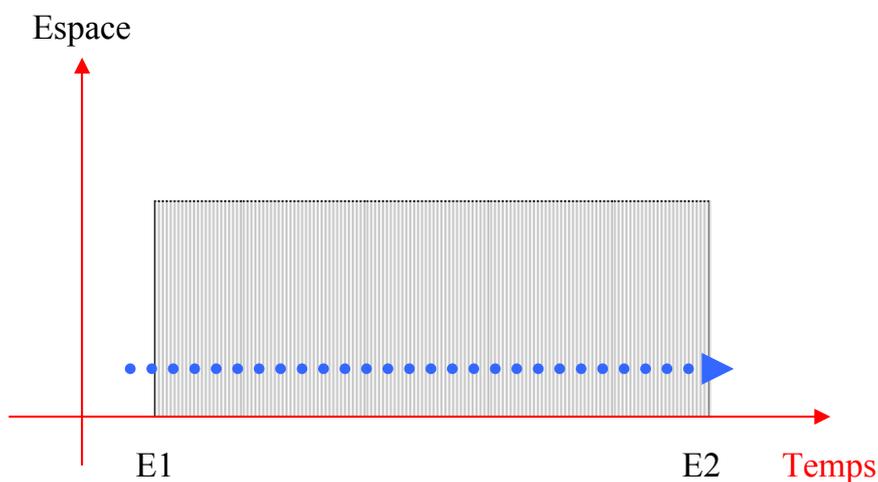
Notons que la simplicité apparente des modèles temporels est due au fait qu'ils ne font agir qu'une seule dimension alors que les modèles spatiaux ont également besoin du Temps pour leur représentation.

4.4.4.2 Traversées temporelles « démétaphorisées » : dépassement d'un événement vu comme ayant une substance temporelle

Jean a traversé la crise.

TR : John went through the crisis.

Figure 29 : Modèle 5bis pour traverser



Cet emploi illustre l'ensemble des traversées que nous appelons temporelles mais qui ne s'effectuent plus au sein d'un objet présenté comme une durée mais au

travers d'un événement temporellement borné : *crise, guerre, etc.* Nous les avons distinguées du précédent pour deux raisons :

- Premièrement, la période traversée possède une substance, une réalité quasi-matérielle que n'ont pas les laps de temps. Ces événements sont intrinsèquement bornés, alors que les précédents ne le sont que par ajout des informations concomitantes.
- En second lieu et conformément au résultat qui nous intéresse, leurs traductions en anglais ne sont pas équivalentes d'après le dictionnaire Robert & Collins.

La flèche de points est toujours perpendiculaire à l'axe spatial. En revanche, les bornes E de l'événement sont clairement identifiées. L'objet traversé a donc ici plus de substance que le Temps lui-même, ce qui lui confère une dimension supplémentaire, certes pas précisément spatiale mais qui y ressemble étrangement, et en fait un emploi métaphorique à l'origine.

Nous relevons pour cet emploi :

- +/-changement car le sujet peut avoir été modifié¹⁸⁴ au cours de cette traversée ;
- -mouvement car le sujet est immobile ou à tout le moins n'induit pas de mouvement repérable qui influencerait sur la nature de la traversée¹⁸⁵ ;
- + télélicité car les bornes de l'événement sont identifiées, et l'événement et l'action sont orthogonales ;
- +/-momentané car le procès est vu dans une certaine durée mais il peut y avoir insistance sur le point culminant de l'action plutôt que sur la traversée elle-même.

A première vue, il peut s'agir d'un accomplissement ou d'un achèvement mais les informations co-textuelles dont nous disposons ne sont pas suffisamment explicites

¹⁸⁴ Nous noterons d'ailleurs que les traversées temporelles de cette catégorie induisent fréquemment un changement d'état du sujet qui est notamment affecté moralement quand il s'agit d'un humain.

¹⁸⁵ Il n'est évidemment pas question d'imaginer un sujet humain attendant « immobile sur une chaise » la fin d'une crise ou d'une guerre. Toutefois, le focus porte exclusivement sur l'événement et son dépassement, non sur tous les mouvements qui ont pu être effectués par le sujet pendant son déroulement.

pour en tirer des conclusions définitives à ce stade. Une analyse contrastive plus approfondie montre que la traduction anglaise peut varier selon la classe aspectuelle dans laquelle nous placerons cet emploi. Cette variation en fonction du co-texte et des traits aspectuels concerne également les autres acceptions de *traverser*.

4.5 Analyse contrastive

4.5.1 Rappel et retour sur le paradigme de traductions

Comme une évidence, nous rappelons que *traverser* hors contexte ne nous donne pas assez d'indices pour conjecturer quelles sont ses valeurs aspectuelles entières. Nous allons montrer maintenant quelles relations les traductions potentielles du verbe entretiennent avec la saturation des traits aspectuels que nous avons tenté d'explicitier.

Comme nous l'avons indiqué, *traverser* peut être traduit par un grand nombre de verbes anglais selon seulement un seul dictionnaire bilingue (Robert & Collins, dernière édition), entre cinquante et cent, probablement davantage (pour une explication au sujet de l'imprécision, nous nous référons à Haton 2003¹⁸⁶).

Rappelons en quelques mots la typologie transitoire de ces traductions en anglais :

- Le plus générique, *to cross*, est essentiellement spatial. Son cas ne sera pas évoqué dans cette partie dans la mesure où nous considérons qu'il neutralise les valeurs aspectuelles correspondantes au même titre que *traverser* tout seul en français.
- Quelques autres sont très spécifiques (*to negotiate, to sail, to ford, etc.*) ; ces verbes, en tant que traductions éventuelles de *traverser*, sont fortement dépendants de la structure argumentale de *traverser* puisque chaque lexème verbal anglais fait référence à un objet bien précis.
- Toutes les autres traductions ou presque sont établies sur l'addition d'une préposition à un verbe de mouvement (y compris *to get*), créant un paradigme « hiérarchisé » : *to come, go* et *get* sont les verbes les plus communs, ils sont génériques pour la traduction de *traverser*, que ses emplois soient spatiaux, temporels ou métaphoriques. D'autre part, *to walk, run, swim* et ainsi de suite, illustrent la compositionnalité de l'anglais quant aux prédicats verbaux en fonction du mouvement

¹⁸⁶ Dans notre premier article de 2003, nous dégageons une hiérarchie des liens de traduction en précisant que les liens les plus distendus concernaient les traductions indirectes, i.e. qui passent par un terme intermédiaire avant leur validation.

effectué lors de la traversée. Nous avons déjà insisté sur le fait que *traverser*, en tant que préverbe¹⁸⁷ apparent¹⁸⁸, porte en lui-même la marque d'une préposition et de l'aspect qu'on associe à cette dernière. Toutefois, ses équivalents en anglais ne possèdent pas cette caractéristique d'inclusion de la préposition (à l'exception peut-être de *to cross* qui se rapporte sans doute à *across*), mais cette langue autorise aisément la mise en relation de verbes de mouvement ou apparentés avec des prépositions plus variées et assez marquées aspectuellement.

Pour la suite de notre analyse contrastive, c'est la distribution de ces prépositions anglaises ainsi que leur impact sur le calcul de l'aspect qui vont nous intéresser.

4.5.2 L'intervention des prépositions *across/through/over* dans les parcours spatiaux

4.5.2.1 Distribution complémentaire apparente

Auparavant, nous avons signalé que l'anglais permet l'utilisation de trois prépositions *over*, *across* ou *through* afin d'exprimer un même concept avec plus de précision temporelle que les monolexèmes comme *to cross*, car leur distribution contextuelle n'est pas aléatoire. En revanche, il est plus difficile de dessiner une carte aspectuelle de *traverser* en français sans observer systématiquement son co-texte large.

L'hypothèse selon laquelle *through*, *over* et *across* coexistent au sein d'une distribution complémentaire est conforme à notre observation du lexique et a été partiellement confirmée par nos relecteurs bilingues¹⁸⁹; le choix de la préposition appropriée dépend fortement de deux facteurs : si l'espace où s'effectue la traversée est bidimensionnel ou tridimensionnel et aussi quel est le sentiment du locuteur au sujet de l'espace ou de la période "à traverser". Ce point de vue induit que c'est le locuteur qui dispose des clés de l'univers aspectuel associé à la situation qu'il décrit. Ainsi, la

¹⁸⁷ Un préverbe est un préfixe placé avant le verbe. Dans le cas de *traverser*, on pourrait être tenté de penser que le verbe est construit à partir de la préposition *à travers*.

¹⁸⁸ En réalité, *traverser* n'est à l'origine ni un dénominal construit sur *travers* ou *traversée* ni le résultat de la verbalisation d'un préverbe. D'après les informations étymologiques du TLF, il procède directement du verbe latin *transversare* qui est lui-même issu de l'adjectif *transversus* qui signifie *transerval* ou *oblique*. Le sens premier de *traverser*, à savoir « être transversal », est donc motivé par le sens de l'adjectif. Dans ces conditions, il s'agit plutôt d'un verbe « désadjectival » bien que cela ne soit plus évident en synchronie.

¹⁸⁹ Merci à Hilary Deneufchâtel-Inwood (Anglaise, bilingue), Judith Kamalski (Néerlandaise, trilingue), Yvon Kéromnès et Emmanuel Haton (Français, bilingues) pour avoir bien voulu relire et critiquer nos données.

préposition sélectionnée doit être conforme à sa vision à lui de ce qui est *traversé*. D'un autre côté, le traducteur devient le seul maître à bord à ce sujet dès lors que le locuteur initial est absent de l'exercice de traduction.

Il faut donc admettre que le dimensionnalisme diffère parfois entre la réalité et le point de vue de celui qui parle. C'est un des critères de détermination de l'aspect dans le discours. Par ailleurs, le récepteur du discours peut également interpréter une occurrence de façon différente en fonction de sa propre vision du monde.

4.5.2.2 Nature de la distribution

À partir des différents relevés que nous avons effectués, des hypothèses que nous avons émises et des commentaires de nos relecteurs, nous sommes parvenus à la distribution suivante pour les trois prépositions concernées :

- *Across* est employé pour des parcours spatiaux « classiques » surtout dans les espaces vus en tant qu'unidimensionnels, plus exceptionnellement bi- ou tridimensionnels. Cette préposition est la moins aspectuellement marquée des trois et la plus intimement liée au verbe d'origine en tant qu'elle dénote le parcours en tant que tel.
- *Through* est employé pour des parcours spatiaux dans un espace à trois dimensions ou considéré comme tel, et également pour des emplois temporels « non marqués », pour les transpercées (typiquement tridimensionnelles) et les vraies métaphores.
- *Over* peut être employé à propos d'espaces n-dimensionnels, pour des sens spatiaux et temporels dans des conditions d'emploi très précises : dès lors que l'accomplissement est plus important que l'activité dans l'esprit du locuteur (avec notions de climax et d'achèvement, voire de dépassement au sens moral, dans la plupart des cas).

Bien que nous la présentions et la « souhaitions » comme telle, la distribution que nous suggérons n'est pas strictement complémentaire. En effet, il peut y avoir recouvrement entre les trois prépositions, notamment entre *across* et *through* d'une part, entre *through* et *over* d'autre part. La confusion entre *across* et *over* est beaucoup moins évidente, les traits qu'ils véhiculent étant trop éloignés. Grâce à nos relecteurs, nous avons pu confirmer nos hypothèses au sujet de cette distribution non-aléatoire mais parfois concurrente. Eux et nous pensons que le sens de *traverser* et ses traductions peuvent changer sous l'influence de la configuration de l'espace.

4.5.2.3 Du prédicat verbal au co-texte : plusieurs façons d'interpréter la même occurrence

Pour illustrer notre hypothèse, observons l'occurrence suivante :

Jean a traversé la forêt (traduction générique : *John crossed the forest*)

Traverser peut ici être traduit par des prédicats contenant chacune des trois prépositions, comme confirmation de ce que Henriette Hendriks¹⁹⁰ disait dans la conférence inaugurale au congrès STLL05 à Rijeka.

Trois interprétations sont disponibles :

- Premièrement, la forêt est une surface, vue comme un sol d'après le locuteur, sur laquelle Jean peut marcher. La présence d'arbres ou d'obstacles au sens large n'est pas pertinente pour borner l'action ; en outre, il n'y a pas trace de limites spatiales dans l'esprit du locuteur dans ce cas. La *traversée* est comparable à une *promenade*. Pour cette interprétation, la forêt doit être vue comme bidimensionnelle. *Across* est le plus susceptible d'être employé, *over* est exclu car le trait d'accomplissement qu'il véhicule n'est pas compatible avec l'activité *promenade*.
- Deuxièmement, la forêt est un volume, pas seulement une terre sur laquelle on peut marcher. Dans ce volume, beaucoup d'obstacles empêchent le voyage d'être linéaire : arbres, broussailles, branchages, ronces, rochers et ainsi de suite. Dans l'esprit du locuteur, la forêt est tridimensionnelle. La traversée est comparable à un passage au travers d'un ensemble d'éléments entravants. Pour cette interprétation, *through* est le plus susceptible d'apparaître, *over* est toujours exclu pour les mêmes raisons que précédemment. *Across* demeure possible à la condition que le co-texte pallie sa relative fadeur aspectuelle par des descriptions supplémentaires.
- Troisièmement, dans l'esprit du locuteur, la forêt n'est pas une surface ni un volume ; la forêt est... juste une forêt. En d'autres termes, elle constitue juste une étape ou un obstacle dans le voyage de Jean, entre le départ d'un point A jusqu'à l'arrivée au point B (*Jean a traversé la*

¹⁹⁰ Professeur à l'Université de Cambridge (UK), Henriette Hendriks inaugurait le congrès STLL05 (« Space and Time in language and literature » à Rijeka en 2005) auquel Laurent Mascherin et moi avons également participé. Sa conférence portait sur l'expression de l'espace en anglais au moyen de certaines prépositions, dont celles que nous présentons ici.

forêt pour atteindre le village voisin = John walked/went over the forest to reach next village). D'après cette troisième interprétation, la reconnaissance des dimensions spatiales de la forêt n'importe pas ; ce qui est pertinent est d'atteindre le point B, qui est en dehors de la forêt. La traversée est comparable à une étape dans un parcours qui l'inclut. C'est un accomplissement et, pour cette interprétation, *over* est approprié si l'on se focalise davantage sur le point d'arrivée que sur le parcours, mais *across* et *through* sont également possibles, le choix dépendant à nouveau du co-texte.

Nous avons rencontré ce genre d'occurrence très souvent dans notre corpus francophone, et évidemment la décontextualisation ne permet pas une lecture complète des traits aspectuels mis en jeu. Il est nécessaire d'observer le co-texte large pour les saturer en français, alors que l'anglais peut partiellement fournir ces informations dans la structure lexicale du verbe. Voyons ce que cela donne en contexte :

« (...) serait qu'ils envoient des patrouilles, mais s'ils le font, ils sont obligés de passer par le gué, près de la ferme Badin, ailleurs ce n'est pas possible, il leur faudrait **traverser** des forêts de ronces, mais dès que Badin les voit, il expédie son fils qui connaît les raccourcis et vient nous avertir. »

Joffo Joseph /Un sac de billes/1973

Dans cet exemple, le co-texte restreint ne montre pas clairement à laquelle des trois interprétations ci-dessus nous avons affaire puisque la phrase seule ne nous permet pas de savoir si le but de périple est bien au-delà de *la forêt de ronces*. Toutefois, la concordance de plusieurs éléments grammaticaux et co-textuels (emploi du conditionnel et de l'auxiliaire modal *falloir*, l'objet complément *forêt de ronces*) nous donne quelques indices qui confirment que *over* est utilisable pour traduire cette occurrence, hypothèse corroborée par le co-texte antérieur qui montre que le point d'arrivée se trouve bien **après la forêt de ronces**.

« (...) étuver dans une petite usine à cinq kilomètres. Nous avons pensé que deux hommes pourraient sortir du camp en se cachant dans ce camion. à trois kilomètres le camion **traversait une forêt assez épaisse sur une route solitaire**. C'était là qu'il fallait sauter du camion. Ensuite, ce n'était plus qu'une affaire de marche à pied, de ravitaillement (...) »

Tournier Michel /Le Coq de bruyère/1978

La précision au sujet de la forêt « assez épaisse » semble indiquer que le tridimensionnalisme est pertinent dans l'analyse de cette occurrence. D'autre part, l'élément co-textuel « c'était là qu'il fallait sauter du camion » montre que la traversée de la forêt n'est pas accomplie car elle s'arrête en son intérieur. Pour ces deux raisons, *through* est sans doute possible le meilleur choix parmi les trois prépositions.

« (...) avec des filles dans des boîtes de Hong-Kong ou de Berlin. Pandora riait. Elle était guérie. Elle était vaccinée. Elle avait subi l'épreuve qui allait lui permettre de **traverser sans encombre les forêts enchantées** et les fournaies ardentes. Le mal qui la rongait était passé à Luis Miguel. De temps en temps, fruit amer, j'imagine, (...) »

Ormesson Jean d' /Tous les hommes sont fous/1986

Cette occurrence de *traverser la forêt* est manifestement une métaphore, laquelle n'est pas comprise dans les acceptions que nous avons relevées dans les différents dictionnaires, à moins de considérer cet exemple comme un emploi temporel pour lequel *les forêts enchantées et les fournaies ardentes* seraient comparables à des crises ou des événements d'un type particulier.

Dans la majorité des cas, les traductions des emplois métaphoriques de *traverser* contiennent la préposition *through*.

Il existe beaucoup d'autres types de parcours influencés par le co-texte dans Frantext et il eut été trop fastidieux d'en faire un relevé exhaustif. En voici tout de même quelques spécimens synthétisés pour compléter les précédents :

- Parcours spatial, description du paysage, pas de connotation d'obstacle ni réellement d'étape → *across*. Co-texte pertinent.

« (...) la terre grasse et rouge verdoie adossée à une chaîne continue de montagnes qui la protège et l'ombre : le blé est déjà haut, la vigne avance ses premières feuilles. On **traverse** des forêts de néfliers en fruits ou des bosquets compacts d'orangers dont l'enivrant parfum assaille par vagues successives. Patras est affreux : une ville écrasée, (...) »

Déon Michel /Le Balcon de Spetsai/1961

- Emploi métaphorique, voyage initiatique. Dépassement de soi → *over*, ou passage mental dans un environnement touffu → *through*. Co-texte nécessaire.

« KADIDJA, regardant Slimane : Cinglée ? SI SLIMANE, souriant : Comme toi tout à l'heure... elle **traverse** la forêt... Elle se dépiaute de la raison, pour arriver pure et comme toi connaître les rapports de l'espace et de la course et le nom des mouches. »

Genet Jean /Les Paravents/1961

« (...) avec des filles dans des boîtes de Hong-Kong ou de Berlin. Pandora riait. Elle était guérie. Elle était vaccinée. Elle avait subi l'épreuve qui allait lui permettre de **traverser sans encombre les forêts enchantées** et les fournaies

ardentes. Le mal qui la rongait était passé à Luis Miguel. De temps en temps, fruit amer, j'imagine, (...) »

Ormesson Jean d' /Tous les hommes sont fous/1986

- Parcours spatial, pas de but hors l'objet, celui-ci vu comme un environnement hostile → *through*. Co-texte nécessaire.

« (...) et vendre n' étaient en réalité qu' une seule et même chose un simple besoin celui de se rassurer, comme des gamins qui sifflent ou chantent fort pour se donner du courage en **traversant** une forêt la nuit, ce qui expliquait pourquoi le chant en chœur faisait partie au même titre que le maniement d' armes ou les exercices de tir du programme d' instruction (...) »

Simon Claude /La Route des Flandres/1960

- État spatial, sujet inanimé. Co-texte pertinent. Dans l'exemple suivant, les adjectifs « épaisse » et « infestée de... » dénotent la présence d'obstacles imposants et proposent une lecture de la forêt comme un véritable espace tridimensionnel. « La majeure partie du trajet » n'est pas un sujet possédant un but d'arrivée au-delà de « la forêt » → *through*.

« La pirogue métallique de terre

Le gouffre de la grande hyène noire À cette époque, le pays était encore fortement boisé. La majeure partie du trajet **traversait** une épaisse forêt infestée d'hyènes et de panthères. À mi-route, il fallait longer sur notre gauche une grande et sombre excavation appelée Dounfing. »

Bœ Amadou Hampôté /Amkoullel, l'enfant peul/1991

- Parcours spatial, nécessité d'aller au-delà et entre-temps de *traverser* un espace jonché d'obstacles → *through*, *over*, ou les deux. Co-texte pertinent.

« (...) serait qu'ils envoient des patrouilles, mais s'ils le font, ils sont obligés de passer par le gué, près de la ferme Badin, ailleurs ce n'est pas possible, il leur faudrait **traverser** des forêts de ronces, mais dès que Badin les voit, il expédie son fils qui connaît les raccourcis et vient nous avertir. »

Joffo Joseph /Un sac de billes/1973

4.5.3 Les parcours temporels

Après avoir bâti nos représentations sur le principe du quadridimensionnalisme, il nous reste à répondre à un certain nombre de questions pour caractériser les emplois temporels comparativement aux emplois spatiaux :

- Le Temps est-il un « objet » dynamique ?

Voyons ce qui se produit quand l'objet traversé est un intervalle de temps. Initialement, nous pensions que les sens spatiaux et temporels devaient être opposés, car pour les parcours spatiaux dynamiques le sujet se déplace et l'objet est statique (avec une valeur durative, (1)), alors que les emplois temporels impliquent un sujet immobile et conceptuellement un objet mobile qui est le temps lui-même (2).

(1) Ils traversèrent une partie de la cour, *puis montèrent au premier étage* de la caserne.

Clavel Bernard /Le Cœur des vivants/1964

(2) (...) elle acquiert une dureté remarquable aux grandes chaleurs ; on *rend donc l'oeuvre quasiment inaltérable* grâce à la cuisson : la forme éphémère *devient ainsi apte à traverser* les siècles.

Fontaine Georges /La Céramique française/1965

Le sujet humain de (1), *ils*, s'oppose à l'inanimé *la forme éphémère* de (2). D'autre part, l'objet inanimé et spatialement repérable de (1), *une partie de la cour*, semble également s'opposer à l'objet *les siècles*, lequel correspond à une période temporelle floue. L'objet est dynamique dans cet emploi car c'est l'avancée du temps, i.e. sa *traversée*, qui est dénotée et non pas la durée elle-même.

- Peut-on transpercer le Temps ?

L'observation des modèles et des paradigmes de traduction nous fait plutôt tendre vers un rapprochement des emplois temporels avec les transpercées statiques. Par la modélisation, ils partagent la particularité de présenter un sujet *traversant* immobile et figuré par une succession de points formant une ligne perpendiculaire à l'axe spatial et parallèle à l'axe temporel, avec comme seule différence le fait que cette ligne est confondue avec l'axe temporel pour les premiers. Par ailleurs, les bornes de l'événement n'apparaissent pas clairement dans les modèles car la mention d'une traversée temporelle ou d'une transpercée statique ne présuppose pas (sans co-texte approprié) que le locuteur connaît précisément le moment où la traversée a débuté. Toutefois, la borne finale de l'événement peut être arbitrairement fixée au présent de l'énonciation, à la borne de l'énonciateur. Malgré cela, nous supposons que la *traversée* continue après l'énonciation et ne s'arrêtera qu'à l'enlèvement du sujet qui transperce pour les uns, par la destruction, l'altération ou l'oubli pour les autres. Pour

cette dernière raison également, nous n'utilisons pas le point de vue du locuteur dans notre travail de modélisation.

Des traductions leur sont également communes, à commencer par l'emploi quasi-systématique de la préposition *through* qui dénote le passage au travers d'un espace ou d'un objet « visiblement » tridimensionnel lorsqu'il est transpercé. Les verbes de mouvement, supports de la préposition, peuvent également être les mêmes, la plupart du temps génériques et peu motivés sémantiquement comme *to get*, *to go*, *to come* ou *to pass*.

- Est-il possible de « passer par la fenêtre du Temps » ?

Nos différentes approches lexicales et formelles prouvent que les emplois temporels sont beaucoup plus proches des (trans)percées que des voyages et parcours. Aussi, nous pensons que le temps ne doit pas être vu comme une ligne - ou un axe simple dans un repère, comme nous le représentons par commodité - ni comme un objet mobile même si la conceptualisation de « l'avancée du temps » nous fait parfois percevoir le contraire, ni comme une rivière impétueuse, ou pas, (comme présenté dans Borillo 1996) parce que cette dernière vision impliquerait que le temps se déplace indépendamment de l'espace et que chaque morceau du monde spatio-temporel ne se déplace pas seul ou marche le long de la rivière.

D'après nos modèles et notre étude contrastive, nous faisons l'hypothèse que le temps peut être perçu comme une fenêtre ou une porte ouverte au-travers de laquelle le sujet va d'un présent... qui est dans le passé jusqu'à un présent... dans le présent. De ce point de vue, les sens temporels sont de vrais emplois dynamiques au sens aspectuel, bien que cela ne résolve pas le problème de la momentanéité.

Tout ceci pourrait expliquer pourquoi les traductions en anglais pour les percées statiques tridimensionnelles et pour les périodes temporelles dénotées par *traverser* sont presque toutes identiques.

Nous devons néanmoins distinguer les périodes temporelles des événements. Avec *Jean a traversé la crise*, il est tout à fait possible de traduire par « *John got **through** a crisis* » ou bien « *John got **over** the crisis* » ou de façon encore plus précise sémantiquement, « *John got **through** a crisis and then get **over** it* »¹⁹¹. Cet exemple ne contredit pas véritablement notre hypothèse, puisque nous admettons que “*getting over an event*” signifie “*successfully and completely getting **through** this event, without damage, etc.*”¹⁹²

La différence est moins explicite en français, les deux occurrences *to get through a crisis* et *to get over a crisis* pouvant être traduites par *traverser la crise* sans distinction co-textuelle.

¹⁹¹ Traduisible par « Jean a traversé la crise et l'a surmontée ».

¹⁹² Ou *getting over an event* correspond en français à la glose « traverser un événement complètement et avec succès, sans dommage, etc. »

Entre les deux traductions, la nuance est à rechercher dans le dépassement de l'événement traversé : s'il s'agit d'une période vécue douloureusement par le sujet traversant, *over* pourra être employé pour indiquer que ce dernier a non seulement survécu à l'événement mais aussi « fait le deuil » de la douleur induite par l'objet, et seulement dans ce cas.

Ainsi, il y a deux conceptualisations distinctes pour les emplois temporels de *traverser*, l'un avec le temps vu comme surface mince et homogène, proche des emplois statifs, et un pour des événements vus comme des volumes à franchir ou à dépasser, plus près des accomplissements voire des achevements.

4.5.4 Les emplois métaphoriques

Nous considérons comme emplois métaphoriques de *traverser* toutes les occurrences qui ne concernent pas un parcours dans un espace identifiable ni les traversées temporelles, qu'elles soient effectuées dans un intervalle défini ou au travers d'un événement.

Pour déterminer ce qui relève de la métaphore lexicale et ce qui n'en est pas, nous acceptons le point de vue suivant :

« Dire que tout langage est métaphorique (...), c'est dire que nombre de mots conçus aujourd'hui comme non-métaphoriques, sont en fait des métaphores éteintes. Quelle qu'ait été l'origine du langage, il reste que dans son emploi, à un moment donné, certaines expressions sont conçues comme figurées, d'autres non. La question de la métaphore doit trouver d'abord une solution synchronique. »

Todorov, 1967

Outre que Todorov suggère que la désignation de la métaphore est de la responsabilité du linguiste synchronicien, nous entrevoyons dans cette citation un indice de l'évolution naturelle du langage qui transforme le « potentiel sémantique » de chaque terme d'une langue en fonction des besoins de ses locuteurs ou du hasard de la créativité. De semblables considérations ont abouti à une étude approfondie au sujet du verbe *aborder* au chapitre précédent.

Comme confirmation de ce que venons de dire, l'observation du corpus textuel FRANTEXT de 1960 à nos jours fait apparaître de grandes disparités dans la fréquence d'apparition et dans la manipulation des acceptions que nous nommons métaphoriques. Pour ce qui concerne *traverser*, les emplois métaphoriques sont tous ceux qui ne permettent pas une visualisation de la *traversée* autrement que par l'imagination et l'abstraction.

Nous avons remarqué que plus les textes de référence nous sont contemporains, plus le nombre de métaphores de la traversée est élevé. D'autre part, nous notons

également une extension de la sphère d'influence de l'unité verbale, laquelle n'est plus cantonnée à l'emploi figuré d'origine : « une idée qui traverse l'esprit ». En ne considérant que les occurrences de 1985 à nos jours, certains emplois sont ininterprétables par la lexicographie et du coup intraduisibles sans faire intervenir l'arbitraire du lecteur.

Tant que les emplois métaphoriques se limitaient à la traversée d'une idée dans un esprit, il était facile de circonscrire l'analyse à un paradigme restreint pour les arguments sujet et objet. Le premier contenait *idée, mot, pensée* et leurs synonymes ; le second était constitué du champ lexical du cerveau en tant que siège de la pensée, *esprit, âme*, etc.

Plus les œuvres consultées sont récentes, plus les métaphores sont variées et difficiles à décrire. D'une part, le paradigme des sujets potentiels augmente considérablement, en partie grâce au phénomène d'ellipse : la métaphore se lexicalisant de plus en plus, il est davantage acceptable de remplacer le syntagme « une idée » par le contenu de l'idée. On peut noter aussi que les synonymes ou mots apparentés à *idée* sont également plus nombreux au fur et à mesure du temps. Enfin, les sens métaphoriques se développent, s'enrichissent et induisent de nouvelles acceptions qui ne sont pas lexicalisées, elles. C'est le cas de ce que nous avons appelé précédemment les « parcours initiatiques » pour lesquels une interprétation immédiate n'est pas impossible mais demeure un peu hasardeuse sinon un peu floue.

L'idée d'une petite voiture me traversa également **l'esprit** à cent à l'heure, avec moi au volant.

Gary (Ajar Émile) Romain /la promesse de l'aube/1960

Cette tordue commence à siphonner ! C'est ce qui en éclair traverse **l'esprit** d'Armand.

Simonin Albert /Du mouron pour les petits oiseaux/1960

C'est à ce moment qu'une **idée lancinante et terrible** lui traverse **l'esprit**. ***Peut-être n'a-t-il trompé personne, ni les geôliers, ni les codétenus, ni surtout lui-même.***

Fanon Frantz /Les Damnés de la terre/1961

Une inquiétude lui traversa pourtant **l'esprit**. - ***Au fond, je décide sans rien te demander, dit-il, mais cela pourrait ne pas t'arranger du tout que je rentre à Paris en même temps que toi.***

Droit Michel /Le Retour/1964

Jusqu'au milieu des années 1960, « l'idée » est le sujet de loin le plus répandu pour les emplois métaphoriques. Ce n'est réellement qu'après que l'on voit apparaître des termes du même champ lexical, ou littéralement le contenu de l'idée.

Je m' étais fabriqué un tas d' *excuses idiotes* et c' est *la plus idiote* qui m' a traversé *l' esprit*: ne me faites pas de mal, je ne suis pas vraiment mauvaise, je voulais voir la mer.

Japrisot Sébastien /La Dame dans l'auto/1966

Que de fois *son cri terrible* me traverse *l' esprit*: "Je ne suis pas votre mère; une mère ce n' est pas ça, ce que je suis!" Ainsi prévenait-elle ses pauvres enfants dans un mouvement éperdu de sincérité.

Monési Irène /Nature morte devant la fenêtre/1966

Deux minutes avant d' entrer, *la personne de Georges* m' a traversé *l' esprit* et je me suis dit: "Surtout ne pas oublier de parler à Marie de la faillite de cet imbécile;

Dutourd Jean /Pluche ou l'amour de l'art/1967

La mala noche... *Le mot* me traversa *l'esprit* et y fit tout à coup un sillage éveillé. Dans la pénombre vacillante des bougies, les images y glissaient sans résistance (...)

Gracq Julien /La Presqu'île/1970

Un souvenir venait de traverser *l'esprit* de la mémé et elle le jetait tout cru. - Mémé, c'est quoi, là-bas? Tout se nommait: les hameaux, les moulins, les fermes, les chemins (...)

Sabatier Robert /Les Noisettes sauvages/1974

Cette conversation et surtout *le trait de lumière* qui traversa *l'esprit* d'Adèle lorsque retentirent les trois coups de sonnette de Bob la soulagèrent définitivement : aucun de ses frères n'aurait aimé une telle femme !

Delay Florence /Le Aïe Aïe de la corne de brume/1975

Après avoir avalé son verre de vin blanc, Edmond eut *l'esprit* traversé par *une subite lueur*. - Dieu existe parce qu'il est ! proclama-t-il avec autorité.

L'Hète Jean /Le Mécréant ou les preuves de l'existence de Dieu/1981

4.5.4.1 Quels critères appliquer pour les métaphores ?

Les emplois métaphoriques habituels de *traverser* sont d'abord caractérisés par les arguments sujets et objets. Les premiers sont à rechercher dans le champ lexical de l'idée, ou de façon plus générale tout ce qui correspond à une image mentale (pensée, souvenir, etc.) ou à leur contenu. Les seconds sont strictement restreints à des syntagmes nominaux dont la tête est *esprit* dans presque tous les cas, ce qui en facilite le repérage.

Ensuite, nous remarquons que les emplois métaphoriques sont des emplois semi-réfléchis impliquant la présence d'un pronom clitique indirect correspondant à la personne dont l'esprit est traversé, ou par l'emploi du déterminant possessif *son*.

La constance de ces critères syntaxiques et sémantiques explique que nous n'avons pas besoin de faire une analyse aspectuelle détaillée des emplois métaphoriques de *traverser*, à tout le moins pour en caractériser le sens. Dans le dernier exemple tiré de Frantext que nous citons, l'adjectif *obsédante* et surtout l'adverbe *périodiquement* montrent que nous avons affaire à un aspect itératif, mais contrairement aux emplois spatio-temporels du verbe cela n'a pas d'influence sur le choix des lexèmes de traduction. En revanche, cette itérativité peut se retrouver dans la sélection des temps verbaux.

Pour les autres emplois métaphoriques qui ne font pas intervenir l'idée comme sujet, l'esprit comme objet ou leurs équivalents synonymiques, il est déjà plus difficile d'en caractériser le sens. Comme nous l'avons déjà montré, la modélisation des sens métaphoriques est rendue impossible par leur caractère abstrait ; d'autre part, le potentiel métaphorique de *traverser* dépasse le cadre des acceptions connues pour le verbe en lexicographie. C'est pourquoi nous devons nous contenter d'essayer d'interpréter ces occurrences particulières au cas par cas, en proposant éventuellement la prise en compte de nouvelles acceptions ou en effectuant le rattachement de certains emplois à des acceptions déjà connues, par exemple le rapprochement entre les « traversées initiatiques » et les parcours temporels de type 2, i.e. celles qui correspondent à la traversée d'une période bien déterminée.

4.6 Conclusion de la traversée

En premier lieu, comme pour tous les autres verbes, le sens du prédicat contenant *traverser* est déterminé par la nature sémantique de ses arguments. En raison de la polysémie du verbe (quatre sens majeurs, eux-mêmes séparables en plusieurs sous-sens) et de la faible lexicalisation du type de traversée effectuée, ce sont d'abord le sujet et l'objet qui nous permettent de déterminer de quel type de traversée il s'agit. Toutefois, cette reconnaissance des arguments ne donne pas pour autant d'information sur le contenu spatial et temporel des traversées effectuées, ce qui nous amène au point suivant.

Comme nous l'avons montré, la recherche des bornes de l'action et de l'événement donne des informations essentielles quant au contenu aspectuel des traversées. Or, ces bornes peuvent parfois être caractérisées par le co-texte large puisqu'il arrive qu'on ne puisse parler d'accomplissement qu'en faisant référence aux propositions voire aux phrases qui suivent l'occurrence de *traverser*.

Les syntagmes précisant le type de mouvement effectué (cf. « en zigzaguant ») complètent également le sens du prédicat du point de vue du mouvement, et influent directement sur la traduction. En anglais, le choix du prédicat verbal pourra être plus volontiers effectué en fonction du circonstant de mouvement, et s'accompagner de la préposition idoine (*to zigzag through* ou *across* (voire *over*) dans le cas que nous citons).

Les modèles que nous avons conçus pour ce chapitre montrent l'importance de ces notions de bornes et de mouvement pour la représentation du verbe *traverser* en contexte.

La plupart des facteurs que nous avons dégagés pour la caractérisation de l'aspect véhiculé par *traverser* sont difficiles à répertorier dans un dictionnaire. Le contenu sémantique et temporel d'un verbe en contexte peut être soumis à l'influence de ses arguments, d'adverbes et de circonstanciels de toutes sortes qui éclairent sur la configuration du temps et de l'espace, de phrases antérieures ou postérieures précisant les bornes de l'action, et de tous les éléments de langue qui apportent des précisions sur l'ensemble du prédicat.

N'oublions pas que le premier discriminateur pour l'aspect est le sens du verbe indépendamment de tous les ajouts qui en précisent la portée ni que l'interprétation que nous faisons des occurrences littéraires peut être remise en question au nom du décalage entre les intentions de l'encodeur et celles du récepteur.

Comme conclusion préliminaire, nous avons voulu démontrer que les structures aspectuelles sont la plupart du temps importantes pour sélectionner la traduction appropriée d'un prédicat, que ces structures soient portées par le prédicat lui-même ou par le co-texte large (sans parler du contexte non discursif). Tout au long de ce chapitre, nous avons tenté de mettre en avant plusieurs phénomènes. Tout d'abord, le n-dimensionnalisme est un facteur très important dans la conceptualisation du verbe : L'espace et le temps ne peuvent pas être séparés pour décrire et modéliser sa structure aspectuelle.

Ensuite, un verbe comme *traverser* a potentiellement beaucoup de structures aspectuelles qui ne sont pas incluses dans le lexème, mais qui sont seulement saturées par le contexte et le co-texte, comme beaucoup de linguistes l'admettent désormais.

Enfin, notre étude contrastive prouve que l'anglais a en apparence plus d'outils directement lexicaux que le français pour distinguer ces structures. En français, seul le contexte permet d'établir la valeur aspectuelle pour *traverser indubitablement* et fixer ses bornes.

Pour terminer, nous espérons avec raison que nos modèles seront utilisables pour décrire d'autres verbes liés à *traverser*, et peut-être tous les prédicats impliquant l'espace et le temps.

La prise en compte des nouvelles données est plus délicate au sujet du verbe *traverser*. Certes, nous réussissons à déterminer de façon quasi-exhaustive quels sont les éléments du discours qui participent à la caractérisation de l'aspect de ce verbe en contexte. D'autre part, l'aspect véhiculé par les traductions anglaises est également repérable dans la majorité des cas. Cela étant, ces nouveaux éléments demeurent difficiles à répertorier de façon ordonnée et, surtout, ne témoignent pas d'un comportement stable d'une occurrence à l'autre.

Dans ce cas, le problème peut être pris à l'envers : si l'environnement langagier des occurrences de *traverser* ne permet pas un repérage systématique de l'aspect en fonction de leur forme, autant faire porter ces informations par les traductions elles-mêmes. Reprenons notre exemple de la « traversée de la forêt à pied » : trois traductions principales se détachent, *walk across*, *walk through* et *walk over*. Le contexte de *traverser* en français est trop instable pour cibler la vedette la plus adaptée, et le verbe ne porte pas en soi de trait aspectuel activé ; de leur côté, les lexèmes verbaux anglais sont implantés avec leurs caractéristiques particulières (*infoscompls*). Pour notre travail de modélisation, c'est en faisant supporter par l'entrée de la traduction les traits sémantiques et aspectuels particuliers que nous avons le plus de chance d'avoir des résultats cohérents lors du transfert interlangue.

A cela nous ajouterons que les données pragmatiques de type *infoscompls* ne sont pas des fragments de langage littéral, ni des types prédéfinis. Leur présence doit permettre à un utilisateur de déterminer son choix en fonction de l'environnement textuel du verbe qu'il cherche à traduire.

Figure 30 : Fragment de CSMU pour *traverser* en contexte

mot-source	mot-cible	syn LS	syn LC	colloc objet	colloc sujet	constr LS	constr LC	registre	infos complémentaires
Ftraverser	Awalk across			territoire					Objet uni- ou bidimensionnel – marche
Ftraverser	Awalk through			territoire					Objet tridimensionnel – marche
Ftraverser	Awalk over			territoire					Aspect d'achèvement – marche
Ftraverser	Arun across			territoire					Objet uni- ou bidimensionnel – course
Ftraverser	Arun through			territoire					Objet tridimensionnel – course

Il est probable que la principale lacune de notre étude est que les emplois du verbe *traverser* ont été analysés sans aucune référence aux autres verbes français qui peuvent se substituer à lui, i.e. ses synonymes et parasynonymes en langue source. C'est pourquoi une autre étude serait utile pour montrer toute l'influence des relations lexicales à l'intérieur d'un champ bien défini. Pour présenter ce phénomène, entre autres considérations, nous avons décrit au chapitre précédent une autre collaboration de recherche impliquant cette fois plusieurs membres de l'ILF dans le cadre d'une réflexion globale sur la transitivité : les verbes de parole et leur comportement en contexte dans les textes littéraires et leur rapport, ou leur apport, à la grammaire traditionnelle et aux dictionnaires.

Pour terminer, revenons aux deux hypothèses formulées dans l'introduction au sujet de ce chapitre et de la fin du précédent :

Première HYPOTHÈSE : Ces analyses produisent des relations lexicales équivalentes et/ou complémentaires aux dictionnaires.

Seconde HYPOTHÈSE : Ce qu'on retrouve comme constantes syntaxico-sémantiques dans une langue se retrouve dans les autres langues de manière équivalente.

Le résultat pour la première hypothèse est que les corpus littéraires font apparaître des compléments en grand nombre aux données lexicographiques, tant syntaxiquement que sémantiquement. La fantaisie des constructions recensées et des acceptions nouvelles est finalement à la mesure de la créativité des auteurs romanesques. Par ailleurs, si beaucoup des occurrences étudiées respectent le canevas d'acceptions observables dans les dictionnaires, nous notons qu'il existe des différences subtiles entre chaque emploi, lesquelles portent aussi bien sur l'aspect du prédicat (comme pour *traverser*) que sur le contexte extralinguistique, sur lequel nous avons hélas peu de prise faute de pouvoir nous mettre à la place des écrivains.

Pour la seconde hypothèse nous répondons : manifestement oui, la créativité lexicale n'est pas l'exclusivité du français, et il est possible de rendre transitif un verbe intransitif d'une autre langue, s'il traduit une forme transitive du verbe français intransitif auquel il est associé habituellement. Toutefois, la fantaisie que l'on observe pour certains verbes conduit celui-ci à changer de sens par rapport aux données lexicographiques et de fait à lui accoler éventuellement d'autres traductions. Ce dernier point nous pose réellement problème car nous n'avons jamais prétendu réécrire les dictionnaires bilingues. Néanmoins, nos modèles autorisent dans un premier temps le recensement de données « isolées », c'est-à-dire ne proposant pas de candidats mais faisant juste apparaître les critères qui permettraient de choisir ceux-ci.

À l'arrivée, le résultat attendu est que l'on peut construire un graphe enrichi des informations syntaxiques et contextuelles en procédant finalement de manière identique à ce qui est décrit au chapitre II. Le résultat constaté est l'apparition d'un graphe encore plus riche que le précédent pour les verbes fortement représentés (voir résultats des hypothèses 1 et 2) et une disparition forcée des verbes sous-représentés ou pas représentés du tout, lesquels ne sont finalisés dans nos bases qu'à partir des acceptions, constructions, traductions et co-textes lisibles dans les dictionnaires monolingues et bilingues. À propos de ce dernier point, nous noterons que c'est ce qui fait une des grandes originalités du TLF, lequel présente en entrée des mots qu'on ne retrouve dans aucun autre ouvrage lexicographique mais en ignore certains, pourtant apparents dans des dictionnaires plus grand public, parce qu'ils n'ont pas été utilisés par les auteurs des textes saisis dans la base littéraire FRANTEXT qui a servi à nourrir l'essentiel du TLF.

Au final, le travail sur corpus n'a pas conduit à modifier les bases de données que nous avons constituées préalablement, au sens où il n'a pas été nécessaire d'en changer la structure pour l'enrichir. Les dictionnaires étant eux-mêmes des bases issues du dépouillement de corpus, au moins pour partie, nous pouvons considérer que les données initiales peuvent être complétées grâce au corpus. Cependant, nous ne nous permettrons pas de modifier les données du dictionnaire qu'un corpus trop restrictif n'a pas permis de répertorier. De ce fait, l'analyse faite à partir de FRANTEXT ne peut servir qu'à enrichir la base.

Dans l'optique de construire un graphe multilingue à partir d'une base de données malléable, il n'y a aucune nécessité à entrer des données exactes et finies dès la première écriture de cette dernière. Des ajouts successifs et des corrections sont implantables à tout moment, ce qui contredit la critique éventuelle selon laquelle « le graphe est faux si le dictionnaire est faux » car rien n'oblige à conserver des données fausses si elles sont reconnues comme telles. Dans le chapitre suivant, nous allons présenter en détail le mode de fonctionnement de ces bases de données ainsi que l'algorithme que nous leur associons pour permettre l'implantation des CSMU.

CHAPITRE IV
Modélisation et algorithme de construction du graphe des Champs
Sémantiques Multilingues Unifiés ou CSMU

1. Objectifs de la modélisation et Exploitation des données

1.1 Problématique originelle

Parvenu à ce stade de l'étude sur les verbes, notre objectif principal est désormais la constitution de graphes lexicaux multilingues illustrant la non biunivocité des correspondances lexicales entre les langues, et permettant de la dépasser. En effet, nous avons déjà insisté sur le fait qu'il est rare qu'à un terme en langue source corresponde une seule traduction en langue cible. Par surcroît, le décalage inter-langue grandit de manière exponentielle à chaque ajout d'une langue dans l'étude.

Pour toutes ces raisons, nous avons besoin d'un mode de représentation et d'exploitation des données souple, modulable et fidèlement représentatif de la réalité du lexique. Notre choix s'est porté sur la constitution de champs lexico-sémantiques implantable à partir d'un algorithme de construction de graphe et apparaissant sous la forme d'une base de données contrôlée par un faisceau de critères distinctifs. Ce sont les Champs Sémantiques Multilingues Unifiés (CSMU) dont nous avons fait une première et brève présentation au chapitre II. De par leur nature automatisée, les CSMU sont des graphes de relations lexicales à lecture multiple : les CSMU pouvant être visualisés dans leur totalité ou bien partiellement par l'appel d'une lexie qui y est présente. Dans l'idéal, ils pourront également être interrogés sur des critères restrictifs comme le choix des langues visées, les registres d'emploi, etc.

Dans la genèse de nos travaux, nous évoquons généralement le terme-source en dehors de toute comparaison avec les unités lexicales de la langue à laquelle il appartient. Or, comme le montrent nos différentes approches linguistiques, il n'est pas envisageable de faire une étude lexicale multilingue cohérente sans étudier les champs sémantiques de chaque langue de l'étude dans leur ensemble. La multiplicité des liens synonymiques et parasyndonymiques des unités d'une même langue entre elles ainsi que des liens de traduction d'une langue à une autre est à l'origine de notre volonté de représenter toutes les informations au sein d'un même espace. Les niveaux de synonymie décrits plus loin sont destinés à distinguer les liens de synonymie décrits dans des ouvrages lexicographiques monolingues des liens plus complexes établis à partir des dictionnaires bilingues. Au final, il s'agit toujours de synonymie, qu'elle soit intra-langue ou inter-langue¹⁹³, cette dernière trouvant une illustration observable dans le langage des enfants bilingues, lesquels sont susceptibles d'insérer des termes de « l'autre langue » qu'ils maîtrisent dans un énoncé, s'ils les jugent plus appropriés dans un contexte précis¹⁹⁴.

¹⁹³ Au nom de la synonymie que nous redéfinissons à notre façon au Chapitre I, afin de permettre la désignation de liens de synonymie entre termes de langues différentes.

¹⁹⁴ Processus de « code-switching ».

1.2 Problématique enrichie

Grâce aux études linguistiques sur corpus, nous avons montré l'intérêt de compléter les données lexicographiques par une étude approfondie des emplois « inhabituels » de la plupart des verbes polysémiques.

L'étude approfondie du sémantisme de *traverser* et de son transfert en langue cible nous a conduit à une recherche de représentation plus formelle de ses traits aspectuels sous la forme de modèles mathématiques idoines.

Les modèles que nous proposons s'inscrivent dans une vision quadridimensionnelle qui réconcilie Temps et Espace dans le même Univers (cf. Einstein 1994 trad., Quine 1960, Denis & Muller 2004, Haton & Mascherin 2005). Dans une perspective beaucoup plus large, ils doivent permettre la description sémantique et aspectuelle de l'ensemble des unités verbales dont l'interprétation suppose l'existence d'une situation d'énonciation basée sur le temps et l'espace, ce qui concerne la grande majorité des prédicats.

Il n'y a pas de séparation de facto entre temps et espace d'où on peut tout représenter par l'intermédiaire d'un repère orthonormé avec des axes spatio-temporels différents. Ce repère ne donne en apparence qu'une seule dimension spatiale, il en manque donc deux. Souvenons-nous que l'objet peut être unidimensionnel (une ligne renouvelée dans le Temps), bidimensionnel (un plan « à l'égyptienne » renouvelé dans le temps) ou tridimensionnel (un volume « à l'égyptienne » renouvelé dans le temps). En revanche, nous rappelons que le sujet qui traverse peut être n-dimensionnel sans incidence sur le sens du procès, c'est-à-dire que ses dimensions ne sont pas pertinentes dans cette représentation.

L'objet traversé existe sur une période temporelle donnée, c'est donc par rapport au temps que nous pouvons représenter les dimensions spatiales de *traverser*. Par exemple, la route (objet traversé ou site) existe à partir du moment où elle a été construite jusqu'au moment où elle est détruite, entre temps il y a l'usure.

Sur l'axe temporel, nous avons repris la perspective de Reichenbach qui introduit 3 points de repérage E, R et S, perspective selon laquelle l'aspect de la prédication est représenté par la structure interne de l'intervalle E. Vet et Gosselin préfèrent parler d'intervalle, en effet, l'aspect lexical (type de situation), n'est pas un point mais un intervalle de temps comme nous avons pu le faire remarquer, compris entre E_1 (début de l'événement) et E_2 (fin de l'événement).

Le lien entre cette modélisation et les CSMU réside dans notre souhait d'intégrer à terme l'ensemble des phénomènes textuels qui ont une incidence sur un parcours de traduction. Compte tenu de la trop faible avancée de nos recherches à propos de la transformation de nos modèles en données directement intégrables dans la base, nous n'en parlerons presque pas dans ce chapitre. Soulignons que ce dernier point constitue notre principal objectif de recherche pour l'année à venir si nous

parvenons à (faire) programmer la transformation de base de données en graphe multilingue¹⁹⁵.

1.3 Méthodologie et Travaux Préliminaires

1.3.1 Vers les CSMU

Pour représenter nos données, nous proposons la création de champs sémantiques multilingues unifiés, dirigés par un réseau qui fait apparaître les liens sémantiques entre les lexies d'une même langue autant qu'entre celles de langues différentes. Chaque unité du réseau est porteuse de ses propres informations linguistiques, supposées stables tous emplois confondus, tandis que les liens font apparaître les données qu'elle partage - ou celles qui en diffèrent - avec les autres unités auxquelles elle est reliée au sein du graphe.

La forme de ces champs permettra de (re)construire des dictionnaires bilingues ou multilingues en utilisant n'importe quelle langue présente dans le graphe comme langue source. De même, toute unité lexicale du graphe pourra être sélectionnée comme lexie centrale de référence pour la visualisation du graphe.

La méthodologie pour les CSMU peut être perçue comme les moyens à mettre en œuvre pour opérer la transformation d'un ou plusieurs dictionnaire(s) bilingues, et éventuellement monolingues moyennant quelques aménagements, en graphe multilingue unique. L'ajout d'une nouvelle langue et de nouvelles données ne doit pas en principe démultiplier le graphe mais en permettre l'insertion le plus naturellement possible. L'un des objectifs finaux demeure de restituer un dictionnaire de langues coopératif pour l'aide à la traduction à partir des CSMU.

Afin de construire nos champs, nous procédons de la façon suivante :

- L'ensemble des informations pertinentes contenues dans un dictionnaire bilingue est organisé dans un tableau dont le fonctionnement a déjà été décrit et sera détaillé plus loin.
- Les données contenues dans ce tableau sont extraites conformément à l'algorithme de construction des CSMU (seconde partie de ce chapitre), chaque ligne créant un ou plusieurs lien(s), chaque colonne correspondant à une étiquette de mot ou à une information linguistique portant sur un mot de la même ligne.

¹⁹⁵ À l'heure où nous imprimons, le programme a été écrit et le passage de la base de données au graphe est acquise.

1.3.2 Préparation d'une entrée

Afin de préparer l'implantation des CSMU à partir d'un dictionnaire de langues, nous allons rappeler l'extrait d'entrée suivant que nous avons détaillé au chapitre II (Robert & Collins 2002) et qui fait apparaître les principaux traits disponibles dans notre ouvrage de référence, en précisant pour chacun de ces traits ce que nous comptons en faire dans les CSMU :

leave¹⁹⁶ /li:v/¹⁹⁷ (vb¹⁹⁸ : prêt, ptp **left**¹⁹⁹)
VT²⁰⁰ a²⁰¹ (= go away from²⁰²) [+ town²⁰³] **quitter**²⁰⁴, partir
de²⁰⁵ ; (*permanently*²⁰⁶) **quitter** ; [+ room, building²⁰⁷]

¹⁹⁶ Entrée lexicale, le mot qui est traduit. Première colonne du tableau intermédiaire.

¹⁹⁷ Forme phonétique de l'entrée lexicale. Critère non utilisé.

¹⁹⁸ Catégorie grammaticale de l'entrée lexicale. Doit apparaître dans un graphe complet mais n'apparaît pas dans nos données car seuls les verbes y figurent, la valeur par défaut étant donc « V ». Toutefois, il peut arriver qu'un verbe soit traduit par une unité lexicale d'une autre catégorie grammaticale, et vice-versa. Dans ce cas, la catégorie « déviante » est mentionnée.

¹⁹⁹ Formes conjuguées particulières de l'entrée, en l'occurrence le prétérit et le participe passé d'un verbe anglais irrégulier. Critère non utilisé pour la construction du graphe mais qu'il serait nécessaire d'implanter en contre-champ pour une éventuelle application des CSMU à des corpus (lemmatiseur).

²⁰⁰ Indicateur de transitivité. Ce trait n'est pas reproduit tel quel dans notre base : la présence de collocations objet y indique que l'on traite une forme transitive (avec une distinction entre la transitivité directe et indirecte, la bitransitivité cumulant les deux, la bitransitivité directe et la bitransitivité indirecte), leur absence dénote les formes intransitives.

²⁰¹ Numérotation propre au dictionnaire pour dégrouper les paradigmes de traduction en fonction des acceptions de l'entrée. Assimilable au dégrouperment polysémique ou homonymique selon le cas. Pas utilisé dans notre base.

²⁰² Mot équivalent en langue-source de l'entrée dans un contexte linguistique donné. Apparaît dans la colonne « synonyme LangueSource » générateur d'un lien de synonymie entre ce terme et l'entrée.

²⁰³ Collocation objet exprimé dans la langue du mot vedette. Apparaît dans une des colonnes « collocations objet » ; il est commun à l'entrée et au mot traduit, bien que seulement exprimé dans la langue source.

²⁰⁴ Traduction de la vedette. Apparaît dans la colonne « mot-cible »

²⁰⁵ Présence d'au moins deux traductions en langue cible de la vedette dans des contextes linguistiques équivalents. Dans ce cas, une ou plusieurs nouvelles lignes (une pour un seul mot nouveau, cinq pour deux mots, etc.) sont créées et les mots-cibles considérés sont placés dans la colonne « synonyme LangueCible ».

²⁰⁶ Information complémentaire d'ordre sémantique portant sur la vedette et ayant une influence sur la traduction (le mot cible étant de son côté porteur de l'information → *leave permanently = quitter*). Présent dans la colonne « infos complémentaires », catégorie éventuelle *partie du discours*. Ce critère peut ne concerner qu'une seule des deux langues comparées, l'autre lexie pouvant contenir implicitement cette information.

sortir de, quitter ;²⁰⁸ [+ *person, job, one's husband, wife*] quitter ; [+ *one's children*] abandonner ♦ **he left Paris in 2001**²⁰⁹ il a quitté Paris en 2001²¹⁰ ♦ **we left Paris at 6 o'clock** nous sommes partis de Paris *or* nous avons quitté Paris à 6 heures...
e (*Math*²¹¹) **three from six ~s**²¹² **three** six moins trois égalent²¹³ trois...
f (*in will*) [+ *money*] laisser {*to à*²¹⁴} ; [+ *object, property*] laisser, léguer {*to à*}²¹⁵ ...
vi (= *go away*) [*person, train, ship etc*]²¹⁶ partir, s'en aller ; (= *resign*) partir, démissionner
♦ **to ~ for Paris** [*person, train*] partir pour²¹⁷

²⁰⁷ Présence d'au moins deux collocations objet distinctes pour un même paradigme de traductions. Ces collocations sont elles-mêmes des classes (et non des fragments de discours littéral) induisant des paradigmes lexicaux sous-jacents. Pour l'instant, il n'y a qu'une ligne par ensemble de collocations dans le tableau.

²⁰⁸ Le point-virgule indique qu'on passe à un autre contexte langagier. Dont acte sans conséquence pour nos bases.

²⁰⁹ Syntagme exemple. Très important car il peut faire apparaître d'autres traductions, d'autres collocations, de nouvelles constructions et d'autres liens de synonymie LangueSource ou LangueCible que ceux proposés précédemment. Leur analyse oblige à un travail de fourmi fastidieux mais nécessaire dans un souci d'exhaustivité.

²¹⁰ Traduction du syntagme exemple. Voir point précédent.

²¹¹ Indicateur de domaine d'emploi : typographie presque identique aux « infos complémentaires » à l'exception de la majuscule initiale. Présent dans la colonne « registre » qui inclut le niveau de langue, le domaine d'emploi et éventuellement la zone géographique.

²¹² Le tilde '~' sert à remplacer la graphie littérale du mot source. Sans effet notable, sauf si nous envisageons de travailler à partir de dictionnaires électroniques balisés.

²¹³ Nouvelle traduction repérée à l'intérieur de la traduction d'un syntagme-exemple. Élément ajouté dans la colonne « mot-cible », informations langagières ajoutées par ailleurs si elles sont repérables (collocs, constr, etc.)

²¹⁴ Constructions syntaxiques particulières associées à l'acception traitée : si le verbe est transitif indirect, on ne fait apparaître que la préposition précédant l'objet, lequel est présent juste avant. L'information portant sur la langue source est en premier et en italique, nous le retrouvons dans notre base dans la colonne « constructionLangueSource » ; la construction pour la langue cible est en second en écriture normale, codé chez nous en colonne « construction LangueCible ».

²¹⁵ La construction en langue cible de cet exemple concerne les deux verbes candidats proposés, ici *laisser* et *léguer*.

²¹⁶ Collocations sujet ou restriction sur la sélection sémantique du sujet pour des emplois donnés. Présent dans la colonne « collocations sujet ». Comme pour les objets, l'information est commune aux deux langues bien qu'il ne soit exprimé qu'en langue source.

²¹⁷ Construction syntaxique particulière non prévue dans le paradigme de traductions principal. A repérer et à ajouter le cas échéant dans les colonnes « construction », si

Paris ; [*ship*] partir *or* appareiller²¹⁸ pour Paris...

► **leave off**²¹⁹ **V**1 (*²²⁰ = *stop*) s'arrêter ♦ **where did**

we ~ off? (*in work, reading*²²¹) où nous sommes-nous arrêtés? ♦ ~ **off !** arrête !, ça suffit ! *

VT SEP a (* = *stop*) arrêter (*doing sth* de faire qch)...

► **leave out** **VT SEP**²²² **a** (= *omit*) (*accidentally*) oublier, omettre ; (*deliberately*) exclure ; [+ *line in text*] (*also Mus*)²²³ [+ *note*] sauter...

l'on considère qu'il s'agit d'une forme syntaxique pertinente (i.e. fréquente et/ou complément, par opposition aux circonstants).

²¹⁸ Nouvelle traduction non présentée d'emblée et mise au même niveau qu'un autre terme-cible dans les syntagmes exemples. Ajouté dans les colonnes « mot-cible » et « synonyme LangueCible », mais sur des lignes différentes, bien entendu.

²¹⁹ Verbe à particule. Pour de nombreux verbes anglais, ces formes sont des lexies à part entière et nous souhaitons les traiter comme telles au sein du graphe. Dans ce cas *leave off* fait l'objet d'un traitement différencié par rapport à *leave* en devenant un « mot-source ».

²²⁰ le * sert à signaler un emploi familier, en l'occurrence celui de *leave off* avec le sens de *s'arrêter*. Plus le nombre d'étoiles est élevé, plus l'expression est considéré comme grossière. Peut faire l'objet d'une colonne à part, non réservée pour l'instant, ou bien être codé dans la colonne *registre*.

²²¹ Information pragmatique sur le contexte langagier, même typographie que les domaines d'emploi ou les informations complémentaires de type adverbiales. On suppose qu'il peut être implicite aussi bien qu'explicite selon les circonstances. Utile lorsque le contexte d'énonciation peut permettre de distinguer deux acceptions. Présent dans la colonne « informations complémentaires ».

²²² SEP indique que le verbe et la préposition sont séparables dans certaines constructions du verbe à particule considéré. Lorsque le verbe et la préposition ne sont pas séparables, le dictionnaire le mentionne par FUS (comme fusionné). Critère non répertorié mais important, à étudier.

²²³ *also* est présent pour préciser que la traduction générique proposée s'applique également à un domaine spécifique, en l'occurrence *la musique*. Dans ce cas, deux lignes de tableau sont nécessaires pour distinguer les collocations alors que la traduction reste la même, car nous ne fusionnons pas d'emblée les données d'emplois différents.

1.3.3 Récapitulation des traits utiles

De tous les éléments présentés ci-dessus, nous ne retenons pour la construction de notre base que ceux de la liste suivante (nous excluons de cette liste les éléments jugés non pertinents pour notre étude (phonétique, etc.) et ceux que nous gardons pour plus tard). En dehors de la catégorie grammaticale, aucun de ces éléments n'est obligatoire, pas même le premier (dans le cas d'une locution intraduisible) :

- la **vedette**, le mot en entrée : *mot source* ;
- une ou plusieurs **traductions** en langue cible pour chaque acception et/ou chaque contexte donné(e(s)) : *mot cible* et *synonymelanguecible* (*synLC*) (pour un nombre de traductèmes ≥ 2) ;
- des **marqueurs d'emploi** pour les vedettes et les candidats (registre de langue, technolecte, idiolecte, régionalisme, zone d'emploi géographique...) : *registre* ;
- les **catégories grammaticales** des vedettes et des traductions (pouvant être différentes) : V (verbe) par défaut. Toutefois, certains adverbess ou noms déverbaux peuvent traduire des verbes dans des contextes particuliers, d'où l'obligation de mentionner la catégorie pour tous les mots ;
- des **synonymes de la vedette** en langue source pour une acception et/ou un contexte donné(s) : *synonymelanguesource* (*synLS*) ;
- les **constructions syntaxiques** associées à la vedette et au(x) candidat(s) : *constrLS* et *constrLC* ;
- des **informations pragmatiques ou des modifieurs** qui affinent le contexte d'emploi : *infos complémentaires* ;
- des **syntagmes exemples** traduits qui illustrent un emploi donné, et qui le cas échéant font apparaître des **traductions** et/ou des **constructions syntaxiques** nouvelles : *mot cible*, *synonymelanguecible*, *constructionLS*, *constructionLC*, *collocations objet*, *collocations sujet* voire *registre* et *infos complémentaires* ; les syntagmes exemples sont une véritable mine d'informations nouvelles ;
- des **collocations sujet ou objet** habituellement associées aux vedettes et candidats présentés : *collocations objet1 et 2*, *collocations sujet*.

1.3.4 Visualisation des critères retenus et de la base d'accueil

Figure 31 : mise en forme du fragment d'entrée "to leave"

mot-source	mot-cible	Syn LS	syn LC	colloc objet	colloc sujet	constr LS	constr LC	registre	infos complémentaires
Aleave	Fabandonner			children					
Aleave	Fappareiller		partir		ship	for obj	pour obj		
Aleave	Fdémissionner	Resign	partir						
Aleave	Fégaler				numbers			math	
Aleave	Flaisser			money		to obj	à obj		in will
Aleave	Flaisser		léguer	object, property		to obj	à obj		in will
Aleave	Fléguer		laisser	object, property		to obj	à obj		in will
Aleave	Fpartir	go away	s'en aller		person, train, ship, etc				
Aleave	Fpartir	resign	démissionner						
Aleave	Fpartir		appareiller		ship	for obj	pour obj		
Aleave	Fpartir				person, train	for obj	pour obj		
Aleave	Fpartir de	go away from	quitter	town					
Aleave	Fquitter	go away from	partir de	town					
Aleave	Fquitter	go away from							permanently
Aleave	Fquitter	go away from	sortir de	room, building					
Aleave	Fquitter			person, job, husband, wife					
Aleave	Fs'en aller	go away	partir		person, train, ship, etc				
Aleave	Fsortir de	go away from	quitter	room, building					
Aleave off	Farrêter		suffire						
Aleave off	Farrêter	stop				vb ing	de vb inf		
Aleave off	Fs'arrêter	stop							
Aleave off	Fsuffire		arrêter		ça				
Aleave off	Fs'arrêter								in work, reading
Aleave out	Fexclure	omit							deliberately
Aleave out	Fomettre	omit	oublier						accidentally
Aleave out	Foublier	omit	omettre						accidentally
Aleave out	Fsauter			line in text					
Aleave out	Fsauter			note				mus	
Aleave out	Ftenir à l'écart		laisser à l'écart	sb					

Chaque ligne du tableau représente les informations liées à une et une seule traduction d'une unité donnée, i.e. l'intersection d'un contexte d'emploi et d'un seul candidat, celui-ci étant éventuellement accompagné d'une autre traduction potentielle

à contexte équivalent. Dans ce dernier cas cependant, une autre ligne sera nécessaire pour mettre en avant le second candidat.

Certaines colonnes nouvellement créées ne figurent pas encore dans cette représentation : la *langue*, désormais séparée des lexies dans l'optique d'un traitement automatique, la séparation entre *collocations objet1* pour les emplois transitifs directs et *collocations objet2* pour les indirects, la colonne *argumentcompl* qui doit rendre compte du nombre d'arguments non sujets dans l'emploi visé. Cette dernière est prévue sous réserve de faisabilité puisque la détermination du degré de transitivité²²⁴ de chaque lexie est une tâche ardue. Ceci n'exclut pas la possibilité de l'ajouter ultérieurement (même de manière parcellaire au cas par cas) car nous proposons des modèles évolutifs susceptibles d'être paramétrés au-delà de leurs possibilités initiales.

Les quatre premières colonnes de la base de données correspondent à des unités lexicales visualisables dans les CSMU et ne faisant pas partie des structures de traits (cf. algorithme) :

- 1 Mot-source : c'est le terme-noyau, celui qui est traduit
- 2 Mot-cible : traduction directe du mot-source
- 3 Synonyme LS : terme présenté dans le dictionnaire comme synonyme direct du mot-source vis-à-vis du mot-cible
- 4 Synonyme LC : terme proposé comme traduction directe du mot-source au même niveau que le mot-cible de la même ligne. Le synonyme LC devient mot-cible à son tour dans la base avec le même paradigme d'informations extralexicales.

Les six dernières colonnes contiennent les informations argumentales, syntaxiques et pragmatiques qui sont exprimées dans le dictionnaire et qui renseignent les liens lexicaux précédemment décrits, futurs constituants des structures de traits susmentionnées :

- 5 Collocation objet
- 6 Collocation sujet
- 7 Construction LS
- 8 Construction LC

²²⁴ Autrement dit de la valence.

- 9 Registre
- 10 Informations complémentaires

Il conviendra d'ajouter à ces six catégories informationnelles celles que nous évoquons plus haut (langue, bitransitivité). D'autre part, nous craignons que la constitution des deux dernières colonnes en « catégories fourre-tout » ne soit sans doute pas satisfaisante dans une analyse complète et sérieuse des phénomènes de synonymie interlangue.

1.4 Construction des bases de données

1.4.1 Méthodologie d'extraction des données

Pour constituer nos paradigmes de traduction, nous employons une méthode d'extraction lexicographique en trois étapes, au moins :

- L'extraction directe, aussi appelée « traduction droite », qui récupère toutes les traductions du mot-source à l'entrée de celui-ci dans le dictionnaire bilingue.
- L'extraction « cachée », ou « traduction inversée », qui ajoute à la précédente toutes les unités lexicales dont le mot étudié est une traduction proposée mais qui n'apparaît pas dans les traductions droites (cf. Ploux, 1997 pour une semblable méthodologie à partir de dictionnaires de synonymes). Les cas de redondance avec les résultats de la première extraction ne constituent pas un problème mais plutôt un complément d'informations. Pour les bases des CSMU comme pour les Lexiques Inversés, c'est le même principe du miroir qui s'applique.
- L'extraction « indirecte » qui nécessite au moins deux étapes d'investigation. Dans notre terminologie, une traduction indirecte d'un mot-source est soit la traduction directe ou cachée d'un de ses synonymes directs, soit un synonyme direct d'une de ses traductions directes ou cachées.

Les deux premières sont manuelles et respectent littéralement la cohérence de l'ouvrage lexicographique choisi, nonobstant les corrections manuelles que nous pouvons être amenés à effectuer en cas d'erreur d'édition. La troisième est souhaitée comme l'aboutissement ou le prolongement du travail de modélisation que nous avons effectué.

1.4.2 Première étape de formalisation

Pour représenter les résultats de l'extraction de façon plus formelle, voici comment nous procédons :

- Tout verbe **Vi** de la base a un nombre **n** de traductions directes.
- Chacune de ces traductions (**V'1... V'j**) a un nombre **n'** de traductions directes dans la langue de **Vi** (ensemble dont celui-ci peut être exclu, cf. les « lexiques inversés » et leur dissymétrie²²⁵).
- Chacune des traductions des traductions de **Vi** a un nombre **n''** de traductions...
- Ainsi de suite. Pour des raisons évidentes de non-redondance des informations saisies, on ne « balaie » pas deux fois à partir de la même lexie.

Par cette méthode, on construit un graphe lexico-sémantique multilingue morceau par morceau. Chaque lien du graphe est un lien objectif, i.e. purement lexicographique.

1.4.3 À propos de la sphère d'influence et des étiquettes des lexies

Le balayage n'a pas de fin, les limites de la « sphère d'influence » du verbe-noyau sont rapidement atteintes puis franchies ; sans restriction, un seul parcours par balayage peut couvrir l'ensemble des verbes des deux langues comparées. Si le champ synonymique du verbe-noyau est contraint préalablement au balayage, l'écueil peut être surmonté (d'où le choix du paradigme de synonymes du noyau). Dans ce cadre, n'importe quel lexème peut devenir le verbe-noyau, selon l'angle de vue du concepteur ou de l'observateur. Néanmoins, le changement du verbe de référence modifie la forme du sous-graphe considéré, car chaque verbe a son propre paradigme de synonymes dans sa langue, très probablement différent de celui du verbe-noyau précédent.

Il faut ajouter que les informations concernant les registres d'emploi, les constructions syntaxiques et les collocations sont associées aux **liens** entre deux lexies

²²⁵ Par exemple, *abandonner* est une traduction directe de *to jettison* mais ce dernier n'est pas une traduction directe de *abandonner*.

et non pas aux **étiquettes lexicales**, étant entendu que ces données diffèrent souvent d'un contexte à l'autre pour un lexème²²⁶ donné.

2. Sur la notion de synonymie

2.1 Préliminaires

Dans la mesure où nous prétendons construire un « graphe de synonymes multilingue », il est nécessaire de revenir sur la notion de synonymie et sur le sens que nous entendons lui donner.

La définition habituelle de la synonymie est : « relation entre deux mots ou deux expressions qui ont des significations très voisines ou le même sens ». Il s'agit d'une relation mathématiquement non orientée car la synonymie marche toujours dans les deux sens, i.e. si le terme *a* est synonyme du terme *b* alors le terme *b* est synonyme du terme *a* :

$$[(aSb \Rightarrow bSa) \wedge (bSa \Rightarrow aSb)]^{227}$$

Pour notre vision de la synonymie, il est nécessaire de prendre en considération deux paramètres :

- la polysémie quasi systématique de la plupart des mots et expressions, ce qui implique que la synonymie ne s'applique pas véritablement (ou très rarement) à des unités lexicales mais plutôt à des entrées lexicales (sous forme de lexies) de chacune d'entre elles ;
- l'influence du contexte langagier. La synonymie que nous envisageons est une synonymie en contexte, une même entrée lexicale en langue source pouvant être traduite de différentes manières selon le contexte d'énonciation. Nous avons remarqué la prédominance du contexte et surtout de sa partie co-textuelle dans les occurrences que nous avons analysées à partir de FRANTEXT.

Les liens de synonymie issus des dictionnaires de langues sont orientés *de facto* et nous ne masquerons pas d'emblée les orientations car les liens orientés nous donnent des informations importantes sur la façon dont les dictionnaires sont conçus

²²⁶ Notons que nous employons toujours « lexie » dans le sens d'un emploi donné d'un mot précisé par Mel'çuk à la suite de Pottier, et que nous utilisons « terme » ou « entrée lexicale » pour désigner un verbe (ou un mot de façon générale) tous emplois confondus.

²²⁷ Où *a* et *b* sont des variables et *S* est le lien de synonymie formalisé. Relation d'implication entre deux synonymes.

ainsi que sur la dissymétrie de traitement et de fonctionnement de la traduction, voire sur la structure conceptuelle même des langues mises en relation. Ce parti pris se retrouve dans le graphe final, bien qu'il soit très facile de transformer un graphe orienté en un graphe non orienté, en modifiant simplement l'algorithme de mise en relation des lexies.

En outre, nous considérons également la synonymie comme un phénomène interlangue en partant du principe que la traduction respecte le principe de base de la synonymie, à savoir que le traductème et le terme en langue-source « ont des significations très voisines ou le même sens ». Pour aller plus loin dans notre propos, il paraît évident que le lien de synonymie est plus fort entre la lexie et ses traductèmes qu'entre deux termes de même langue, selon le principe d'économie du langage qui engendre peu de doublons systématiquement substituables l'un à l'autre tous contextes confondus dans une langue donnée.

Nos travaux de dépouillement de données font émerger différents niveaux de synonymie en fonction du statut des lexies derrière l'entrée vedette. Si la traduction idéale est le niveau de synonymie le plus évident et le plus immédiat, la plupart des paradigmes de traduction ne sont pas conçus de façon homogène et strictement bijective : tel contexte entraînera l'émergence de plusieurs candidats et, *a contrario*, tel candidat pourra être présenté comme la traduction potentielle d'un mot de langue identique à l'entrée tout en n'étant pas l'entrée....

2.2 Les niveaux de synonymie

2.2.1 Notion de lien

La forme des dictionnaires sur lesquels nous avons travaillé a conduit à la distinction entre plusieurs types de synonymie (nous évitons de dire « niveau » pour ne pas établir - pour l'instant - de hiérarchie dans la valeur des liens choisis).

Les quatre premiers sont des liens directement accessibles, i.e. visibles pour le lecteur lors d'une consultation simple dans le(s) dictionnaire(s) ainsi que dans le tableau de données. Nous entendons par lien à la fois une relation de synonymie et/ou de traduction au sens linguistique (en l'occurrence lexicographique), ainsi qu'une relation univoque ou biunivoque au sens mathématique, celle-ci marquée par des liens orientés ou non au sein d'un graphe implantable. Les relations linguistiques sont ainsi formalisées mathématiquement dans un algorithme destiné à la programmation. L'expression la plus prosaïque du « lien » sera par conséquent un arc de graphe orienté, et les objets reliés des sommets.

Par ailleurs, les liens que nous présentons dans notre algorithme ne sont pas tous issus de l'observation d'un seul et même ouvrage lexicographique. Il convient ici de distinguer la synonymie sans lien avec la traduction (pour la délimitation du sous-champ sémantique de la lexie noyau en langue source) de la synonymie observée dans notre dictionnaire de référence. Pour cette dernière, plusieurs types de liens de

synonymie différents ont été relevés : des liens de traduction directe ou indirecte qui raccorde des lexies de langues différentes, et des liens de synonymie via la traduction qui mettent en relation des lexies de la même langue (source ou cible selon les cas).

Pour caractériser les liens que nous allons définir, des réponses doivent être apportées aux trois questions suivantes : qu'est-ce qui établit chaque type de synonymie, est-ce que la langue est la même pour les deux lexies, l'arc est-il orienté ?

2.2.2 Les liens synonymiques

2.2.2.1 Liens synonymiques sans lien avec la traduction : synonymie directe (SD)

Il s'agit du lien de synonymie par les synonymes et quasi-synonymes en langue source. C'est le seul lien qui n'est pas interprété à partir des dictionnaires de langues ; il s'agit de liens de synonymie répertoriés dans des dictionnaires de synonymes monolingues. Leur utilisation nous permet de visualiser des sous-graphes à partir d'une ou plusieurs lexies de référence (ou lexie-noyau) en réglant le problème des limites du champ. Ce lien est orienté pour les mêmes raisons que le suivant, en ce qui concerne les dictionnaires de synonymes.

a SD b ssi :

- b est synonyme de a dans le paradigme des synonymes et quasi-synonymes de a ;
- Langue a = Langue b ;
- arc orienté de a vers b.

2.2.2.2 Liens Synonymiques De Traduction

2.2.2.2.1 Liens repérables en première lecture

2.2.2.2.1.1 *Synonymie de Traduction (ST)*

C'est le lien de traduction directe correspondant à la note 1 de l'entrée et aux deux premières colonnes de notre tableau : il y a ST lorsqu'un terme est proposé comme candidat de traduction pour une vedette donnée. Le lien est orienté car il n'y a pas forcément symétrie entre les deux parties d'un même dictionnaire bilingue (cf. une nouvelle fois le commentaire sur les Lexiques Inversés).

a ST b ssi :

- b traduit a
- Langue a \neq Langue b
- Arc orienté de a vers b

2.2.2.2.1.2 Synonymie Partielle de Traduction en langue Cible (SPTC)

Le SPTC est un lien de synonymie entre deux lexies de langue identique qui traduisent le même terme de la langue source dans des contextes équivalents (note 10 de l'entrée-exemple, deuxième et quatrième colonnes de notre tableau). La présence d'au moins deux traductions d'une entrée lexicale à contextes équivalents conduit à établir une connexion forte entre deux entrées lexicales de deux lexèmes différentes en langue cible.

Ce lien est non orienté car le dictionnaire n'établit aucune hiérarchie ni aucune comparaison discriminatoire entre les deux termes par rapport à la lexie d'origine. On peut donc en déduire que les deux traductions sont interchangeables, en principe.

a SPTC b et b SPTC a ssi :

- a et b traduisent c à contexte équivalent, (contexte(a) = contexte(b)) \Leftrightarrow contexte(c)²²⁸
- (Langue a = Langue b) \neq Langue c
- Arc non orienté

2.2.2.2.1.3 Synonymie Partielle en langue Source (SPS)

Le SPS est le lien de synonymie entre la lexie à traduire et un autre mot de la langue source proposé comme équivalent dans un contexte donné, en rapport avec un ou plusieurs mot(s) en langue cible traduisant la première (note 7 de l'entrée-exemple, première et troisième colonnes du tableau).

Le lien est orienté car il n'y a pas là non plus de symétrie systématique entre les deux parties du dictionnaire.

²²⁸ contexte(a) est à comprendre comme « le contexte dans lequel le mot désigné par la variable a est employé ».

a SPS b ssi :

- c ST b et a est proposé comme équivalent à b dans le contexte de traduction qui mène à c, (contexte(a) = contexte(b)) \leftrightarrow contexte(c)
- (Langue a = Langue b) \neq Langue c
- Arc orienté de b vers a

2.2.2.2 Liens indirects de synonymie

Les niveaux suivants nécessitent la mise en place d'au moins une règle d'extraction intermédiaire entre le dictionnaire et le tableau d'un côté, et le graphe de l'autre.

2.2.2.2.1 *Synonymie indirecte (SI)*

Le SI est un lien de synonymie entre l'équivalent SPS de la vedette à traduire et la lexie qui traduit cette dernière (deuxième et troisième colonnes du tableau).

Le lien SI est orienté car il n'existe pas de symétrie observable pour ce cas. Toutefois, la même synonymie indirecte peut être observée en sens inverse, il y aurait donc dans ce cas neutralisation de l'orientation du lien.

a SI b ssi :

- a SPS c et b ST c
- (Langue a = Langue c) \neq Langue b
- Arc orienté de a vers b

2.2.2.2.2 *Synonymie absolue (SA)*

Nous employons le terme de synonymie absolue pour dénoter un double lien de traduction directe entre deux lexies. Autrement dit, il y a lien non orienté SA entre deux mots de langues différentes dès lors que le premier est proposé comme traduction directe du second et vice-versa.

De même, tous les autres liens orientés de la première liste peuvent être renforcés en cas de symétrie.

a SA b ssi :

- a ST b et b ST a
- Langue a \neq Langue b
- Arc Non Orienté

2.2.3 Préformalisation des liens de synonymie, réinterprétation

En résumé notre pré-algorithme (non encore implantable) se présente de la façon suivante :

Soit deux mots M1 et M2,

M1 est synonyme direct de M2 si :

- M2 traduit M1, Langue M1 (LM1) \neq Langue M2 (LM2). *Ex : to leave(English (E)) traduit **abandonner**(Français (F))*
- ou M1 traduit M2, LM1 \neq LM2
- ou M1 et M2 sont synonymes lexicographiques en langue source, LM1 = LM2

M1 est synonyme partiel de M2 si :

- M1 et M2 traduisent le même mot dans des conditions d'emploi analogues, LM1 = LM2. *Ex : **quitter**(F) et **partir de**(F) traduisent to leave(E)*
- ou M1 et M2 sont traduits par le même mot dans des conditions d'emploi analogues, LM1 = LM2. *ex : to leave(E) traduit **quitter**(F) et **partir de**(F)*

M1 est synonyme absolu de M2 si :

- M2 traduit M1 et M1 traduit M2, LM1 \neq LM2. *ex : **abandonner**(F) traduit to leave(E) et to leave(E) traduit **abandonner**(F)*
- ou M1 et M2 traduisent le même mot et sont tous deux traduits par le mot en question dans des conditions d'emploi analogues, LM1 = LM2.

ex : *quitter(F)* et *partir de(F)* traduisent *to leave(E)* **ET** *to leave(E)*
traduit *quitter(F)* et *partir de(F)*

M1 est synonyme indirect de M2 :

- Chaque fois qu'ils font partie du même champ synonymique sans lien direct ou absolu. Exemples : *abandonner(F)* traduit *to leave* et *quitter* traduit *to leave*. Autre cas de figure, *Abandonner(F)* traduit *to leave* et *to leave* traduit *quitter*. Dans ces deux cas, *quitter* et *abandonner* possèdent un lien de synonymie indirecte via *to leave*. Les liens de synonymie indirecte nécessitent l'élaboration d'une solide méthodologie complémentaire pour être correctement distingués les uns des autres.

3. Recensement des variables et lecture du tableau

La liste ci-dessous résume en les explicitant toutes les données présentes dans notre base synthétique. Le cas échéant, nous tentons de porter un regard sur une possible ontologie des éléments de chaque catégorie vue dans sa totalité.

3.1 Étiquettes de mot

Les étiquettes de mot concernent tous les verbes et toutes les périphrases verbales de chaque langue à insérer dans le graphe. Dans la mesure du possible, nous souhaitons implanter une seule étiquette par unité lexicale, même en cas de forte polysémie de celle-ci. Les périphrases verbales figées ainsi que les verbes à particule sont considérés comme des lexèmes à part entière dans notre étude.

Cette variable est valable pour *motcible*, *motsource*, *synonymemotsource* et *synonymemotcible*, lesquels renvoient dans le tableau à des objets de type *Mot*. L'ontologie de cette catégorie est équivalente à une nomenclature exhaustive des verbes de chaque langue en y incluant les périphrases les plus sémantiquement pleines.

3.2 Liens de synonymie

En complément aux niveaux de synonymie décrits dans le chapitre précédent, nous noterons que tout lien implanté concerne des lexies présentes dans la même ligne du tableau de données. Toutefois, des liens indirects peuvent émerger à partir du graphe. L'ensemble des liens de synonymie répertoriés a déjà été décrit dans ce chapitre.

3.3 Synonymes et quasi-synonymes

Les synonymes et quasi-synonymes sont extractibles pour chaque lexème visé, leur paradigme est restreint et borné. Afin de ne pas construire dans le flou les listes de synonymes et quasi-synonymes de chaque verbe, nous proposons d'utiliser, pour le français, les données fournies par la fusion des dictionnaires de synonymes réalisée à l'INALF et exploitée à Caen (Jacques François et son équipe) et à Lyon (Sabine Ploux). Lors de l'extraction de nos données bilingues par balayage, nous ne balayons plus dès lors que nous rencontrons une lexie n'appartenant pas à ce paradigme. Pour l'anglais, nous envisageons d'utiliser les paradigmes de composantes connexes anglaises utilisés par Sabine Ploux en relation avec les paradigmes en langue française. Aucune solution définitive n'est sélectionnée pour les autres langues pressenties puisque nous ne connaissons pas de travaux équivalents à ceux que nous venons d'évoquer en espagnol ou en italien.

Nous voyons le réseau final comme une grande sphère à l'intérieur de laquelle s'interconnectent autant de « sphères-bulles » qu'il y a d'unités lexicales, toutes langues confondues. Chacune de ces bulles a comme noyau central un lexème qui possède sa « sphère d'influence » sur tous les autres termes de ladite bulle. Cette vision est purement lexico-conceptuelle et s'oppose *de facto* à une modélisation sous la forme de traits aspectuels et événementiels. Toutefois, il n'y a pas de réelle incompatibilité entre les deux visions, celle qui nous est propre se rapproche de la volonté de Bruno Gaume (IRIT, Toulouse) de mettre des artistes à contribution pour transformer ses modélisations lexicales en œuvres artistiques.

3.4 Collocations sujet

La constitution d'une ontologie est probablement nécessaire, à partir du dictionnaire ou dans l'absolu (utilisation d'ontologies déjà existantes) pour dégager des classes de collocations, de la même façon que le dictionnaire, lequel ne propose pas des nomenclatures de collocations mais exclusivement des étiquettes de classes (*e.g. véhicule*) ou des objets prototypiques (*ex : voiture, camion, bicyclette*). Dans la mesure où notre recueil de données relève de la synthèse d'informations lexicographiques, nous envisageons de conserver telles quelles les données présentes dans les dictionnaires.

Détail important, les collocations sujet mentionnées concernent presque exclusivement les verbes signalés comme intransitifs. De ce fait, la pertinence sémantique du sujet semble être ignorée par les rédacteurs pour les verbes intransitifs, comme si la présence d'un objet suffisait à sélectionner le candidat approprié, ce qui correspond assez peu à la réalité de la langue.

3.5 Collocations objet

3.5.1 Ontologiquement parlant

La situation des collocations objets est la même que celle des sujets concernant la constitution éventuelle d'une ontologie, à ceci près que les collocations objet ne sont proposées que pour les verbes signalés comme transitifs. Ceci implique la caractérisation de la transitivité par l'objet et celle de l'intransitivité par le sujet, ce qui peut surprendre à première vue mais s'explique probablement en raison de la trop grande diversité des arguments sujets potentiels pour la plupart des lexies.

3.5.2 Structure argumentale : Généralités et lexicographie

En réalité, l'étude systématique de la structure argumentale des prédicats montre que la saturation complète des traits sémantiques et aspectuels de ceux-ci repose sur une connaissance globale de tous les éléments qui composent leur co-texte. De ce point de vue, les dictionnaires de langues que nous utilisons présentent une lacune importante pour le passage à une modélisation précise. Pour le verbe *traverser*, par exemple, il peut y avoir conflit entre les traductions proposées pour la collocation sujet [*projectile*] et celles proposées pour la collocation objet [*pont*], car les deux ensembles sont fondés sur des données stochastiques et prototypiques et ne distinguent pas des occurrences de ce type :

- 1) *L'obus a traversé le mur*
- 2) *Le passant a traversé le pont*
- 3) *L'obus a traversé le pont*

La troisième phrase est ininterprétable du point de vue du dictionnaire, car les traductions qu'elle induit sont contradictoires, selon qu'on pointe sur *le pont* comme édifice représentatif ou sur *l'obus* en tant que projectile. Comme nous le montrons dans le troisième chapitre, les données lexicographiques restreignent les conditions d'emploi de chaque lexie en fonction de critères génériques non généralisables. Dans l'exemple de l'obus qui traverse le pont, il y a prééminence de la transpercée (par l'obus) sur la traversée standard du pont, ce dernier étant vu en tant que parcours. Si l'utilisateur peut dans la plupart des cas effectuer cette inférence, cela n'est absolument pas transparent ni formalisable dans notre algorithme.

3.5.3 Séparation des Collocations Objet

Le second problème posé par les collocations objet réside dans la bivalence ou la trivalence de certains prédicats. La présence de deux compléments obligatoires (ou attendus) implique une duplication des collocations objet dans le tableau de données, à moins de faire apparaître une même traduction sur deux lignes successives, solution

peu efficace à notre avis. Par conséquent, nous y avons ajouté une colonne *collocobj2* en complément de la *collocobj1*.

Par souci de cohérence, les deux paradigmes collocationnels ne doivent pas contenir le même type de données. Nous allons montrer leurs différences à l'aide des exemples ci-dessous :

- 4) *Pierre parle de Paul à Marie*
- 5) = *Pierre parle à Marie de Paul*
- 6) *John is giving a book to Mary*
- 7) = *John is giving Mary a book*

Plusieurs constats syntaxiques sont à prendre en compte pour la comparaison entre français et anglais. Tout d'abord, l'ordre des compléments non pronominaux est souvent libre en français (4 et 5), alors que cette inversion n'est pas possible en anglais à moins d'un changement de structure infra-syntagmatique (6 et 7). Ensuite, le français ne permet jamais la présence de deux compléments syntagmes nominaux (SN) derrière le même prédicat verbal ; le cas des circonstants SN tels que « l'an dernier » n'entrant pas dans le calcul de la valence verbale. Ainsi, **donner Marie un livre* ou **donner un livre Marie* est un modèle impossible à reproduire depuis l'anglais et *donner un livre l'an dernier* fonctionne mais n'est donc pas à prendre en compte. Enfin, la présence d'au moins deux compléments dans la portée d'un seul prédicat implique de les discriminer en termes de rôles argumentatifs, lesquels serviront de critères pertinents pour la caractérisation de *collocobj1* et 2.

Tout ceci nous montre la forte interaction entre les données collocationnelles et les données structurales. Aussi, le tableau comme l'algorithme devront la faire apparaître en mentionnant le type d'objet visé pour chaque construction proposée.

3.6 Constructions syntaxiques indirectes

Dans notre tableau de données, mention est faite des constructions propres à chaque langue pour les transitifs indirects. Leur liste exhaustive est en apparence assez facile à constituer en raison du nombre limité de prépositions et conjonctions, et du nombre également limité de constructions qui les suivent (propositions complètes, syntagmes nominaux, groupe infinitif, etc. Voir en annexe).

3.7 Indices non langagiers

3.7.1 Registres et domaines

Leur ontologie est restreinte et relativement facile à constituer si l'on respecte une distinction rigoureuse entre les différents registres répertoriés. En effet, il convient de faire deux distinctions :

- Il ne faut pas confondre régionalisme, niveau de langue et domaine d'emploi, bien que ces informations soient typographiquement peu différenciées dans les ouvrages.
- Certains registres affichés peuvent être subsumés par d'autres. Rien ne distingue à première vue le domaine *sport* de *rugby*, cet exemple démontrant la nécessité probable de recourir à une ontologie hiérarchisée en ce domaine.

3.7.2 Informations complémentaires

Cet ensemble est également très délicat à répertorier, les informations n'étant pas toujours très homogènes. En particulier, il convient de faire la différence entre les données liées au contexte langagier et celles qui ne se rapportent qu'à la situation d'énonciation, voire à l'état d'esprit du locuteur. Les premières sont souvent des adverbes (pas nécessairement exprimés dans le discours et généralement valables ou pertinents pour une seule des deux langues mises en correspondance) et les seconds brassent beaucoup plus large. Néanmoins, la présence de ces informations demeure marginale dans l'ensemble du dictionnaire, ce qui n'en fait pas pour autant des éléments négligeables.

L'observation d'occurrences sur corpus montre que ces informations complémentaires sont souvent propres à un emploi du verbe en contexte mais ne présume pas de l'opportunité de sa généralisation dans les dictionnaires, ce qui fait de ses informations des traits prototypiques avant tout.

4. Unification des contextes

4.1 Pourquoi unifier

Pour déterminer la validité des liens de synonymie absolue et de synonymie indirecte (SA et SI), il faut procéder à un examen de compatibilité des contextes (comprendre : « structure contenant les informations sur les collocations, les constructions syntaxiques et le(s) registre(s) et domaine(s) d'emploi ») et en réaliser la fusion ou –à défaut- la réunion.

4.2 Comment unifier

4.2.1 Méthode

Deux contextes seront considérés comme compatibles s'ils sont identiques ou si l'un subsume l'autre. Pour tous les autres cas, il y a doute et besoin de vérifier la réelle non-compatibilité (notamment pour la synonymie absolue (SA), laquelle devrait découler d'une fusion systématique des contextes par souci de cohérence, mais est-ce toujours bien le cas ?).

L'unification des constructions syntaxiques ne doit pas être un problème : les emplois intransitifs et/ou transitifs directs sont explicitement mentionnés dans les deux langues, ainsi que les constructions syntaxiques particulières en cas de présence d'un objet indirect ou d'une complétive (infinitive ou phrastique). Sur ce point comme sur les autres, il n'y a guère de symétrie interlangue ; l'unification doit respecter le principe selon lequel une construction syntaxique ne se calque pas littéralement d'une langue sur l'autre.

4.2.2 Polémique sur le sens et la forme

Cette dernière réflexion en appelle une autre : la question de la prééminence de la syntaxe sur la sémantique, et/ou vice-versa.

De notre point de vue, la question ne peut se poser en ces termes. La syntaxe est ce qui est le plus directement perceptible dans le langage reçu (i.e. par l'auditeur) tandis que le sens est le « matériau cognitif » indéfinissable qui préexiste à sa mise en forme langagière (pour le locuteur). Toutefois, ceci étant posé, l'un ne va pas sans l'autre et inversement. Dire que la forme préexiste au sens reviendrait à percevoir le langage comme une succession de structures qui « prennent sens » après leur production, de même que prétendre que le sens prime sur la forme oblige à donner un aspect explicite - et donc explicitable en termes langagiers - à ce qu'est le premier, chose impossible à réaliser sans recourir au langage, fût-il métalinguistique.

Certaines théories de la philosophie du langage (Pascal, Guillaume, etc.) postulent qu'il n'y a pas de frontière entre la pensée et la parole, entre le 'méta-sens' et son expression langagière, entre la substance et la structure. « L'homme est un être de parole parce qu'il pense, et il pense car il est un être de parole. » (Pascal, 1662). Le renversement de l'axiome initial s'explique difficilement d'un point de vue mécanique ; pourtant, c'est la mécanique que Gustave Guillaume invoque pour expliquer sa « théorie psychomécanique du langage ».

Aussi, nous admettons que forme et sens (syntaxe et sémantique) sont fortement corrélés et interdépendants, et qu'ils n'entretiennent pas de relation de prééminence ni de préexistence l'un par rapport à l'autre. Cela suppose que nous nous posions le problème en termes de structure au sens informatique : quels sont les traits distinctifs d'une lexie en usage, comment les représenter dans une structure unique et comment y avoir accès. La hiérarchie des traits s'établit dès lors *de facto* par la linéarité du

langage : ce qui prime, c'est ce qui est en tête de la structure implantée, ce qui ne signifie nullement que le trait rencontré en premier subsume les autres. Il en va de même pour les traits de *registre*, de *domaines* voire de *données pragmatiques*, qui tous participent à et de la parole. Tout au plus peut-on relever certaines pistes de réflexion sur le fait que l'aspect figé et limité des structures d'une langue conditionne le sens, mais on signale que l'inverse est également vrai dans la mesure où les structures subissent la loi de la nécessité naturelle : si besoin est d'exprimer quelque chose, le langage qui est l'outil de cette expression fournit les matériaux de base pour enrichir le paradigme sémantique. En somme, la langue fonctionne par auto-enrichissement ou tout simplement par évolution naturelle.

4.2.3 Clôture de la polémique et solutions pour l'unification

Cette vision des choses peut évidemment poser certains problèmes méthodologiques dans la mesure où nous ne dégageons pas de hiérarchie de traits. Au demeurant, les besoins d'une hiérarchie de traits ne se posent pas dans le cadre d'une structure comme définie plus loin : une lexie se définit par un ensemble de traits pertinents, lesquels doivent tous être présents au sein de la même structure compositionnelle (un trait non pertinent ou non précisé est tout de même explicité en tant qu'ensemble vide).

Pour dépasser les obstacles induits par les vellétés d'unification de structures de traits compatibles dans les faits mais contradictoires en apparence, nous envisageons de revenir à l'idée d'union des données (au sens mathématique), au risque d'introduire dans le graphe des données redondantes.

Procéder de cette façon suppose que nous désorientions le graphe à un moment donné et surtout que nous admettions qu'un mot qui en traduit un autre est lui-même traduisible par celui-ci (principe du miroir). Linguistiquement parlant, cela reste discutable et nous y reviendrons.

5. Écriture de l'algorithme

5.1 Choix de la procédure de mise en œuvre

5.1.1 Outillage

Pour décrire précisément l'algorithme d'implantation du graphe, nous aurons recours aux notions et outils d'objets structurés, d'ensembles et de fonctions.

Les premiers contiennent l'ensemble des traits distinctifs pour une lexie donnée (globalement ou en contexte), les seconds sont utilisés en guise de listes de traits ou de valeurs associées à un attribut d'objet donné et les troisièmes permettront l'écriture de sous-programmes nécessaires à la bonne implantation des liens de synonymie.

Reprenons pour exemple quelques lignes de notre tableau de données légèrement adaptées. Nous avons choisi deux verbes pour illustrer la construction de

l'algorithme : *to leave* (A) et *parler* (F). Leurs caractéristiques respectives, notamment argumentales, sont très complémentaires et justifient ce choix.

5.1.1.1 Fragment de *to leave*

Figure 32 : seconde mise en forme d'un fragment d'entrée de *to leave*

<i>motsource</i>	<i>LS</i>	<i>Motcible</i>	<i>LC</i>	<i>SynLS</i>	<i>SynLC</i>	<i>obj1</i>	<i>obj2</i>	<i>subj</i>	<i>constrLS</i>	<i>constrLC</i>	<i>Registre</i>	<i>infoscomp ls</i>
leave	a	laisser	f		léguer	object, property			obj1 to obj2	obj1 a obj2		in will
leave	a	quitter	f			job, husband, wife						
leave	a	partir	f	go away	s'en aller			person, train, ship, etc				
leave	a	partir	f	resign	démissionner							

Dans les classifications verbales prototypiques, *to leave* est plutôt présenté comme un verbe de valence 1, c'est-à-dire que sa structure prédicative attend deux arguments : le sujet et un objet direct.

Toutefois, la polysémie de *to leave* le classe également parmi les verbes de valence 2 (*to leave sth to sb*) ou de valence 0 en tant qu'intransitif à sujet spécifique. Les trois cas de figure sont présentés dans l'exemple ci-dessus.

Outre la mise en exergue des constructions intransitives, transitives directes et bitransitives de ce verbe, l'autre intérêt de ce fragment est de présenter une possible présence de *synLS* et *synLC*.

5.1.1.2 Fragment de *parler*

Figure 33 : fragment sélectionné de *parler*

<i>motsource</i>	<i>LS</i>	<i>Motcible</i>	<i>LC</i>	<i>synLS</i>	<i>SynLC</i>	<i>obj1</i>	<i>obj2</i>	<i>subj</i>	<i>constrLS</i>	<i>constrLC</i>	<i>Registre</i>	<i>infoscomp ls</i>
parler	f	speak	a		talk							faculté phys
parler	f	talk	a		speak							faculté phys
parler	f	tell	a			quelque chose	quelqu' un		de obj1 obj2	about (2) obj1		pour informer

Parler présente des caractéristiques très différentes des prédicats verbaux appartenant au champ de l'abandon. En premier lieu, il s'agit d'un verbe naturellement

intransitif (valence 0) en tant que prédicat dénotant l'activité de parole, i.e. la faculté physique qui y est associée ainsi que l'activité qu'elle permet²²⁹.

Nous avons déjà largement insisté sur la dichotomie de surface entre *dire* et *parler* ; aussi, nous nous contenterons de rappeler que de nombreux verbes de parole peuvent être classés dans les deux ensembles selon leur emploi, y compris *parler* lui-même. Les emplois que nous présentons ci-dessous illustrent cette ambivalence :

- ou *parler* se rapporte à l'activité physique, à la capacité langagière et il est intransitif (deux premières lignes du tableau exemple) et se classe dans le paradigme qu'il chapeaute lui-même ;
- ou bien *parler* est employé pour introduire le thème et le destinataire du contenu de parole (ligne 3), auquel cas il n'est pas prototypiquement un verbe support de paroles rapportées (comme le sont les « cousins » de *dire*), mais il n'en devient pas moins un verbe bitransitif dont les traits sémantiques attendus (valence 0) sont sursaturés.

Ce fragment est également très intéressant du point de vue contrastif, car *to tell* est un verbe qui peut traduire aussi bien *dire* que *parler*, ainsi que *raconter*, moyennant des constructions syntaxiques différentes. Il est bitransitif doublement direct lorsqu'il se rapporte à *dire*, bitransitif doublement ou simplement direct pour *raconter* et bitransitif direct et indirect lorsqu'il traduit *parler*.

Cette dernière constatation le rapproche des nombreux verbes de parole français dont les emplois en contexte se rapportaient tantôt à l'un tantôt à l'autre des deux génériques, de manière souvent imprévisible.

5.1.1.3 Contraste entre les paradigmes verbaux

Il existe d'autres caractéristiques importantes dans la syntaxe des verbes qui nous intéressent ici. En particulier, nous remarquons que les verbes de type *parler* et *laisser* (cf. *abandonner*) sont potentiellement bitransitifs direct et indirect. Or, si l'objet direct est effaçable sans influence sur la présence de l'objet 2 avec *parler* (*parler de quelque chose, parler à quelqu'un*), cela n'est pas possible avec *laisser* (**laisser à quelqu'un, *laisser de quelque chose*) car la répartition thématique n'est pas la même. D'autre part, ces divergences syntaxico-sémantiques ont une influence (et inversement, sens ou forme d'abord ?) sur la sélection de la nature sémantique des arguments du verbe.

²²⁹ Faculté physique qui est précisée par les dictionnaires et reproduite dans les informations complémentaires de notre base.

En guise de conclusion intermédiaire, nous constatons une fois de plus que l'extrême complexité de la langue et du rapport interlangue rend ardu voire impossible le relevé exhaustif des manifestations langagières observables.

5.1.2 Objectifs

5.1.2.1 Justifications

L'algorithme dont nous avons besoin doit pouvoir transformer chaque ligne du tableau en, respectivement :

- un objet structuré de type mot qui contient des objets de type « caractéristiques du mot-source » comprenant *langue*, *collocobj1 et 2*, *collocsubj*, *constrLS*, *registre* et *infoscompls* ;
- un objet structuré qui contient des objets de type « caractéristiques du mot-cible » comprenant *langue*, *collocobj1 et 2*, *collocsubj*, *constrLC*, *registre* et éventuellement *infoscompls*. L'unification entre ces deux structures est automatique, seule la construction étant éventuellement différente (rappelons qu'elle n'est précisée que si le verbe est transitif indirect) ;
- éventuellement un ou deux objets structurés qui contiennent des objets de type « caractéristiques synonymes en langue source ou en langue cible » comprenant exclusivement les informations contextuelles et co-textuelles (*collocs*, *registre* et *infoscompls*) ainsi que la *langue* ;
- un lien de synonymie directe entre le mot-source et le mot-cible ;
- des liens de synonymie partielle qui relient les termes des quatre premières colonnes en dehors du lien de synonymie directe précité.

5.1.3 Définitions

5.1.3.1 Les objets structurés, attributs et valeurs

En informatique, un objet structuré est un ensemble de couples attribut/valeur. Un attribut est assimilable à un type générique auquel est attribuée une valeur spécifique. La valeur en question est une variable au type prédéfini, obligatoire ou optionnelle (valeur par défaut) : chaîne de caractères, élément d'un ensemble, ensembles d'objets, classes ontologiques (issues d'un ensemble), catégories, types en extension, etc.

Les objets que nous manipulons sont des lexies, auxquelles sont associés autant de couples attribut-valeur qu'il y a d'informations pertinentes dans notre tableau de synthèse. Les lexies seront implantées sous la forme de structures sous forme d'objets dont nous allons préciser les attributs.

5.1.3.1.1 Récapitulation des valeurs, type, facultativité, ontologie et valeurs par défaut

Les attributs utilisés et leurs valeurs attendues sont les suivants à partir des données strictement lexicographiques. Pour chacun, nous présentons le type de la valeur, la présence ou non d'une nomenclature et la nature de celle-ci le cas échéant, et enfin la valeur par défaut éventuelle :

Lemme

- chaîne de caractères obligatoire, correspondant au lexème lemmatisé (infinitif pour les verbes, singulier pour les substantifs, masculin singulier pour les adjectifs, etc.),
- nomenclature finie mais extensible (conforme au lexique d'une langue à un instant donné et sensible à son évolution),
- pas de valeur par défaut. Si le lemme n'est pas renseigné, il est impossible de construire une structure de mot-fantôme.

Catégorie

- type obligatoire,
- nomenclature finie et peu extensible, la catégorie par défaut est normalement V comme verbe, mais deux facteurs obligent à préciser la valeur pour toutes les lexies. En premier lieu, un verbe d'une langue source peut être traduit par un terme d'une autre catégorie grammaticale en langue cible. Deuxièmement, un certain nombre de nos lexies sont des locutions verbales. Quel que soit le degré de figement de ces dernières, elles n'ont pas toutes les propriétés syntaxiques liées à l'absolu figement. En particulier, il est toujours possible d'insérer un adverbe entre le lexème verbal et les autres constituants de la locution, ne serait-ce que le « pas » de négation (cf. chapitre I à ce propos),
- pas de valeur par défaut. L'idée de mettre *verbe* comme valeur par défaut a été abandonnée pour les raisons énoncées ci-dessus.

Langue

- type obligatoire, une lexie appartient nécessairement à une langue.
- nomenclature finie et extensible (si ajout de nouvelles langues dans les champs),
- pas de valeur par défaut.

Collocobj1

- chaîne de caractères facultative,
- nomenclature potentiellement infinie, difficile à délimiter par une ontologie issue ou non du dictionnaire. Le problème le plus épineux est la confusion entre les collocations englobantes (déjà ontologiques) et les collocations contextuelles, occurrenceielles et/ou prototypiques. Les différents types apparaissent dans le dictionnaire sans distinction typographique.
- en cas d'absence de complément, la valeur par défaut correspond aux constructions intransitives, ou transitives avec absence du complément hiérarchiquement supérieur (emploi absolu).

Collocobj2

- chaîne de caractères facultative, pas nécessairement liée à la présence d'un *collocobj1*,
- nomenclature potentiellement infinie, même problématique que pour *collocobj1*,
- la valeur par défaut correspond aux constructions intransitives si *collocobj1* n'est pas explicitée non plus, ou transitives avec absence du complément second.

Collocsuj

- chaîne de caractères facultative, bien que la présence d'un sujet soit obligatoire dans toutes les phrases, à l'exception des énoncés au mode impératif,
- nomenclature potentiellement infinie, comme les *collocobj*,
- pas de valeur par défaut. Si cet attribut n'a pas de valeur, nous devons en déduire que la nature du sujet « n'importe pas ». Malheureusement pour nous, ce parti pris est indéfendable.

Constr

- type et chaîne de caractères facultatifs, renvoie le cas échéant à l'une des deux collocations objet,
- nomenclature finie et relativement facile à constituer par recoupement des données répertoriées dans la base. La nomenclature de *Constr* comprend toutes les constructions syntaxiques complexes,
- la valeur par défaut renvoie soit à l'intransitivité si *collocobj* n'est pas remplie, soit à la transitivité directe sans préposition ni complémentateur.

Registre

- type facultatif, pouvant être de plusieurs sous-types : nous pouvons distinguer les domaines de spécialité des niveaux de langue et des zones géographiques d'emploi.
- nomenclature finie et assez facile à réaliser à partir des données. Pour bien faire, il ne faut pas qu'il y ait de recouvrement dans les étiquettes des différents sous-types,
- la valeur par défaut doit correspondre au langage courant non spécialisé ni régionalement marqué.

Infoscompls

- type ou chaîne de caractères facultatif, les *infoscompls* apportent des informations liées au contexte ou à un approfondissement du registre précédemment instancié,
- nomenclature potentiellement infinie et impossible à réaliser. Une information complémentaire est par nature imprévisible et inclassable. De plus, elles sont propres à chaque lexie et peuvent très bien n'apparaître qu'une fois dans la totalité de la base. Enfin, les informations complémentaires apportées par les études linguistiques sur corpus sont construites et augmentent d'autant la liste.
- la valeur par défaut n'enlève ni n'ajoute aucune information et ne signifie pas que la lexie n'a « rien à dire » en dehors de ses caractéristiques strictement structurelles et argumentales.

Les résultats des analyses à partir de corpus n'ayant pas motivé la création de nouvelles valeurs distinctives, la liste qui précède est par conséquent exhaustive en ce qui concerne notre travail en l'état actuel des choses.

5.1.3.1.2 Structure des lexies

Soit l'exemple suivant, ligne extraite de notre base :

Figure 34 : Ligne sélectionnée de *to leave*

<i>motsource</i>	<i>LS</i>	<i>motcible</i>	<i>LC</i>	<i>synLS</i>	<i>SynLC</i>	<i>obj1</i>	<i>obj2</i>	<i>Suj</i>	<i>constrLS</i>	<i>ConstrLC</i>	<i>registre</i>	<i>Infoscompls</i>
leave	a	laisser	f		léguer	object, property			obj1 to obj2	obj1 a obj2		in will

En nous mettant à la place de notre algorithme, cette ligne peut être glosée de la façon suivante :

« Le verbe anglais *to leave* peut être traduit par le verbe français *laisser* pour le cas où l'action serait effectuée *par testament (in will)*, prenant comme objet direct un individu de type *objet (object)* ou *propriété (property)*, et comme objet indirect n'importe quel individu précédé de *to* en anglais et de *à* en français. Dans ce contexte, le verbe français *léguer* est également une traduction potentielle de *to leave* ainsi que le synonyme contextuel (SPTC) de *laisser* en français. D'autre part, le dictionnaire ne prévoit pas pour cette acception de contrainte sur la sélection sémantique du sujet ni de l'objet 2. Il ne propose pas non plus de synonyme du terme source (SPS). »

Schématiquement, les structures contextuelles²³⁰ correspondant à notre ligne d'exemple se présentent comme suit. L'accès se fait à partir d'un identifiant d'un couple attribut-valeur pouvant lui-même être structuré :

Leave1 { lemme = « leave » ;
 Cat = V ;
 Lang = A ;
 Constr = collocobj1 + “ to ” + collocobj2²³¹ ;
 Collocobj1 classe = [objet, propriété] ;
 Collocobj2 classe = [] ;
 Collocsuj classe = [] ;
 Registre = 0 ;
 Infoscompl = ‘in will’ }

Laisser1 { lemme = « laisser » ;
 Cat = V ;
 Lang = F ;
 Constr = collocobj1 + « à » + collocobj2 ;
 Collocobj1 classe = [objet, propriété] ;
 Collocobj2 classe = [] ;
 Collocsuj classe = [] ;
 Registre = 0 ;
 Infoscompl = ‘par testament’ }

Léguer1 { lemme = « léguer » ;
 Cat = V ;
 Lang = F ;
 Constr = (Équivalent « laisser ») ;
 Collocobj1 classe = ([objet, propriété]) ;
 Collocobj2 classe = ([]) ;
 Collocsuj classe = ([]) ;
 Registre = (0) ;
 Infoscompl = (‘par testament’) }

Comme nous le voyons, une lexie est caractérisée par son lemme, sa catégorie grammaticale, la langue à laquelle elle appartient, puis par les constructions syntaxiques qu'elle induit, les collocations qu'elle admet dans un contexte de

²³⁰ Pour nous, une « structure contextuelle » est équivalente à un objet structuré représentant une lexie et ses conditions d'apparition.

²³¹ Cette ligne représente la concaténation d'une valeur et de deux attributs. Ces derniers sont représentés également dans l'objet structuré principal sous la forme de classes ontologiques.

traduction donné, et enfin le(s) registre(s) de langue lié(s) à ce contexte plus éventuellement des informations complémentaires sur la situation particulière d'énonciation et plus rarement sur l'état d'esprit du locuteur, ce dernier nous paraissant essentiel bien que presque totalement ignoré en lexicographie (voir chapitre III sur l'analyse du verbe *traverser*).

Les informations ainsi codées sont de natures diverses ; il appartient au modèle de faire apparaître ces modèles distinctement grâce à la typographie utilisée :

- Les chaînes de caractère qui correspondent à des fragments du discours sont présentées entre guillemets (« laisser »).
- Les types généraux destinés à situer la lexie dans son contexte métalinguistique sont en majuscules (Langue, Catégorie).
- Les classes d'objets notoirement englobantes sont mises entre crochets car elles ne correspondent que rarement aux chaînes de caractère réellement produites dans le discours ([objet, propriété]). Ce sont principalement des classes ontologiques dont le relevé est conditionné par l'ontologie des dictionnaires de référence.
- Les informations pragmatiques sont mises entre apostrophes. Elles relèvent à la fois du discours (*par testament* et *in will* peuvent très bien se retrouver derrière les verbes cités) et du métalangage, car ce sont surtout des indicateurs situationnels.
- Tout ce qui n'est pas précisé mais considérable comme valeur par défaut est présenté comme un ensemble vide entre crochets.
- Enfin, les données ininterprétables à partir du dictionnaire sont mises entre parenthèses. C'est le cas ici pour *léguer* dont le fonctionnement syntaxique n'est pas apparent, non plus que les contraintes de sélection sémantique des arguments. Toutefois, ces données sont récupérables dans une autre ligne de la base, celle qui présente *léguer* comme synonyme de traduction directe (ST) de *to leave* et *laisser* comme SPTC. Conséquemment, la ligne du tableau que nous décrivons sert avant toute chose à préciser ce lien SPTC entre *léguer* et *laisser* par l'intermédiaire du verbe anglais.

Certaines informations propres aux lexies de référence ne sont pas mises en exergue dans la structure synLC ou LS. En effet, les constructions syntaxiques en particulier ne conservent pas nécessairement les mêmes contraintes après substitution d'un équivalent. En l'occurrence, *léguer* se construit comme *laisser*, mais ce n'est

qu'un hasard (voir *doter (qqn de qqch)* par opposition à *léguer (qqch à qqn)*, par exemple). De même, l'information complémentaire *par testament* peut être considérée comme redondante lorsqu'on l'accrole au verbe *léguer*, alors qu'elle manquerait fâcheusement avec *laisser* qui ne comporte pas le trait *testamentaire* dans sa structure sémantique intrinsèque.

Certains critères de reconnaissance lexicale sont invariants : lemme, catégorie grammaticale et langue. Tous les autres (construction, collocations, registre et informations complémentaires) sont propres aux contextes visés et naturellement subsumés par la superstructure, laquelle correspond pour nous à la fusion de toutes les structures contextuelles d'un lemme donné. Ceci n'exclut pas l'invariabilité potentielle des critères dits fluctuants, comme l'objet direct pour *traverser*, qui est constant dans sa forme syntaxique à l'exception des emplois absolus.

Il convient d'ajouter à la superstructure de *to leave* toutes les données qui concernent cette lexie en tant que mot-cible. La construction des Lexiques Inversés montre qu'un complément d'information lexicale non négligeable est récupéré par ce biais.

Autre remarque importante : notre structure initiale ne rendait pas compte des constructions bitransitives, alors même que les exemples que nous présentons sont souvent des verbes à objets multiples. Par conséquent, il s'agit de faire apparaître les constructions de chaque verbe en contexte, et également de préciser si les collocations exprimées s'appliquent à la forme transitive directe ou à la forme indirecte. À titre d'exemple, notons qu'il n'y a pas de double transitivité directe en français du type « *raconter une histoire quelqu'un » mais possibilité en anglais, « to tell somebody something » versus « to tell something to somebody », formes qui impliquent l'introduction de notions nouvelles de portée syntaxico-sémantique du verbe ainsi que sa visée aspectuelle.

En premier lieu, nous avons choisi de séparer les arguments complémentaires en deux colonnes, l'une pour les objets « de type 1 » (objets directs ou objets indirects pour les verbes monotransitifs) et l'autre pour les objets de type 2 (le plus souvent objets indirects seconds, y compris ceux qui apparaissent après effacement de l'objet 1 attendu).

5.1.3.2 Unification des structures contextuelles : superstructure et sous-structures

5.1.3.2.1 Problèmes liés à l'unification

Après avoir défini les structures de chaque lexie, il reste à les unifier. En cas de synonymie directe, absolue et partielle, l'unification est automatique en cela que nous considérons (à juste titre) qu'une traduction respecte *a priori* le principe d'unification des traits qui composent les éléments lexicaux mis en jeu.

Nous relevons un nouvel écueil : les objets structurés que nous définissons contiennent des informations contextuelles pour une lexie et non pas des données fixes

pour une unité lexicale tous emplois confondus. En d'autres termes la *structure du mot-source* ne s'applique que pour un lien de traduction donné en direction d'une lexie possédant la *structure mot-cible* correspondante. Partant du principe que nous souhaitons « une seule étiquette de mot par lexème au sein du graphe final », il est donc nécessaire soit de simplifier les structures au maximum avec perte d'informations ou récupération ultérieure d'une sous-structure, soit d'appliquer les structures aux liens²³² et non aux étiquettes, voire de proposer des structures génériques²³³ modulables (i.e. prévoyant toutes les conditions d'emploi de l'unité lexicale considérée, avec application du principe d'unification sur une sous-structure de la structure générique).

En d'autres termes, nous entendons déterminer s'il est préférable d'implanter une structure d'attributs à valeurs multiples ou un ensemble unifié de structures d'attributs à valeurs contextuelles.

5.1.3.2.2 Une structure d'attributs à valeurs multiples

La structure générique incomplète et unifiée de « to leave » pourrait se présenter comme suit, les virgules indiquent la disjonction entre plusieurs éléments concurrents :

Superstruct mot-source { lemme = « leave », « left » ;
Cat = V ;
Lang = A ;
Constr = collocobj1, collocobj1 + “to” + collocobj2, 0 ;
Collocobj1 = [person], [thing], [place], [object, property],
[human being], [animal], etc. ;
Collocobj2 class = [person],[], etc. ;
Collocobj class = [person], [animal], [], etc. ;
Registre = 0 ;
Infoscompl = ‘in will’, 0 }

Il reste à la relier à d'autres structures génériques en ne perdant pas de vue que les liens ne concernent que des sous-structures :

SuperStruct mot-cible { lemme = « laisser », cf.< conj « laisser »> ;
Cat = V ;

²³² Ce que nous entendons par « Appliquer les structures contextuelles aux liens » suppose que nous demandions à l'algorithme de relier ces structures aux liens de synonymie qu'elle génère. Cette solution est celle que nous avons retenue au final par l'intermédiaire d'indices incrémentés à chaque fois qu'on change de lexie.

²³³ Structures génériques et superstructures sont équivalentes en terminologie et correspondent à l'ensemble unifié des structures contextuelles pour une unité lexicale donnée.

```

Lang = F ;
Constr = collocobj1, collocobj1 + « à » + collocobj2 ;
Collocobj1 class = [qqn], [qqchose], [objet, propriété2] ;
Collocobj2 class = [qqn] ;
Collocsuj class = [] ;
Registre = 0 ;
Infoscompl = 'par testament'=>collocobj1 class [objet,
propriété]
Indice = x}

```

```

Struct synLC { lemme = léguer ;
Cat = V ;
Lang = français ;
Constr TI = (à + collocobj) ;
Collocobj class = objet, propriété ;
Collocsuj class = [] ;
Registre = [] ;
Infoscompl = ('par testament')
Indice = y }

```

L'unification des sous-structures à partir des structures génériques suppose que le système soit capable d'extraire les informations liées à un emploi, et à lui seul. De ce fait, chaque fois qu'une donnée est encodée, il faut lui associer un indice qui la relie aux autres données contextuelles associées, d'où la présence de *indice* en fin de sous-structure.

Ce système de renvois systématiques revient finalement à « cacher » des sous-structures contextuelles dans un ensemble complet. Dans une perspective d'implémentation, il nous paraît possible de réaliser cette unification des sous-structures idoines à l'aide d'un système de pointeurs, bien que cette méthode soit particulièrement coûteuse en temps de programmation.

```

Unification Struct mot-source – Struct mot-cible { lemmes = « to
leave » |234 « laisser » ;
Cat = V ;
Lang = anglais | français ;
Constr TI = monotrans collocobj1 | ditrans collocobj1, « à » +
collocobj2 ;
Collocobj1 class = [qqn], [qqchose], [objet, propriété2] ;
Collocobj2 class = [qqn] ;

```

²³⁴ Le séparateur '|' marque le passage des caractéristiques de la langue de la première lexie à celles de la deuxième lexie.

Collocsuj class = [];
Registre = [];
Infoscompl = 'par testament'=>collocobj1 class [objet,
propriété] }

5.1.3.2.3 « Prétendre à l'exhaustivité », est-ce trop prétentieux ?

La solution de l'exhaustivité initiale nous paraît peu envisageable en raison du trop grand nombre de formes et de collocations subsumées par des lexies aussi polysémiques que celles sur lesquelles nous travaillons. N'oublions pas que nous étudions d'abord les rapports lexico-conceptuels entre mots de langues différentes, afin entre autres et avant tout de faire émerger des liens de sens entre ces dernières. En théorie, l'emploi d'un mot en contexte induit la subsumption d'une structure spécifique de traits langagiers par une structure plus vaste qui prévoit l'ensemble des traits pertinents (i.e. potentiels) pour un lexème donné. En ce sens, il est tout autant impossible d'imaginer l'intégralité des structures contextuelles possibles que la superstructure qui les engloberait toutes. En théorie toujours, la représentation contexto-conceptuelle étendue d'une seule unité lexicale est hors des compétences humaines.

En pratique, les ouvrages lexicographiques synthétisent les sens spécifiques pour les rendre accessibles à notre entendement, notre cerveau se chargeant de « désynthétiser » les données acquises, à chaque emploi d'un mot. Toutefois, l'étude en corpus nous fait rapidement franchir puis éclater les bornes imposées par la norme lexicale en ramenant le sens (la « structure spécifique») à un contexte langagier non synthétique. C'est le chemin qu'emprunte François Recanati (1997) pour élaborer sa théorie observable de la « polysémie absolue »²³⁵.

5.1.3.2.4 La superstructure comme ensemble de sous-structures

Grâce à la constitution de nos bases de données, nous pouvons élaborer des superstructures qui unifient automatiquement toutes les sous-structures d'un lexème donné, ces sous-structures étant elles-mêmes générées automatiquement à partir de la base de données d'origine, conformément à l'algorithme de construction des CSMU que nous décrivons ici.

Subséquentement, la superstructure est une structure à la fois « subsumante » et « résultante » des sous-structures. L'origine est le fruit de ses avatars, paradoxalement, et pourtant à la base de notre théorie structurelle.

Les structures contextuelles et les superstructures sont tout aussi importantes : les premières déterminent les liens lexico-sémantiques entre deux lexies et les secondes constituent la structure réelle, i.e. complète, d'une unité lexicale dans le

²³⁵ Revoir sur ce sujet « la polysémie, ça n'existe pas ? », chapitre I du présent document.

graphe. Loin de s’opposer, les données contextuelles et lexicographiques trouvent ainsi un terrain d’unification par l’intermédiaire de l’algorithme.

Si nous reprenons notre fragment-exemple du verbe *to leave*, nous pouvons proposer une représentation simple de la superstructure de la lexie :

Figure 35 : Fragment sélectionné de *to leave*

<i>motsource</i>	<i>LS</i>	<i>motcible</i>	<i>LC</i>	<i>synLS</i>	<i>SynLC</i>	<i>obj1</i>	<i>obj2</i>	<i>subj</i>	<i>constrLS</i>	<i>constrLC</i>	<i>registre</i>	<i>infoscompls</i>
leave	a	laisser	f		léguer	object, property			obj1 to obj2	obj1 a obj2		in will
leave	a	quitter	f			job, husband, wife						
leave	a	partir	f	go away	s’en aller			person, train, ship, etc				
leave	a	partir	f	resign	démissionner							

Superstruct mot-source {

{ lemme = « leave » ;
Cat = V ;
Lang = A ;
Constr = collocobj1 + “ to ” + collocobj2 ;
Collocobj1 classe = [object, property] ;
Collocobj2 classe = [] ;
Collocsuj classe = [] ;
Registre = 0 ;
Infoscompl = ‘in will’
Indice = x },

{ lemme = « leave » ;
Cat = V ;
Lang = A ;
Constr = collocobj1 ;
Collocobj1 classe = [job, husband, wife] ;
Collocobj2 classe = [] ;
Collocsuj classe = [] ;
Registre = 0 ;
Infoscompl = 0
Indice = y },

{ lemme = « leave » ;
Cat = V ;
Lang = A ;
Constr = 0 ;
Collocobj1 classe = [] ;

```

Collocobj2 classe = [] ;
Collocsuj classe = [person, train, ship, etc.] ;
Registre = 0 ;
Infoscompl = 0
Indice = z },

{ lemme = « leave » ;
Cat = V ;
Lang = A ;
Constr = 0 ;
Collocobj1 classe = [] ;
Collocobj2 classe = [] ;
Collocsuj classe = [] ;
Registre = 0 ;
Infoscompl = 0
Indice = w }
}

```

Cet ensemble ne rend pas compte des liens lexicaux que chaque sous-structure induit. Cela est le rôle de la base de données telle qu'elle est conçue, ainsi que celui de l'algorithme de construction du graphe synonymique que nous allons décrire.

Dans la superstructure de type verbe, il apparaît judicieux d’essayer de co-écrire les attributs de type construction et objets pour éviter de sur-spécifier dans la structure et parce qu’une construction donnée se rapporte à des types sémantiques et ontologiques d’arguments. Pour ce faire, nous pouvons élaborer une liste exhaustive des constructions possibles en partant du principe que toute lexie possède une construction de type objet1-objet2 pour laquelle la présence d’un des deux objets (ou des deux à la fois) n’est pas obligatoire :

Figure 36 : Typologie des constructions syntaxiques associées au verbe²³⁶

Type	Objet1	Objet2
<i>Intransitif</i>	^	^
<i>Monotransitif direct</i>	/	^
<i>Monotransitif indirect</i>	^	/
<i>Bitransitif doublement direct</i>	//	^
<i>Bitransitif direct et indirect</i>	/	/
<i>Bitransitif doublement indirect</i>	^	//
<i>Bitransitif indirect et direct</i>	/	/ (*avec inversion de l’ordre des compléments*)

Cette typologie ne rend pas toujours compte de l’ordre des arguments dans le cas d’une bitransitivité doublement directe ou doublement indirecte. Dans certains cas, cet ordre n’a aucune importance : on peut tout aussi bien *parler de quelque chose à quelqu’un* que *parler à quelqu’un de quelque chose* sans nuance réellement perceptible. Toutefois, il existe des cas de bitransitivité double pour lesquels l’ordre des arguments n’est pas fortuit : *to tell somebody something* n’est pas équivalent à **to tell something somebody* dont la forme correcte est *to tell something to somebody* (bitransitif direct et indirect) ; une telle dichotomie n’est pas possible en français : **raconter quelqu’un quelque chose* est une forme exclue. Conséquemment, la question ne se pose pas en français et dans les autres langues romanes, où l’ordre des compléments habituels n’est que très rarement contraint autrement que pour des raisons narratives et stylistiques (thématisation, emphase voire esthétique de la phrase), mais il faut parfois la prendre en compte pour l’anglais.

En plaçant constructions et objets dans le même type, nous obtenons un attribut *constrelementlangue* qui est soit un couple {type, info type d’objet} soit un « ménage à trois » {type, objet1, objet2}. Dans le premier cas, *info type d’objet* reste un attribut à type multiple qu’il nous appartient de séparer en {objet1, objet2}. Dans le second, *type*

²³⁶ Dans ce tableau, le symbole ^ correspond à « absence de (objet 1 ou 2) » et le / signifie « présence de ». Si la barre oblique est doublée (//), on doit comprendre « présence deux fois de l’objet considéré. »

correspond à la construction relevée dans le tableau ci-dessus et *objet1-objet2* sont les attributs correspondant à la forme et à la nature de chacun des arguments compléments.

Il ne reste plus qu'à définir la liste exhaustive des constructions possibles dans chacune des langues étudiées afin de constituer une bibliothèque de types à laquelle l'algorithme pourra se référer si nécessaire.

5.1.3.3 Implantation des liens

5.1.3.3.1 Travail sur les structures

Les liens que nous implantons aboutissant essentiellement à des unifications de structures de données, la manipulation de listes nous paraît une des meilleures manières de procéder. La première étape d'unification concerne la concaténation des structures contextuelles en une même superstructure, comme nous l'avons décrit dans les paragraphes précédents. Deux solutions s'offraient à nous :

- Soit la superstructure est conçue comme une liste d'objets structurés, ceci permettant en particulier d'implanter les sous-structures dans leur ensemble sans mélanger leurs traits avec ceux des autres. Cette solution est la plus facile à transformer en code mais est nettement plus coûteuse en cases mémoire.

Rappelons ce que cela donne :

```
Superstruct mot-source { [lemme = « leave », cf.conj “leave”];  
  Cat = V;  
  Lang = anglais;  
  Constrelementlangue = {BitransDI237, objet1, “to” objet2} ;  
  Objet1 = [] ;  
  objet2 = [ object, property];  
  Sujet = [];  
  Registre = [];  
  Infoscompl = ‘in will’  
  Indice = x ],  
  
[lemme = « leave », cf.conj “leave”];  
  Cat = V ;  
  Lang = anglais ;  
  Constrelementlangue = {MonotransD238, objet1, objet2} ;
```

²³⁷ Bitransitif direct et indirect, pour lequel seule le contenu sémantique l'objet indirect second est précisé.

Objet1 = [somebody];
objet2 = [];
Sujet = [];
Registre = [];
Infoscompl = ‘in will’
Indice = y],

[etc.] }

- Soit la superstructure est construite comme un « superobjet structuré » contenant pour chaque attribut autant de valeurs qu’il y a de traits dans une structure contextuelle. Cette dernière solution est préférable pour une simplification de la superstructure, ainsi qu’en termes d’économie de mémoire ; toutefois, elle contraint le programmeur à mettre des pointeurs sur chaque trait afin de le rattacher aux autres traits dont il est le voisin en contexte²³⁹.

Superstruct mot-source { lemme = « leave », cf.conj “leave”

Cat = [V]

Lang = [anglais]

Constr TI = [monotrans collocobj1 ; ditrans collocobj1, “to” + collocobj2]

Collocobj1 class = [[person] ; [sth] ; [place] ; [object, property²], etc.]

Collocobj2 class = [[person], etc.]

Collocsuj class = []

Registre = []

Infoscompl = [‘in will’=>collocobj1 class [object, property]] }

Le choix que nous avons fait est d’établir les sous-structures comme traits relatifs à la synonymie contextuelle reliant deux termes, et d’autre part d’affecter aux étiquettes de mots la réunion de toutes les structures contextuelles les concernant. Par conséquent, c’est la solution de la concaténation des objets structurés (et non celle de leur fusion) que nous avons choisie.

²³⁸ Monotransitif direct pour lequel seul l’objet1 est attendu. Si nous choisissons de faire de *constrelementlangue* un attribut à trois types, *objet2* est mentionné même s’il n’est pas pourvu.

²³⁹ plus des pointeurs de chaque élément sur la structure contextuelle subsumante

5.1.3.3.2 Notions d'ensembles et d'objets structurés

Un ensemble est une suite d'éléments qui peuvent être définis de deux manières :

- soit en extension, auquel cas l'ensemble correspond à une liste d'éléments distincts et disjoints du type $\{x_1 x_2 x_3 \dots x_n\}$;
- soit en intension, ce qui revient à manipuler un même élément un nombre de fois défini : $\{x^*\}$ (x répété un nombre de fois quelconque), $\{x^+\}$ (x répété au moins une fois), $\{x^*\}^{\omega_4}$ (x répété quatre fois), etc.

Un objet structuré est un ensemble de couples attribut-valeur, comme nous l'avons déjà dit précédemment. Une valeur peut elle-même être un couple attribut-valeur, ce qui permet une grande modularité des objets et sous-objets manipulés.

Dans notre nomenclature, un mot est défini par un ensemble sous-forme de quadruplet contenant quatre éléments, lesquels sont des attributs :

- mot = {lemme, catégorie, langue, constr}

Une lexie, au sens où nous employons ce terme, est caractérisée par un ensemble contenant les mêmes informations que le mot (au moins les trois premières), augmenté d'informations portant sur l'emploi précis de la lexie en contexte :

- lexie = {lemme, catégorie, langue, constrelementlangue, registre, infoscompls}

5.1.3.3.3 Listes et liens de synonymie

Les synonymes d'un terme source sont implantés au fur et à mesure de leur détection dans une liste dont il est la tête.

Lors de l'appel de fonction qui permet la création de liens de synonymie, le terme cible est placé en queue de liste et sa structure lui est associée. Notre première intention fut de déterminer les traits inhérents à un lien donné sous la forme d'une liste chaînée. De cette façon, il y aurait inclusion de listes chaînées (les structures de traits) dans une liste chaînée principale (les synonymes de type X d'un terme source), laquelle serait elle-même un élément d'une liste chaînée d'ordre supérieur (la liste des listes de synonymes, chacune de ces dernières se rapportant à un type de lien différent (cf. différents liens de synonymie)).

Pour de simples raisons pratiques, nous pensons préférable de traiter les structures de traits à part : superstructures (ou liste d'objets structurés) pour les étiquettes lexicales, et les objets structurés (ou structures contextuelles) pour chaque ligne de la base de données. Dans cette perspective, il sera peut-être nécessaire de

créer des pointeurs associés aux lexies présentes dans les listes et pointant sur lesdites structures. Ce parti pris nous ramène à la réflexion autour du sens du mot : est-ce qu'un pointeur est à même de représenter le sens de manière globale ?

5.1.3.4 Les fonctions

De nombreuses fonctions de programmation peuvent être nécessaires pour l'implantation des CSMU : fonctions d'extraction des données depuis la base, fonctions de construction des structures, fonctions de mise en correspondance des lexies avec les structures, fonctions de synonymies, etc.

Nous allons à présent en décrire les plus importantes, en laissant de côté les plus usuelles, à savoir les fonctions déjà existantes dans les bibliothèques de fonctions préétablies que chaque compilateur doit avoir en mémoire. Tout ce que nous allons décrire est indépendant d'un langage de programmation quelconque, bien que nous fussions de prime abord inspiré par celui que nous connaissons le mieux, le langage C.

Nous commençons par établir des profils de représentation. Toute fonction se caractérise par quatre éléments : son nom qui précise ce qu'elle fait, les paramètres qu'elle utilise pour fonctionner, le résultat du traitement des paramètres et le programme qui lui permet d'obtenir ce résultat. Dans la suite, nous ne parlerons que des trois premiers car nous nous limiterons à l'écriture d'un algorithme simple sans présumer de sa programmation. Dans l'ordre, nous écrirons les fonctions sur le modèle suivant :

(RESULTAT :type du résultat) : NOM_FONCTION (PARAMÈTRES :type du paramètre)

Les trois fonctions dont nous avons normalement besoin pour raccorder les données sont *Ajouter*, *Rechercher* et *Supprimer*. La fonction *Modifier* est également indispensable mais elle se résume à *Rechercher* + *Supprimer* + *Ajouter*. Il n'est donc pas nécessaire de la redéfinir.

Dans les lignes de code d'arrivée, les fonctions *rechercher* et *supprimer* ont été ignorées ou plus exactement fondues dans les fonctions *ajouter* qui seront au nombre de trois.

5.2 L'algorithme

Dans la première version de l'algorithme que nous avons conçue, nous avons séparé distinctement les étapes suivantes : extraction des données de la base, construction des superstructures pour chaque lexie, constitution des entrées par la concaténation des superstructures, implantation des liens de synonymie et raccordement des liens aux lexies. Ce mode d'écriture linéaire a posé d'insolubles problèmes, notamment dans la mise en correspondance des liens avec les bonnes

superstructures puis avec les lexies. Nous vous épargnerons la lecture de cet algorithme demeuré inachevé.... En définitive, toutes les opérations devaient pouvoir être effectuées en une seule manœuvre, ce qui nous a amené à concevoir une autre version que nous allons présenter ci-dessous.

Tout d'abord, nous présentons l'algorithme final sans le commenter. Puis nous le détaillerons avec des commentaires ligne par ligne afin d'en décrypter le fonctionnement. Ensuite, nous poursuivrons le travail de description en expliquant d'abord les fonctions et les objets que nous avons manipulés, puis en détaillant toutes les étapes de l'algorithme en langage naturel, explications supplémentaires incluses. Enfin, nous terminerons ce chapitre en essayant de montrer ce que nos bases de données et l'algorithme associé peuvent apporter aux représentations multilingues, à la lexicographie et éventuellement à l'aide à la traduction automatique et semi-automatique assistée par ordinateur.

5.2.1 L'algorithme sans commentaire

Quelques explications sur la typographie choisie sont nécessaires pour faciliter la lecture de l'algorithme : Les fonctions sont en petites majuscules ; les ensembles d'objets de même type sont notés entre crochets { } ; les ensembles de couples attribut-valeur hétéroclites sont entre parenthèses () ; les opérateurs habituels (boucles, actions, conditions, etc.) sont soulignés ; les constantes en position de valeur sont entre guillemets « » ; Les variables sont écrites avec une majuscule initiale (et peuvent être intégralement écrites en grandes majuscules pour certaines) ; chaque ligne de code active de l'algorithme se termine par un point virgule.

(Dico : Ensemble d'objets) : EXTRACTDONNEES (BD : xls)

(Valeur : Chaîne de caractères) : EXTRACT (TAB[x, y] : tableau)

(Dico : Ensemble d'objets) : AJOUT (Objdico : (Lemme, Cat, Lang, Econstr), Dico : Ensemble d'objets)

(Liens : Ensemble de liens) : AJOUTLIEN (Lien : (Indice, Indice, Type), Liens : Ensemble de liens)

(Econstr : Ensemble de constructions) : AJOUTCONSTR (Eltconstr : (Constr, Collocobj1, Collocobj2, Collocobj, Registre, Info compl), Econstr : Ensemble de constructions)

Ensemble-d'objets : (Objdico, Liens)

Objdico : (Lemme, Cat, Lang, Econstr)

Liens : {lien}

Lien : (Indice, Indice, Type)

(Econstr : ensemble de constructions) : {Eltconstr}

Eltconstr : (Constr, Collocobj1, Collocobj2, Collocobj, Registre, Info compl)

Dico = {^} ; Liens = {^} ;

Ligne = 1 ;

Motsource = TAB[ligne, 1] ;

LS = TAB[ligne, 2] ;

Numcons = 0 ;

Tant que *motsource .NE. ^ répéter*

Econstr = {^} ;

Tant que (*Motsource, LS*)= (*TAB[ligne, 1], TAB[ligne, 2]*) répéter

Numcons = Numcons + 1 ;

IndiceMS = Numcons ;

Numcons = Numcons + 1 ;

EltconstrMS = (Constr = EXTRACT (TAB[ligne,10]), Collocobjet1 = EXTRACT (TAB [ligne, 7]), Collocobjet2 = EXTRACT (TAB [ligne, 8]), Collocsuj = EXTRACT (TAB [ligne, 9]), Registre = EXTRACT (TAB [ligne, 12]), Infoscompls = EXTRACT (TAB [ligne, 13]), IndiceMS) ;

MC = EXTRACT (TAB[ligne, 3]) ;

LC = EXTRACT (TAB[ligne, 4]) ;

SMC = EXTRACT (TAB[ligne, 6]) ;

SMS = EXTRACT (TAB[ligne, 5]) ;

IndiceMC = Numcons ;

Numcons = Numcons + 1 ;

EltconstrMC = (Constr = EXTRACT (TAB[ligne,11]), Collocobjet1 = EXTRACT (TAB [ligne, 7]), Collocobjet2 = EXTRACT (TAB [ligne, 8]), Collocsuj = EXTRACT (TAB [ligne, 9]), Registre = EXTRACT (TAB [ligne, 12]), Infoscompls = EXTRACT (TAB [ligne, 13]), IndiceMC) ;

EconstrMC = {EltconstrMC} ;

AJOUTLIEN (Liens, (IndiceMS, IndiceMC, « ST ») ;

Objdico = (Lemme = MC ; Cat = « V » ; Lang = LC ; Constr = EconstrMC) ;

Dico = AJOUT (Dico, Objdico) ;

Si *SMS .NE. ^ alors début*

IndiceSMS = Numcons ;

Numcons = Numcons + 1 ;

EltconstrSMS = (IndiceSMS) ;

EconstrSMS = {EltconstrSMS} ;

AJOUTLIEN (Liens, (IndiceSMS, IndiceMC, « SI ») ;

AJOUTLIEN (Liens, (IndiceSMS, IndiceMS, « SPS ») ;

objdico = (lemme = MC ; cat = « V » ; lang = LC ;

Constr = EconstrMC) ;

Dico = AJOUT (Dico, objdico) ;

fin

Si *SMC .NE. ^ alors début*

IndiceSMC = Numcons ;

Numcons = Numcons + 1 ;

EltonstrSMC = (IndiceSMC) ;

EconstrSMC = {EltconstrSMC} ;

AJOUTLIEN (Liens, (IndiceMS, IndiceSMC, « ST ») ;
AJOUTLIEN (Liens, (IndiceSMC, IndiceMC, « SPTC ») ;
AJOUTLIEN (Liens, (IndiceMC, IndiceSMC, « SPTC ») ;
Objdico = (Lemme = SMC ; Cat = « V » ; lang = LC ;
Constr = EconstrSMC) ;
Dico = Ajout (Dico, Objdico) ;

fin

Si SMS.NE. ^ et SMC.NE. ^ alors AJOUTLIEN (Liens, (IndiceSMS,
IndiceSMC, « SI »)

EconstrMS = AJOUTCONSTR(EconstrMS, EltconstrMS)

ligne = ligne +1

Fintantque1

objdico = (lemme = Motsource ; cat = « V » ; lang = LS ; Constr = Econstr)

Dico = Ajout (Dico, objdico)

Fintantque2

5.2.2 L'algorithme commenté

Ci-dessous, nous allons expliquer le fonctionnement de l'algorithme ligne par ligne comme si nous étions à l'intérieur du programme en train de tourner. À chaque opération, nous allons voir ce qui se passe, ce qui est créé ou ce qui est ignoré. Les commentaires sont en écriture normale et placés entre balises /* (...) */ :

/**Déclaration des fonctions et des objets utilisés dans l'algorithme***/

(Dico : Ensemble-d'objets) : EXTRACTDONNEES (BD :xls)

/*EXTRACTDONNEES est une fonction qui transforme la base de données en un ensemble d'objets structurés sous l'appellation Ensemble-d'objets, défini plus bas. Cette fonction correspond aux actions de l'algorithme tout entier*/

(Valeur : Chaîne de caractères) : EXTRACT (TAB[x, y] : tableau)

(Dico : Ensemble d'objets) : AJOUT (Objdico : (Lemme, Cat, Lang, Econstr), Dico : Ensemble d'objets)

(Liens : Ensemble de liens) : AJOUTLIEN (Lien : (Indice, Indice, Type), Liens : Ensemble de liens)

(Econstr : Ensemble de constructions) : AJOUTCONSTR (Eltconstr : (Constr, Collocobj1, Collocobj2, Collocobj, Registre, Info compl), Econstr : Ensemble de constructions)

Ensemble-d'objets : (Objdico, Liens)

/*Le type du résultat Ensemble-d'objets, c'est un couple composé de l'ensemble des constructions de chaque mot et de la liste des liens*/

Objdico : (Lemme, Cat, Lang, Econstr)

/*Un ObjDico est caractérisé par son lemme, sa catégorie et sa langue (invariables pour un mot donné), et par l'ensemble des constructions d'un mot donné*/

Liens : {lien}

/*L'ensemble *Liens* est constitué de tous les liens de synonymie qui ont été implantés à partir de l'algorithme*/

Lien : (Indice, Indice, Type)

/*L'objet *Lien* est un triplet constitué de deux indices (le premier se rapporte au premier élément et le second au second élément réunis par un lien de synonymie) et d'un type (c'est-à-dire de quel lien de synonymie il s'agit)*/

(Econstr : ensemble de constructions) : {Eltconstr}

/*Définition de l'ensemble des constructions qui est composé de la liste de toutes les constructions qui correspondent à une lexie donnée*/

Eltconstr : (Constr, Collocobj1, Collocobj2, Collocsubj, Registre, Info compl)

/*Définition d'une construction, *EltConstr* est un élément de construction d'une lexie et se compose des informations présentes dans la ligne de la lexie et qui sont utiles pour la distinguer des autres lexies : Les collocations objet et sujet, la construction syntaxique, le registre et les informations complémentaires s'il y a lieu. *Eltconstr* ne comprend pas les valeurs lemme, langue, catégorie qui sont communes à toutes les lexies d'un même terme*/

/**Fin de déclaration des fonctions et des objets utilisés dans l'algorithme***/

Dico = {} ; Liens = {} ;

/*Initialisation du dictionnaire et de l'ensemble des liens*/

Ligne = 1 ;

Motsource = TAB[ligne, 1] ;

LS = TAB[ligne, 2] ;

Numcons = 0 ;

/*Les quatre dernières lignes permettent d'indiquer que l'on va opérer sur la première ligne du tableau, dans laquelle le *Motsource* correspond au contenu de la première colonne et la langue source (*LS*) au contenu de la deuxième. Par ailleurs, le compteur *Numcons* destiné à donner des indices à chaque construction et à chaque lien est également initialisé*/

Tant que motsource .NE. ^ répéter

/*Début de la première boucle : traitement de la base de données dans sa totalité : tant que la case *motsource* n'est pas vide, on continue les opérations qui suivent dans l'algorithme. En d'autres termes, l'algorithme tourne tant que l'on n'est pas arrivé au bout de la base de données*/

Econstr = {} ;

/*L'ensemble des constructions est initialisé pour chacun des termes traités*/

Tant que (Motsource, LS) = (TAB[ligne, 1], TAB[ligne, 2]) répéter

/*Début de la deuxième boucle, traitement des lignes concernant un mot source et un seul : tant que *Motsource* correspond au même lemme et à la même langue que celui

de la ligne précédente, on effectue les opérations qui suivent à l'intérieur de la boucle²⁴⁰*/

Numcons = *Numcons* + 1 ;

IndiceMS = *Numcons* ;

Numcons = *Numcons* + 1 ;

/*L'indice *Numcons* est incrémenté une première fois et sa valeur est transmise au mot source (via *IndiceMS*) puis on l'incrmente à nouveau en prévision de sa prochaine utilisation*/

EltconstrMS = (*Constr* = *extract*(*TAB*[*ligne*,10]), *Collocobjet1* = *extract* (*TAB* [*ligne*, 7]), *Collocobjet2* = *extract* (*TAB* [*ligne*, 8]), *Collocsubj* = *extract* (*TAB* [*ligne*, 9]), *Registre* = *extract* (*TAB* [*ligne*, 12]), *Infoscompls* = *extract* (*TAB* [*ligne*, 13]), *IndiceMS*) ;

/*À une lexie du mot source, on affecte *constrMS*, i.e. toutes les informations sémantiques et syntaxiques associées à cette lexie dans la ligne parcourue par l'algorithme. On y intègre également l'*indiceMS* qui a été initialisé plus haut.*/

MC = *TAB*[*ligne*, 3] ;

LC = *TAB*[*ligne*, 4] ;

SMC = *TAB*[*ligne*, 6] ;

SMS = *TAB*[*ligne*, 5] ;

/*De même que pour le mot source dans la boucle supérieure, on fait correspondre aux objets mot cible, langue cible, synonyme langue source et synonyme langue cible le contenu des cases qui leur correspondent, en guise de valeur. La langue de *SMS* et celle de *SMC* n'ont pas besoin d'être déclarées ici car elles sont identiques respectivement aux *LS* et *LC* de *MS* et *MC**/

IndiceMC = *Numcons* ;

Numcons = *Numcons* + 1 ;

/*L'indice incrémenté plus haut est affecté au mot source et il est incrémenté pour sa prochaine utilisation*/

EltconstrMC = (*Constr* = *extract*(*TAB* [*ligne*,11]), *Collocobjet1* = *extract* (*TAB* [*ligne*, 7]), *Collocobjet2* = *extract* (*TAB* [*ligne*, 8]), *Collocsubj* = *extract* (*TAB* [*ligne*, 9]), *Registre* = *extract* (*TAB* [*ligne*, 12]), *Infoscompls* = *extract* (*TAB* [*ligne*, 13]), *IndiceMC*) ;

/*A la traduction d'une lexie du mot source, on affecte *ConstrMC*, i.e. toutes les informations sémantiques et syntaxiques associées à cette traduction dans la ligne parcourue par l'algorithme. On lui affecte également l'*IndiceMC* qui a été initialisé plus haut.*/

EconstrMC = {*EltconstrMC*} ;

/*étant donné que le mot cible sera traité à part en tant que mot source, l'ensemble des constructions de *MC* est circonscrit à l'objet *EltconstrMC* qui ne comprend qu'un seul

²⁴⁰ On suppose pour cela que les lexies sont classées par ordre alphabétique du mot source dans la base initiale.

ensemble d'informations syntaxiques et sémantiques. Si on rencontre le même terme d'autres fois en traduction du mot source traité dans la boucle, la même opération sera répétée.*/
fin

AJOUTLIEN (Liens, (IndiceMS, IndiceMC, ST)) ;

/*Cette fonction ajoute à la liste des liens le lien de synonymie *ST* qui existe entre le *MS* et le *MC* (lien orienté).*/

Objdico = (Lemme = MC ; Cat = V ; Lang = LC ; Constr = EconstrMC) ;

Dico = AJOUT (Dico, Objdico) ;

/*Ajout dans le dico de l'entrée *MC*. Si elle n'existe pas encore, l'entrée est créée. Si elle existe déjà, la fonction *AJOUT* se contente de vérifier si le *Constr* fait déjà partie de l'entrée et si non, elle l'y ajoute*/

/***/ Traitement des synonymes *LS* et *LC* présents dans la base***/

Si SMS .NE. ^ alors début

/*La présence d'un *SMS* n'étant pas systématique dans une ligne, on vérifie d'abord sa présence : les lignes suivantes ne sont opérationnelles que s'il y a un *SMS**/

IndiceSMS = Numcons ;

Numcons = Numcons + 1 ;

/*La valeur du compteur *Numcons* est affectée à *SMS* via *IndiceSMS* et on l'incrémente pour sa prochaine utilisation*/

EltconstrSMS = (IndiceSMS) ;

EconstrSMS = {EltconstrSMS} ;

/**EltconstrSMS* ne contient que l'indice qui permet de référer au *SMS* car aucune information syntaxique et sémantique ne lui correspond directement dans la ligne où il apparaît. L'information est répercutée sur *EconstrSMS* afin de permettre la création de l'entrée*/

AJOUTLIEN (Liens, (IndiceSMS, IndiceMC, SI)) ;

AJOUTLIEN (Liens, (IndiceSMS, IndiceMS, SPS)) ;

/*Les trois liens de synonymie liés à la présence d'un *SMS* sont implantés dans la liste *Liens*, sauf si ces liens existent déjà*/

objdico = (lemme = MC ; cat = V ; lang = LC ; Constr = EconstrMC) ;

Dico = AJOUT (Dico, objdico) ;

/*Ajout dans *Dico* de l'entrée *SMS*. Si elle n'existe pas encore, l'entrée est créée. Si elle existe déjà, la fonction ajout se contente de vérifier si le *Eltconstr* fait déjà partie de l'entrée et si non, elle l'y ajoute */

fin

Si SMC .NE. ^ alors début

/*La présence d'un *SMC* n'étant pas systématique dans une ligne, on vérifie d'abord sa présence : les lignes suivantes ne sont opérationnelles que s'il y a un *SMC**/

IndiceSMC = Numcons ;

Numcons = Numcons + 1 ;

/*La valeur du compteur *Numcons* est affectée à *SMC* et on l'incrmente pour sa prochaine utilisation*/

EltconstrSMC = (*IndiceSMC*) ;
EconstrSMC = {*EltconstrSMC*} ;

/**EltconstrSMC* ne contient que l'indice qui permet de référer au *SMC* car aucune information syntaxique et sémantique ne lui correspond directement dans la ligne où il apparaît. L'information est répercutée sur *EconstrSMC* afin de permettre la création de l'entrée*/

AJOUTLIEN (*Liens*, (*IndiceMS*, *IndiceSMC*, *ST*) ;
AJOUTLIEN (*Liens*, (*IndiceSMC*, *IndiceMC*, *SPTC*)) ;
AJOUTLIEN (*Liens*, (*IndiceMC*, *IndiceSMC*, *SPTC*)) ;

/*Les trois liens de synonymie liés à la présence d'un *SMC* sont implantés dans la liste *Liens*, sauf si ces liens existent déjà*/

Objdico = (*Lemme* = *SMC* ; *Cat* = *V* ; *lang* = *LC* ;
Constr = *EconstrSMC*) ;
Dico = *Ajout* (*Dico*, *Objdico*) ;

/*Ajout dans *Dico* de l'entrée *SMC*. Si elle n'existe pas encore, l'entrée est créée. Si elle existe déjà, la fonction *AJOUT* se contente de vérifier si le *Eltconstr* fait déjà partie de l'entrée et si non, elle l'y ajoute */

fin

Si SMS.NE. ^ et SMC.NE. ^ alors AJOUTLIEN (*Liens*, (*IndiceSMS*,
IndiceSMC, *SI*))

/*Seulement en cas de présence conjointe d'un *SMS* et d'un *SMC* dans la même ligne, un nouveau lien de synonymie est implanté*/

EconstrMS = *AJOUTCONSTR*(*EconstrMS*, *EltconstrMS*)

/*Dernière opération de la boucle *tantque* qui traite une lexie : *EltconstrMS* est ajoutée à la liste *Econstr* des constructions de *MS**/

ligne = *ligne* + 1

/*On incrmente *ligne* pour traiter la ligne suivante dans la base, puis retour au début de la boucle*/

Fintantque1

/* Une fois qu'on sort de la boucle et en cas de passage à un autre *MS*, le mot source qui vient d'être traité est ajouté à *Dico**/

objdico = (*lemme* = *Motsource* ; *cat* = *V* ; *lang* = *LS* ; *Constr* = *Econstr*)
Dico = *Ajout* (*Dico*, *objdico*)

Fintantque2

/****Fin de l'algorithme, il n'y a plus de lexie à traiter dans le tableau****/

5.2.3 Description des fonctions et objets manipulés

5.2.3.1 Les fonctions utilisées

Cinq fonctions apparaissent dans l'algorithme, la première qui sert de chapeau à tout le programme est *EXTRACTDONNEES* ; la deuxième *EXTRACT* permet de formater le contenu de chaque cellule du tableau et d'effectuer sa transformation en couple attribut-valeur si la cellule est non vide ; les trois autres sont des fonctions d'ajout d'objets à un ensemble, *AJOUTMOT*, *AJOUTLIEN* et *AJOUTCONST*. Ces trois dernières fonctions peuvent effectuer des opérations complexes qui ne sont pas détaillées dans l'algorithme parce qu'elles sont facilement explicables :

- **EXTRACTDONNEES.** Cette fonction prend en paramètre la base de données au format Excel et la transforme en un graphe multilingue achevé. Il s'agit en fait de la fonction directrice de l'algorithme et en comprend la totalité des opérations. Sous *EXTRACTDONNEES* sont effectuées toutes les étapes de récupération des données de chaque ligne de la base, de la construction des structures contextuelles à l'établissement des liens de synonymie en passant par la constitution des ensembles : superstructures pour chaque terme, liste des liens, ajout des mots dans le dictionnaire principal, etc.
- **EXTRACT.** Cette fonction a pour but de récupérer les valeurs de chaque attribut (*Constr*, *Collocobj* et *subj*, etc.) d'une lexie ou d'un mot. Sa présence n'est théoriquement pas indispensable dès lors qu'on considère qu'on peut affecter une valeur nulle à un attribut. C'est précisément pour cette raison que nous utilisons *EXTRACT* car l'action de la fonction se décompose comme suit : *EXTRACT* recherche d'abord si l'élément visé existe dans la base. S'il n'existe pas, elle passe à l'attribut suivant sans affecter de valeur. S'il existe, la valeur est extraite du tableau et affectée à l'attribut en cours de traitement.
- **AJOUTCONSTR.** Cette fonction permet d'ajouter une nouvelle structure contextuelle d'une lexie (*EltconstrMS*) à la superstructure d'un mot qui contient toutes celles qui ont déjà été repérées et ajoutées précédemment (*EconstrMS*). Autrement dit, elle permet l'ajout à un lexème de toutes les informations qui permettent d'en différencier les emplois. La fonction ne s'applique qu'au mot source puisque la base est préalablement triée par ordre alphabétique du mot source et qu'on part d'abord du principe qu'un mot cible n'est présent que dans une seule ligne par mot source (nous verrons que cela n'est pas forcément le cas, et comment résoudre le problème).

- **AJOUTLIEN.** Cette fonction a pour but d'ajouter tout lien de synonymie nouvellement créé à l'ensemble des liens de synonymie (*Liens*). Si le lien existe déjà dans *Liens*²⁴¹, il n'est pas réimplanté.
- **AJOUTMOT.** Dès qu'une structure contextuelle est créée, *AJOUTMOT* ajoute le lexème visé (*Objdico*) au dictionnaire (*Dico*). Si l'entrée existe déjà, le programme se contente de réactualiser la superstructure *Econstr* de *Objdico* en vérifiant que la nouvelle structure contextuelle n'y figurerait pas déjà²⁴². Dans ce dernier cas, seul l'indice de la structure *Eltconstr* est ajouté à la structure déjà répertoriée dans l'ensemble.

Bien que nous ne les ayons pas détaillées plus avant, les trois fonctions *Ajouter* ont le même comportement face à la redondance : avant d'inclure l'élément visé à l'ensemble où il doit apparaître, la fonction vérifie qu'il n'a pas déjà été implanté et ne l'ajoute que s'il n'existe pas encore. En ce sens, la fonction *Ajouter* contient une fonction *Rechercher*, comme nous l'avons annoncé avant la présentation de l'algorithme.

5.2.3.2 Les autres paramètres manipulés

Nous allons présenter ici les différents objets nécessaires pour l'écriture de l'algorithme. Tout ce qui a été présenté précédemment (par exemple les différents liens de synonymie) ne sera pas redétaillé ici.

- **Dico.** *Dico* correspond au dictionnaire intégral qui concentre toutes les données mises en relation au moyen de l'algorithme. Il est le résultat de la fonction principal *EXTRACTDONNEES*. C'est un ensemble de type *Ensemble d'objets*.
- **Ensemble-d'objets.** Comme son nom l'indique, c'est un ensemble d'objets et se décompose en deux : la liste des superstructures de chaque mot (*Objdico*) et la liste des liens de synonymie (*Liens*). La réunion de ces deux ensembles constitue le *Dico*.

²⁴¹ En général les liens non orientés sont construits deux fois par l'algorithme, une fois pas mot source. Ceci implique que la fonction s'assure d'abord que le lien n'existe pas encore avant de l'ajouter.

²⁴² En théorie, ce cas de figure ne se rencontre jamais pour le mot source, une structure contextuelle ne pouvant correspondre qu'à une lexie, et donc à une ligne. Dans la pratique, si un verbe peut être traduit par plusieurs verbes de la langue cible dans des emplois identiques, la structure est répétée sur plusieurs lignes. Il faudrait donc le prévoir dans la définition de *Ajoutmot*.

- **Objdico.** *Objdico* est un objet qui possède quatre paramètres qui correspondent pour les trois premiers aux critères de dégroupement propres à chaque lexème : lemme, catégorie, langue (*Lemme*, *Cat*, *Lang*, qui sont des valeurs stables et prédéfinies) ainsi qu'à l'ensemble des structures contextuelles possibles pour cette entrée (*Econstr*, qui est construite par l'algorithme).
- **Liens.** *Liens* est un ensemble qui contient la totalité des liens de synonymie créés par l'algorithme et enregistrés après leur création. Son paramètre est *{lien}*.
- **Lien.** Le lien de synonymie est constitué de trois paramètres : l'indice qui renvoie à la lexie visée du premier terme du lien, l'indice qui renvoie à la lexie du second terme et le type de synonymie dont il s'agit (celui-ci est une constante et se note entre guillemets, « ST », par exemple). Lorsqu'on interroge un lien de synonymie entre deux termes, l'algorithme renvoie directement aux lexies équivalentes grâce aux indices.
- **Econstr.** C'est la liste des structures contextuelles *Eltconstr* (qu'elle a en paramètre) pour chaque lexème.
- **Eltconstr.** L'élément de construction *Eltconstr* correspond à une structure contextuelle d'une lexie. Elle se constitue des informations linguistiques propres à une lexie (constructions, collocations, registre, informations complémentaires) et également d'un indice identique aux liens de synonymie qui lui sont associés. La liste des *Eltconstr* d'un lexème donné est contenue dans la superstructure *Econstr*.
- **Numcons.** Le compteur *Numcons* permet d'affecter des indices à chaque structure contextuelle et aux liens de synonymie qui vont avec. Il est incrémenté à chaque utilisation.
- **Indice.** Les indices prennent la valeur de *Numcons* après chaque incrément et sont affectés à la structure contextuelle en cours de construction dans l'algorithme (de *MS*, *MC*, *SMS* ou *SMC* selon le cas).

5.2.4 L'algorithme en langue naturelle : détails et explications

À présent que nous avons posé et expliqué chaque ligne de l'algorithme ainsi que les fonctions et les objets qui nous ont servi à l'élaborer, nous allons en faire la

description en langage naturel. Les explications qui vont suivre sont conformes à la progression de l'algorithme et sont avant tout destinées à en expliciter certains points.

Tout d'abord et avant d'entrer dans l'algorithme proprement dit, la base de données initiale existe sous forme d'un tableau, chaque cellule étant par conséquent caractérisée par un indice de type « élément de tableau à deux paramètres » $TAB[i,j]$. Le paramètre nécessaire pour l'**extraction** est la base de données elle-même, sous la forme d'une page Excel. On peut supposer à tort ou à raison que la fonction de récupération d'une feuille de tableur est prédéfinie dans la bibliothèque de tout compilateur, au même titre que pour les fichiers de type texte. Toutefois, même si elle ne l'était pas, notre propos n'est pas ici d'en préparer la programmation.

La première fonction à appeler (*EXTRACTDONNEES*) est celle qui va extraire les données de la base, se présentant donc pour nous sous forme de tableur, et les transformer en structures contextuelles. Autrement dit, il s'agit de la fonction qui va exploiter les données de la base ligne par ligne.

La fonction d'extraction nécessite la définition d'au moins trois autres ensembles de procédures : la transformation de la base de données en un tableau exploitable par la programmation (pas explicité ici puisque nous partons du principe qu'il s'agit d'une fonction prédéfinie), la recherche de la présence d'éléments dans une case (*ExisteElément* par exemple..., notons que cette dernière n'est pas indispensable en cas d'appel négatif du contenu d'une case du tableau) et la construction des structures contextuelles, laquelle consiste en l'affectation du contenu de la cellule du tableau qui lui correspond à chacun des attributs d'une lexie.

Les fonctions de recherche ne posent pas de problème particulier, si ce n'est dans l'emplacement de chaque élément recherché. Dans cette optique, il ne faut pas oublier de préciser à chaque tour de boucle l'indice de la colonne où doit se trouver l'élément que l'on recherche. Toutefois, cet indice est constant tout au long de la base, ce qui ne devrait pas poser le moindre problème de détection (à un critère donné correspond un numéro de colonne donné). La recherche de l'existence d'un mot peut être utilisée dans la fonction qui transforme la base pour signifier qu'on ne remplit des cases mémoire qu'en cas de présence d'un objet au moins dans la cellule correspondante. Elle est également utile - voire indispensable - pour déterminer s'il y a lieu d'associer une structure à une cellule ; sans cette condition, le programme créera des structures pour toutes les cellules vides... et donc des structures vides et inutiles, mais demandeuses de mémoire. Dans l'algorithme, la fonction de recherche est absorbée, et ce doublement :

- par les fonctions *AJOUT* car nous avons décidé que dans le fonctionnement interne de celles-ci, il est prévu que l'élément sera recherché avant d'être joint à l'ensemble correspondant ;

- par l'affectation des valeurs aux attributs qui constituent les structures contextuelles. La fonction *EXTRACT* est conçue pour d'abord rechercher une case non vide avant « d'autoriser l'existence » de l'attribut.

Après la déclaration des fonctions et des ensembles que nous avons définis un peu plus haut, les premières opérations de l'algorithme sont l'initialisation du dictionnaire (*Dico*) et de l'ensemble des liens (*Liens*), la mise en position à la ligne 1 du tableau, l'affectation au mot source et à la langue source de leurs valeurs respectives (sans fonction *EXTRACT* car *Motsource* et *LS* sont **toujours** renseignés dans le tableau) et l'initialisation du compteur *Numcons* qui va nous servir pour affecter des indices à chaque élément construit par l'algorithme, qu'il s'agisse des structures des lexies ou des liens de synonymie. Une fois que tout est initialisé, la procédure d'extraction et de mise en relation des informations peut commencer.

L'algorithme est régi par deux boucles de type *tant que* qui autorisent les opérations à se dérouler jusqu'à ce qu'il n'y ait plus de données à traiter dans le tableau :

- La première boucle principale concerne le tableau tout entier, elle stipule que « tant qu'une ligne du tableau contient des informations, et avant toute chose tant que la cellule *Motsource* contient une valeur non nulle, les lignes qui suivent sont activées » ; autrement dit, on s'arrête dès que le parcours de l'algorithme rencontre une ligne vide ou plus exactement une cellule *Motsource* vide. À l'intérieur de cette boucle principale, l'ensemble des structures contextuelles est initialisé pour chaque lexème traité. Encore une fois, nous insistons sur le fait que le tableau est préalablement classé par ordre alphabétique pour autoriser cette procédure.
- La seconde boucle contenue dans la première se rapporte à chaque terme, ce qui suppose donc que nous traitons les entrées une par une en position de *Motsource*. Elle stipule que « tant que le mot source et la langue source de la ligne en cours de traitement sont identiques au mot source et à la langue source de la ligne précédente²⁴³, on effectue les opérations qui suivent jusqu'au renvoi au début de la boucle ».

²⁴³ Ce faisant, on se prémunit du fait qu'une seule graphie peut être commune à des entrées de langues différentes. Les deux paramètres sont donc nécessaires pour éviter que l'algorithme n'associe pour le même mot des structures contextuelles de langues différentes (par exemple, les verbes français et espagnol *salir* qui ont la même graphie et la même catégorie grammaticale pour le lemme de référence et qu'il faut par conséquent distinguer l'un de l'autre par la langue).

Ceci étant posé, on effectue l'incrémentation du compteur *Numcons* et l'affectation de sa nouvelle valeur au mot source via *IndiceMS*, celui-ci restant constant pour une lexie donnée (i.e. à la fois pour la structure unique d'une lexie et pour tous les liens de synonymie qu'elle génère). Ce compteur sera incrémenté à chaque fois qu'il faudra donner un indice de référence à une lexie quel qu'en soit le type (*MC*, *SMS* et *SMC* inclus). Grâce à cela, on ne retrouvera jamais dans *Dico* deux lexies différentes avec le même indice, ce qui facilitera la recherche d'informations pour l'utilisateur à l'appel du programme. Dans la grande majorité des cas, une ligne du tableau contient au moins un mot source et un mot cible. Pour la suite de cette description, nous considérerons ce principe comme constant²⁴⁴.

La première opération d'extraction de données consiste en la fabrication de la structure contextuelle du mot source dans la ligne considérée. Pour ce faire, on affecte à cette structure *EltconstrMS* les couples attribut-valeur suivants : « Construction-contenu de la case 10 », « Collocation objet 1-contenu de la case 7 » et ainsi de suite pour les autres critères retenus, collocation objet 2, registre et informations complémentaires. Il faut noter que *EltconstrMS* contient également en valeur l'indice qui a été affecté à *MS*, sans quoi on ne pourrait pas récupérer cette structure contextuelle ultérieurement.

S'ils existent, le mot cible, la langue cible (qui vont nécessairement de pair), le synonyme en langue source et le synonyme en langue cible prennent comme valeur le contenu de la case où ils apparaissent normalement. S'il n'y en a pas dans la ligne traitée, la fonction *EXTRACT* a pour mission de les ignorer et de renvoyer une « valeur nulle » qui entraîne incessamment la suite des opérations. Il est à noter que la présence d'un *MC* ne préjuge pas de la présence d'un *SMS* ou d'un *SMC*. En revanche, dès qu'il y a un *SMC*, il y a forcément un *MC*. Il n'y a pas besoin de récupérer l'information concernant la langue pour les deux synonymes éventuels car celle-ci est identique à celle de *MS* ou de *MC*. De ce fait, l'information sera simplement récupérée par le biais de la superstructure de *SMS* et de *SMC* en fin d'algorithme.

S'il y a un mot cible, donc le cas le plus fréquent, l'indice est d'abord incrémenté pour être différencié de l'indice du *MS*, puis on effectue la construction de la structure contextuelle de *MC* qui correspond à la lexie de *MS*. En l'état de notre base de données, *MS* et *MC* partagent les mêmes informations s'ils apparaissent sur la même ligne, sauf en ce qui concerne la partie construction qui est justement destinée à montrer des différences syntaxiques entre les deux langues mises en rapport. Ainsi, les collocations, le registre et les informations complémentaires sont valables aussi bien

²⁴⁴ Par l'intermédiaire de plusieurs sources (synonymes de la même langue, analyses linguistiques des chapitres III et IV), nous avons suggéré d'enrichir la base par des données qui n'ont pas de lien avec la traduction. Ce cas de figure n'est pas pris en compte dans notre algorithme principal mais le sera en conclusion à ce chapitre. Pour l'instant, nous ne nous intéressons qu'aux cas de traduction « réelle ».

pour la lexie source que pour celle qui la traduit²⁴⁵. Dès qu'elle est créée, la structure contextuelle du mot source *EltconstrMC* est ajoutée à la superstructure de *MC EconstrMC*. La présence d'un élément MC dans la base indique qu'il s'agit d'une traduction de *MS*, ce qui induit la création d'un lien de synonymie de type « synonyme de traduction (*ST*) » qui est immédiatement ajouté à la liste des liens grâce à la fonction *Ajoutlien*, pour laquelle nous rappelons qu'elle commence par vérifier que le nouveau lien à implanter n'existe pas déjà. Comme nous l'avons précisé pour cette fonction, un lien nouveau est constitué de trois paramètres : les indices de chaque lexie (se rapportant à leurs structures contextuelles) et le type du lien dont il s'agit (en l'occurrence *ST*). Enfin, l'entrée *MC* est ajoutée à *Dico* sous la forme de *Objdico*. Comme pour les autres ajouts, l'entrée est créée si elle n'existe pas encore ; si elle existe déjà, la fonction *AJOUT* se contente de vérifier si la construction *Constr* qui vient d'être formatée existe elle aussi, et l'y ajoute seulement si elle est déjà répertoriée à l'identique. Si cet *Constr* existait déjà dans l'*Objdico* se rapportant à *MC*, seul les indices des *EltconstrMC* sont reportés dans *Objdico*.

Toujours à l'intérieur de la seconde boucle, il reste à traiter les synonymes LS et LC dont la fréquence d'apparition n'est pas très élevée mais est loin d'être négligeable. Quatre cas de figure se présentent sous conditions :

- soit il n'y a ni *SMS* ni *SMC*, auquel cas les procédures conditionnelles sont ignorées (*si non-SMS alors on passe..., si non-SMC alors on passe...*) ;
- soit il y a un *SMS* et pas de *SMC*. *SMS* subit alors les mêmes opérations que *MS* et *MC* précédemment avec quelques adaptations : le *SMS* reçoit à son tour un indice différent des précédents, on construit son *Eltconstr* qui ne contient pas d'autre information que l'indice, on ajoute les liens de synonymie induits par la présence d'un *SMS* et l'entrée du *SMS* est créée par *AJOUT* sur le même modèle que *MS* et *MC*²⁴⁶ ;
- soit il y a un *SMC* et pas de *SMS*. Ce que nous venons de décrire pour *SMS* est appliqué à *SMC* ;

²⁴⁵ Nous admettons ce principe pour une écriture simplifiée de l'algorithme et de la base de données, mais nous admettons plus encore que le principe ne peut pas être valable dans tous les cas de figure.

²⁴⁶ Nous faisons remarquer que c'est essentiellement l'indice qui est reporté dans *Dico*, et non pas les informations linguistiques éventuellement liées au *SMS* dans un contexte précis, celles-ci faisant l'objet d'un traitement à part lors de la présence du même lemme en position *MS* dans d'autres lignes de la base de données.

- soit les deux sont présents. Aux commandes ci-dessus, effectuées dans l'ordre établi, on ajoute simplement le lien de synonymie indirecte (*SI*) qui caractérise dans notre typologie la relation entre un synonyme langue source et un synonyme langue cible.

Une fois que les opérations conditionnelles ont été passées en revue, il ne reste plus qu'à effectuer les dernières opérations d'enrichissement du dictionnaire, à savoir l'ajout à *EconstrMS* de la structure contextuelle de *MS*, laquelle n'a pas encore été adjointe à la superstructure de *MS*. Cette dernière opération, avant l'incrémentation de la ligne, permet de stocker sous un même lemme l'ensemble des informations relatives à chaque lexie. Comme nous l'avons déjà dit, la superstructure fonctionne comme un ensemble d'objets structurés, ce qui donne par exemple pour un lemme quelconque :

```
(lemme = «   »
Cat = V
Lang = A
Constrementlangue = {(Typeconstr = MonotransD);
  Objet1 = [x1], [x2], ... [xn] ;
  Objet2 = *
  Sujet = *247 ;
  Registre = * ;
  Infoscompls = ... ) ;
Constrementlangue = {(Typeconstr = BitransDI;
  Objet1 = * ;
  Objet2 = [x1], [x2], ... [xn] ;
  Sujet = * ;
  Registre = * ;
  Infoscompls = [] ) ;
etc.}
```

En résumé, la fonction *EXTRACTDONNEES* extrait les éléments de chaque cellule, elle les formate comme valeur d'attributs (en fonction du deuxième paramètre du tableau, celui qui correspond à la colonne) et enfin crée des couples attribut-valeur formatés qui vont se placer dans les ensembles auxquels on les fait appartenir. Toutes ces opérations s'effectuent au sein d'une boucle dont la condition d'arrêt correspond à l'arrivée en fin de ligne du tableau (une ligne = une lexie dans la base de données). Ensuite, on répète l'opération tant qu'il y a des lignes non formatées dans la base, jusqu'à la fin de la boucle principale qui dit que « tant qu'une ligne contient des informations, on répète les opérations ».

²⁴⁷ On fait correspondre ici le signe '*' à un élément quelconque par opposition à [] qui est l'ensemble vide.

Pour chaque ligne, l'algorithme crée jusqu'à quatre structures contextuelles (selon le nombre de lexies présentes), enrichit autant de superstructures en créant éventuellement les entrées concernées, implante tous les liens de synonymie toujours en fonction du nombre et du type de lexies présentes, et enfin renvoie le tout muni d'indices de repérage dans le dictionnaire principal. Et ce jusqu'à ce qu'on parvienne au bout de la base de données.

Il faut noter que pour n langues implantées dans le graphe, il faut utiliser les données de factoriel- n dictionnaires. Cela étant, il y a possibilité de transformer les CSMU une fois programmés en dictionnaires bilingues avec n'importe quelle correspondance, en vertu du fait que chaque objet structuré est muni de son étiquette de langue.

5.2.5 Quelques ajouts facilement envisageables

Pour des raisons purement didactiques, la proposition que nous avons faite se résume à un algorithme simple qui tient compte du plus petit nombre de contraintes possible qui pèseraient sur le processus de transfert interlangue. Au demeurant, nous pouvons citer d'ores et déjà quelques points pour lesquels un aménagement immédiat de la base de données et/ou de l'algorithme est facilement envisageable.

Pour commencer en ce qui concerne les données, la solution de rendre communes aux deux langues mises en rapport les informations sémantiques et argumentales n'est évidemment pas suffisante dans l'optique d'un système opérationnel en aide à la traduction. Chacun des critères que nous avons retenus pourrait donner lieu à de nouvelles formes de traitement au sein du tableau de données :

- Étant donné que nous n'avons pas projeté d'associer à nos travaux un analyseur syntaxique pour chaque langue, nous ne remettons pas en cause le dégroupement des colonnes *Constr* en deux, une pour le mot source, une pour le mot cible. Toutefois, la forme initiale de nos cellules ne permet pas à première vue de traiter les constructions vraiment différentes d'une langue à l'autre.
- L'unification des collocations sujet et objet pour les deux langues confrontées pose un petit problème d'ontologie et de traduction à l'intérieur de l'ontologie. En effet, une collocation n'est pas traduite de façon systématique dans les dictionnaires. Encore plus embêtant est la différence de traitement entre les collocations : soit la collocation est ontologique et n'apparaît qu'en langue cible, soit elle est intégrée à la phrase exemple et est visible dans ce cas dans les deux langues. Le

statut de la collocation n'est d'ailleurs pas le même dans ces deux cas de figure ; dans le premier, il y a une insistance sur la présence obligatoire d'une collocation et sur son appartenance nécessaire à un paradigme lexical (non explicite), et dans le second la phrase exemple est là surtout pour préciser l'acception qui donne lieu à la traduction sans qu'il y ait d'insistance réelle sur la collocation elle-même.

- Le registre et les informations complémentaires véhiculent des informations disparates qui n'ont pas toutes le même degré d'appartenance aux deux langues en rapport. Lorsqu'il s'agit de terminologie, technolecte et domaine d'emploi spécifique, l'information est sans problème commune aux deux. En revanche, les parlars nationaux ou régionaux, les niveaux de langue et certains prolongements de sens (souvent propres à une seule des deux lexies) ne peuvent subir l'unification interlangue sans être vérifiées au préalable.

Dans presque tous les cas, nous pouvons envisager de doubler les colonnes afin de faire correspondre à chaque lexie son propre paradigme d'informations linguistiques. Les modifications à reporter dans l'algorithme seraient mineures dans ce cas de figure ; il s'agit avant tout de savoir où vont se placer les informations nouvellement encodées et quels seraient leurs indices. En tout état de cause, l'alourdissement de la base de données peut être tempéré par l'usage de dictionnaires électroniques entièrement balisés, avec des balises de langues susceptibles d'isoler des informations qui ne concernent qu'une seule des langues.

Concernant l'algorithme, quelques ajouts peuvent être effectués rapidement sans gêner la structure que nous lui avons donnée.

Tout d'abord, dans le cadre de la saisie des données relatives à la synonymie intralangue aussi bien qu'à l'analyse linguistique monolingue, il ne faut pas ignorer les cas où l'information mot source n'est pas renseignée dans un certain nombre de lignes, à moins de faire bénéficier à ces lignes d'un traitement à part (sous la forme d'un sous-algorithme qui ne traiteraient que la construction de structures contextuelles et de liens synonymiques en langue source). Les deux solutions sont intégrables à notre algorithme, soit en considérant que nous faisons fonctionner à part la sous-base de données qui ne contient pas de parcours de traduction, soit en intégrant des procédures conditionnelles avant de traiter les *MC*. C'est cette dernière solution que nous allons décrire maintenant.

À l'intérieur de la deuxième boucle de l'algorithme, nous intégrons la condition « si *MC* existe alors nous créons sa structure contextuelle et le(s) liens qui vont avec ». Mais comme les présences de *SMS* et de *SMC* ne sont pas garanties non plus, il faudra répéter plusieurs procédures conditionnelles, « Si *MC* et *SMS* existent... », « si *MC* et *SMC* existent... » et « si *MC*, *SMS* et *SMC* existent... ». Notons que nous pouvons

intégrer le traitement des lignes sans traduction à l'intérieur de la première condition car s'il n'y a pas de *MC*, c'est que nous traitons uniquement un synonyme en langue source (donc un *SMS*) ou une simple structure contextuelle du *MS* (sans renvoi synonymique).

EltconstrMS = (*Constr* = *EXTRACT* (*TAB*[*ligne*,10]), *Collocobjet1* = *EXTRACT* (*TAB* [*ligne*, 7]), *Collocobjet2* = *EXTRACT* (*TAB* [*ligne*, 8]), *Collocsubj* = *EXTRACT* (*TAB* [*ligne*, 9]), *Registre* = *EXTRACT* (*TAB* [*ligne*, 12]), *Infoscompls* = *EXTRACT* (*TAB* [*ligne*, 13]), *IndiceMS*);

MC = *EXTRACT* (*TAB*[*ligne*, 3]);

LC = *EXTRACT* (*TAB*[*ligne*, 4]);

SMC = *EXTRACT* (*TAB*[*ligne*, 6]);

SMS = *EXTRACT* (*TAB*[*ligne*, 5]);

si MC .NE. ^ alors début

IndiceMC = *Numcons*;

Numcons = *Numcons* + 1;

EltconstrMC = (*Constr* = *EXTRACT* (*TAB*[*ligne*,11]), *Collocobjet1* = *EXTRACT* (*TAB* [*ligne*, 7]), *Collocobjet2* = *EXTRACT* (*TAB* [*ligne*, 8]), *Collocsubj* = *EXTRACT* (*TAB* [*ligne*, 9]), *Registre* = *EXTRACT* (*TAB* [*ligne*, 12]), *Infoscompls* = *EXTRACT* (*TAB* [*ligne*, 13]), *IndiceMC*);

EconstrMC = {*EltconstrMC*};

AJOUTLIEN (*Liens*, (*IndiceMS*, *IndiceMC*, « *ST* »));

Objdico = (*Lemme* = *MC*; *Cat* = « *V* »; *Lang* = *LC*; *Constr* = *EconstrMC*);

Dico = *AJOUT* (*Dico*, *Objdico*);

sinon

si SMS .NE. ^ alors début

etc.

sinon

continuer

si SMS .NE. ^ alors début

etc.

5.3 Atouts et perspectives

À l'heure de faire le bilan des recherches que nous avons menées dans le cadre de cette thèse, il est d'abord indispensable de nous pencher sur ce que les modèles et l'algorithme que nous défendons peuvent réellement apporter à la linguistique et en particulier au domaine de la traduction assistée par ordinateur.

Bien que nous soyons parfaitement conscients des manques et insuffisances de notre travail liés à la complexité de chaque langue et plus encore de leur

interconnexion asymétrique, nous avons souhaité montrer qu'un dictionnaire multilingue entièrement paramétré pouvait être généré de manière entièrement automatique par l'intermédiaire de l'algorithme et des améliorations qui pourront lui être apportées.

Notre objectif initial était la représentation de lexiques de plusieurs langues au sein d'un même graphe. Au final, nous proposons une méthodologie qui permet non seulement de construire ce graphe mais également d'aboutir à un mode de représentation des données intermédiaire entre des dictionnaires édités et des lexiques électroniques paramétrables. Les bases de données construites à l'appui de notre méthodologie doivent autoriser aussi bien leur transformation en un format lexicographique traditionnel, constituant en cela une « mémoire de traduction » reproduisant fidèlement et exhaustivement les données de plusieurs dictionnaires de référence, que le passage à un format électronique avec possibilité de générer une aide à la traduction efficace.

Outre son format intermédiaire, ce qui ne signifie pas qu'il n'appartient ni à l'un ni à l'autre des formats papier et électronique, la base de données reste ouverte et ce à deux niveaux : toute nouvelle information peut être ajoutée (ou retirée) à tout moment (ligne horizontale) et un nouveau critère général peut être inséré sans problème (ligne verticale). Dans ce dernier cas, l'apparition d'un critère générique devra être répercutée dans le programme qui viendra à l'appui de notre algorithme, celui-ci étant conçu pour être le plus évolutif possible.

Les deux plus grands atouts de notre travail sont à notre avis la simplicité et la malléabilité. En effet, il n'y a rien de plus trivial qu'une base de données ordonnée qui intègre l'ensemble des lexies de toutes les langues ; l'algorithme lui-même répercute cette simplicité apparente dans son mode de fonctionnement, en exploitant les données ligne par ligne sans se soucier de subtilités - pour l'instant - qui ne rentreraient pas dans les cases prévues. Du côté de la malléabilité, nous avons déjà insisté sur le fait que la base et l'algorithme sont tous les deux conçus pour recevoir des critères et des aménagements ultérieurs. Bien entendu, il faudra veiller à l'heure de programmer que nos velléités d'amélioration constante ne se heurtent pas à la complexité de la programmation et ne mènent pas à la pratique du « pansement » qui consiste à bidouiller un sous-programme à l'endroit adéquat plutôt que d'intégrer les nouvelles informations au bon endroit. Compte tenu de la forme actuelle de notre algorithme, ce dernier point ne suscite pas chez nous de réelle inquiétude.

CHAPITRE V

Conclusion
Bilan final, Perspectives et Obstacles

1. Rappel des objectifs et des résultats

Pour atteindre l'objectif que nous nous étions fixé au départ, à savoir proposer une méthodologie permettant de transformer des bases de données bilingues en un graphe multilingue informatisé, nous avons effectué quatre étapes principales :

- transformation des données issues de dictionnaires bilingues et de dictionnaires de synonymes en une base de données multilingue ;
- prise en compte des données issues de corpus et réflexion sur la façon d'extraire et de représenter les informations susceptibles d'avoir de l'influence sur une traduction éventuelle ;
- réunion dans la mesure du possible des deux ensembles de données, lexicographiques et littéraires, au sein d'une base unique ;
- écriture d'un algorithme pour un programme informatique capable de transformer la base de données finale en un graphe multilingue complet où tous les liens lexicaux présents dans la base sont reproduits. Ce programme doit servir de base au passage du graphe ou de la BD au format lexicographique traditionnel ou au contraire au format électronique paramétrable.

Le principal attrait de nos modèles est double : réunir toutes les informations entre langues dans un même modèle de représentation et induire une « collaboration tacite » de la part de l'utilisateur d'un tel lexique, en lui proposant automatiquement et de manière ordonnée toutes les traductions envisageables dans un contexte précis.

Notre problématique d'origine ne comprenait que l'exploitation des dictionnaires édités mais nous avons rapidement pris conscience que la réalité de la langue en contexte n'en dépendait guère autrement qu'en s'en servant comme point de départ à la création originale de sens et de structures, d'autant que les dictionnaires ne sont que le reflet de la langue et ne précèdent pour ainsi dire jamais l'apparition de nouveaux sens. Sur ce dernier point, il est manifeste que les auteurs littéraires, et dans une moindre mesure les journalistes, sont les plus grands pourvoyeurs publics de sens nouveaux généralement métaphoriques, ou de constructions syntaxiques nouvelles qui s'adaptent à des sens déjà connus. Notre base de données lexicale intègre ces dernières sans adaptation nécessaire dans la mesure où une lexie est implantée avec toutes les informations qu'elle véhicule explicitement mais ne nécessite pas qu'on lui associe d'emblée une ou plusieurs traductions dans une langue cible. Ce dernier travail peut être purement collaboratif grâce à la participation et à la validation des utilisateurs à partir de leur propre connaissance de la langue.

D'un point de vue pratique, la programmation d'un système de transformation de la base de données en graphe synonymique ne pose pas de problème (cf. algorithme). Nous pouvons annoncer que le programme de création de graphes a été écrit à l'ATILF par Étienne Petitjean directement à partir de notre algorithme et qu'il est d'ores et déjà opérationnel pour des expérimentations futures.

Au-delà de ce résultat concret, il nous reste à préciser un certain nombre de points dans l'optique de poursuivre le projet, et en particulier nous devons nous pencher sur la question des liens indirects qui ne sauraient être générés automatiquement sans l'aval d'un utilisateur expérimenté. C'est donc la principale difficulté de notre modélisation : comment qualifier et caractériser les liens de synonymie (ou autres) qui ne sont pas directement apparents dans le graphe ? Quels sont les moyens à mettre en œuvre pour activer les liens invisibles ?

À ces questions nous n'avons répondu que partiellement car nous n'avons pas les outils théoriques ni la visibilité dont nous aurions besoin pour une application de qualité constante. Toutes les expérimentations de liens indirects que nous avons tentées mènent à une seule et même conclusion : tout lien qui ne serait pas directement perceptible par les dictionnaires ou les corpus doit nécessairement être validé par un utilisateur avisé. En l'état actuel des choses, les fameux liens invisibles ne pourraient être évalués qu'au cas par cas, ce que nous avons refusé de faire compte tenu du temps que prennent et qu'ont pris déjà les autres analyses linguistiques et la création de nos bases. Cet aveu d'échec provisoire n'exclut pas néanmoins que nous reprenions la problématique pour un prolongement utile et efficace de nos graphes, bien au contraire, d'autant que cette réflexion sera de toute façon indispensable si nous avons la chance de continuer cette recherche.

2. Contribution pour le dictionnaire du XXIème siècle

Lorsque nous songeons au lien qui unit dictionnaire et graphe, nous pensons qu'un dictionnaire contient toutes les données nécessaires pour construire un graphe sémantique. Les CSMU dont nous avons posé les principes sont à l'origine entièrement conçus à partir du matériau lexicographique. Aucune donnée utile à la discrimination lexicale n'est ignorée, de même qu'aucune information superflue n'est initialement présente. Les CSMU se posent donc d'abord en reflet absolument fidèle d'un outil textuel préexistant à sa création, hors de tout arbitraire de la part du programmeur.

Comme nous le suggérons en sous-titre, un dictionnaire contient toutes les données nécessaires pour construire un graphe sémantique. Certes, la formule fonctionnerait tout aussi bien pour dire qu'un dictionnaire de correspondance directe, comme un dictionnaire de synonymes non commenté, se suffit déjà à lui-même dans la perspective de bâtir un graphe. Ce que nous voulons vraiment exprimer est qu'il ne devrait pas y avoir de limitation dans l'implantation des critères destinés à faciliter ou à expliciter le passage d'une langue à une autre. Autrement dit, dès lors qu'une

information linguistique est jugée pertinente en traduction, nous devrions pouvoir l'exploiter et la représenter. C'est tout l'enjeu de la collaboration entre la linguistique de corpus que nous avons mise en avant et la lexicographie éditée. Au-delà de ce simple constat, nous devons également songer aux compléments à apporter aux CSMU pour englober un dictionnaire dans son intégralité. Les données que nous n'avons pas prises en compte, mais qui figurent dans les dictionnaires bilingues en tant qu'informations pertinentes, peuvent être facilement ajoutées à nos bases de données moyennant l'insertion de nouvelles colonnes. Il s'agit notamment de la forme phonétique de l'entrée (dans le Robert & Collins), de certains indicateurs de dégroupement neutralisés par la forme linéaire et séquentielle de nos bases, etc. Rien n'interdit de les intégrer à une base de données complète et ensuite de moduler le programme d'extraction des données selon les besoins. Ce que nous envisagerions dans l'idéal serait l'écriture de plusieurs programmes parfaitement distincts à partir du même algorithme modulable, un programme pour la création des CSMU tel que nous l'avons annoncé, un programme pour simplement extraire un fragment des CSMU (et dans ce cas paramétrable « à la main » grâce à des menus informatisés) et également un programme qui prenne la totalité des informations de la base de données pour en faire un dictionnaire absolument complet sans filtre de la connaissance déjà saisie.

La base de données est ainsi à prendre comme le résultat de l'algorithme appliqué au dictionnaire. En étendant avec enthousiasme la portée de notre algorithme de création de graphes, rien n'empêche d'imaginer que celui-ci sera capable de transformer directement le dictionnaire (dans sa forme électronique) en cette même base. C'est d'ailleurs la motivation profonde du travail que nous défendons.

Cela suppose évidemment que l'on puisse disposer d'un dictionnaire entièrement électronique, et balisé qui plus est. La reconnaissance de chaque balise, correspondant à un type d'informations et un seul, peut être aisément repérée et intégrée à sa place dans le tableur. D'une certaine façon, l'algorithme que nous avons écrit est un baliseur mais ce n'est pas suffisant. Ce que nous souhaitons, mais nous ne pouvons que l'appeler de nos vœux, est que nous puissions exploiter des dictionnaires déjà balisés pour les transformer automatiquement en base de données. Si une telle chose était possible, nous nous ferions une joie d'écrire l'algorithme préalable à l'application de notre algorithme principal.

Malheureusement, il est fort peu probable que nous aurons un jour l'accès aux sources et aux balises cachées des dictionnaires commerciaux à moins d'envisager une collaboration directe avec les éditeurs. Du coup, notre travail s'inscrit avant tout dans une perspective purement méthodologique et verra ses premiers aboutissements et sa première confirmation dans l'expérimentation manuelle, c'est-à-dire en écrivant nous-même notre base. Pour l'instant, plusieurs milliers de lexies y ont été implantées par nos soins mais elles ne représentent qu'une infime partie de tout ce qui est présent dans un seul dictionnaire.

Toutefois, la solution de baliser nous-même un dictionnaire numérisé demeure envisageable dans la mesure où les formats typographiques diffèrent selon ce que les

champs désignent à l'intérieur de l'ouvrage. Hélas, dans certaines éditions, l'identification des champs est rigoureusement impossible de manière automatique car les formats typographiques ne sont pas clairement différenciés (cas des éditions Larousse et Bordas évoqués au chapitre II). En ce qui concerne le Robert & Collins, ce travail d'automatisation en aval est peut-être plus aisé mais se heurtera néanmoins à la concurrence occasionnelle de plusieurs critères pour une même typographie. Par conséquent, l'expérience mérite d'être tentée mais nous n'espérons pas une automatisation intégrale de ce type de balisage par nos soins, à tout le moins pas sans une vérification systématique des champs concurrents. Il reste que l'opération ne pose aucun problème sur le plan théorique : tout dictionnaire entièrement balisé est directement exploitable moyennant l'algorithme adapté, et tout dictionnaire n'utilisant qu'une seule typographie par champ peut être facilement balisé par un simple programme de conversion.

Dans la lignée des lexiques miroirs dont nous avons imaginé la représentation et décrit le fonctionnement, partons du principe que nous sommes capables de générer entièrement une base de données lexicale et son graphe de représentation à partir d'un simple dictionnaire, sans autre outillage que l'algorithme de création des CSMU. Par extension, rien ne semble s'opposer à ce que cet algorithme soit inversé pour générer un dictionnaire complet à partir des bases. Toujours en partant de l'hypothèse que les CSMU pourraient être conçues à partir d'un matériau textuel balisé, les graphes que nous envisageons garderaient, dans les faits ou potentiellement, la trace de ces balises de sorte que leur restitution inverse soit immédiate.

Encore une fois, c'est le principe du miroir que nous appliquons pour retourner au matériau d'origine, le dictionnaire en tant que tel. Dans la mesure où nous intégrons des données de sources multiples, le dictionnaire restitué serait donc d'une richesse très supérieure à ce que chacun recelait au départ. Le procédé d'inversion de l'algorithme permettant de restituer, à l'identique ou de manière fusionnée, les données présentes dans un dictionnaire bilingue d'origine, nous pouvons ainsi imaginer ce que cela pourrait apporter à la lexicographie en gain de temps et de lisibilité.

Le lexique miroir et son corollaire base de données présentent la particularité de proposer la fusion des deux parties d'un dictionnaire bilingue sans toutefois l'imposer ; tout dépend de la partie de l'algorithme qui est activée. Ce dernier point présuppose que l'algorithme fût complété par cette potentialité de fusion des données, chose assez facile à ajouter en désorientant le graphe, comme nous l'avons déjà expliqué.

L'un des avantages des bases issues du dictionnaire est de pouvoir être enrichies, appauvries ou modifiées à loisir. Par l'évolution de la langue ou par décision des lexicologues, toute information lacunaire ou obsolète peut à tout instant être

ajoutée ou ôtée. Conséquemment, le nouveau graphe est recalculé par le passage du même algorithme de construction sur la base modifiée.

On imagine sans peine ce que cette méthode pourrait apporter en gain de temps aux rédacteurs de dictionnaires : un ouvrage entier sous la forme d'un ensemble de lignes, chacune de celles-ci correspondant à une lexie ou plus précisément à un emploi précis donnant lieu à une traduction spécifique. Dès lors, si la base est suffisamment complète pour embrasser tout l'ouvrage lexicographique, sa réédition serait la simple réévaluation du graphe à partir des modifications qu'on aura fait subir à celle-là.

Une fois encore, nous nous heurtons à l'opacité parfaitement compréhensible des concepteurs de dictionnaires industriels en la matière. En effet, il est tout à fait possible qu'une partie de notre méthodologie, et en particulier de celle qui sous-tend la création des CSMU, fût déjà exploitée par les maisons d'édition. Mais sur ce point, nous ne pouvons hélas que supputer.

Nous venons de le dire, il est difficile d'arriver à savoir comment travaillent les rédacteurs à visée commerciale ; on n'offre pas facilement ses méthodes de travail en ce milieu. Il nous paraît néanmoins probable que les données sont stockées numériquement d'une année sur l'autre et que les modifications que les lexicographes leur font subir sont directement transmises sous la forme informatique.

Il nous semble également plausible qu'aucune base linéaire telle que la notre n'existe pas dans le monde industriel, hormis l'exemple remarquable du DI de Dominique Dutoit (cf. chapitre I) dont le principe de réévaluation-algorithme-réévaluation-etc. est nettement conforme à notre orientation de recherche, même si nous essayons de faire intervenir notre arbitraire le moins souvent possible.

Devant toutes ces hypothèses sur les méthodes de travail de chacun et en l'absence de visibilité en ce domaine, nous ne pouvons qu'espérer que le travail que nous avons effectué est simplement original à défaut d'être directement exploitable, et qu'il sera possible d'en tenter l'application pour mesurer son impact sur le dictionnaire idéal dont nous rêvons tous, celui qui contient tout et qui répond à toutes les questions. Bien sûr, compte tenu de l'évolution rapide de toutes les langues, cela n'est pas gagné....

3. Des difficultés en grand nombre

La polysémie, la synonymie, la terminologie lexicologique et toutes les notions que nous avons traitées au cours de nos travaux n'ont pas eu la « vie facile » et ne nous l'ont pas rendu de même. De toute évidence, nous ne parvenons à un consensus terminologique qu'à partir du moment où nous figeons toutes les notions à l'intérieur de la méthodologie de construction des CSMU.

Si elle n'a pas ou n'avait pas déjà été effectuée, l'entreprise de saisie des données intégrales d'un dictionnaire représente un travail monumental que nous

sommes très loin d'avoir pu mener à terme au-delà d'un seul lexème, lequel correspond déjà à plusieurs milliers de lignes (pour *abandonner*). Et encore, nous nous sommes contenté d'en proposer la sphère d'influence sémantique en restreignant autant que faire se peut l'ensemble des composantes de son graphe dans la limite des contraintes que nous avons établies au cours de ce travail.

La tentation d'aller explorer les bases littéraires pour en extraire du jus d'acceptions nouvelles a entraîné ce que nous appelons la « double peur du vide ».

Tout d'abord, s'intéresser au contexte, et donc au corpus, est une manifestation de l'angoisse qui nous anime face aux simples données non actualisées. La question est largement débattue depuis plusieurs années et la polémique ne s'est pas encore essoufflée.

Ensuite, une fois que la première peur a été surmontée par l'observation des données attestées, une prise en compte large et systématique de ces données se heurte inmanquablement à l'écueil de leur exploitation par une application informatisée. Il existe deux principaux obstacles relativement faciles à repérer, beaucoup plus malaisés à dépasser. En premier lieu, la créativité langagière des auteurs dépasse de loin ce que les dictionnaires proposent. Il est à noter que ces derniers ne font jamais de prospection, se contentant par nécessité de refléter l'usage. L'émergence très récente de dictionnaires plus « personnels » d'auteurs littéraires, de linguistes ou d'amuseurs tient plus du divertissement que d'une véritable volonté de renseigner sur la langue. Nous pouvons rappeler à ce sujet que Grevisse en fait une critique féroce dans son manuel. D'autre part, il n'est plus possible d'ignorer la créativité massive issue des données de la toile Internet, laquelle subit des processus de création certes anarchiques mais tout aussi « réels » que tout ce que nous avons observé dans d'autres ensembles de données. Ensuite, le décalage entre le sens encodé et l'interprétation est parfois difficile à surmonter. En effet, il nous est arrivé à plusieurs reprises de ne pas savoir comment interpréter un énoncé tiré de Frantext, comme par exemple certains exemples de métaphore mentale avec *traverser*, sujet bien épineux effleuré au chapitre III. En l'absence de l'indispensable commentaire de l'auteur sur ses intentions lexicosémantiques, il n'est pas possible de faire à sa place le travail de désignation d'une nouvelle acception.

De ce point de vue, l'arbitraire des exemples construits peut rassurer le linguiste lexicologue car celui-ci est sûr du sens des énoncés qu'il a lui-même produits avec une finalité précise. Il lui reste simplement à convaincre le reste de la communauté du bien-fondé de ses créations.

Il existe des formes langagières qui ne sont pas des verbes au sens grammatical mais qui fonctionnent comme des prédicats verbaux d'un point de vue sémantique. Le cas des locutions verbales a déjà été traité dans cette thèse, nous n'y reviendrons pas. En revanche, nous avons ignoré les cas des substantifs déverbaux dont le discours est particulièrement riche. En marge de l'analyse des verbes de parole que nous avons

tentée au sein du PDR ILF « Transitivité », nous n'avons pas exploré la productivité des substantifs qui jouent le rôle de prédicats de production de parole. Cela n'empêche pas ces noms d'avoir une productivité sémantique et structurelle comparable à celle de leurs équivalents verbaux et nous ne pourrions pas les ignorer bien longtemps si nous poursuivons notre travail au-delà de ce que nous avons déjà accompli.

4. Conclusion finale

Nous n'avons parlé dans cette thèse que de la catégorie verbale. Il est bien entendu qu'un dictionnaire électronique complet ne saurait contenir que les éléments de cette catégorie sans présenter de fâcheuses lacunes. C'est pourquoi la méthodologie de création des CSMU devra être augmentée et enrichie par l'étude des autres catégories grammaticales, chacune d'entre elles avec les caractéristiques qui lui sont propres. En plus de tous les prolongements que nous entrevoyons à nos recherches en cours, ce dernier point est sans doute le plus important si nous assumons notre désir de participer à l'élaboration d'un véritable « dictionnaire du futur ».

De nombreux facteurs de détermination des emplois et des traductions sont nécessaires et ont été volontairement simplifiés dans notre travail, à l'instar de la nécessaire vulgarisation induite par l'écriture des dictionnaires de langues. Rien ne nous oblige cependant à pousser cette vulgarisation jusqu'à l'ignorance volontaire de tel ou tel critère linguistique apparemment peu pertinent. Toujours dans l'espoir de contribuer à bâtir un dictionnaire idéal, nous espérons parvenir à la modélisation du plus grand nombre de critères qu'il se pourra désigner.

Enfin, comme promis en introduction à ce document, nous rappelons toute l'importance que nous attachons à la vision globale du travail linguistique et en particulier pour ce qui concerne la sémantique lexicale : dictionnaires et corpus doivent pouvoir cohabiter dans le vaste monde des lexiques informatisés. La généralisation de notre méthodologie à toute source de données lexicales est dorénavant l'objectif majeur que nous nous fixons pour la poursuite du projet.

Glossaire

Glossaire des acronymes

CSMU : Champs Sémantiques Multilingues Unifiés. Définition pages 230 et suivantes.

DI : Dictionnaire Intégral. Définition page 74.

LC : Langue Cible

LD : Lexique Droit. Définition page 11.

LI : Lexique(s) Inversé(s). Définition page 11 et 98.

LS : Langue Source

MC : Mot Cible

MS : Mot Source

SA : Synonymie Absolue. Définition page 244.

SD : Synonymie Directe. Définition page 242.

SI : Synonymie Indirecte. Définition page 244.

SMC : Synonyme du Mot Cible

SMS : Synonyme du Mot Source

SPS : Synonymie Partielle en langue Source. Définition page 243.

SPTC : Synonymie Partielle de Traduction en langue Cible. Définition page 243.

ST : Synonymie de Traduction. Définition page 242.

Glossaire Lexical restreint

Lexie. Définition page 55.

Polysémie absolue. Définition page 38-41.

Polysémie contextuelle. Définition page 40.

Structure contextuelle. Définition page 259.

Superstructure. Définition page 262.

Bibliographie

Articles et Ouvrages

ALBERTAZZI L. ed., 2000. *Meaning and cognition*. John Benjamins Company.

BALTAZARD D., TRIGANO P., 1993. "Des homographes aux références pronominales". Thème : III (Langage Naturel).

BARON I., 2000. « La lexicalisation des noms composés déverbaux. Traits généraux de la composition nominale ». In *Argomenti per una Linguistica della Traduzione*, a cura di Iorn Korzen e Carla Marellò. Edizioni dell'Orso, pp. 43-52

BASSAC C., BOUILLON P., 2002. « middle transitive alternation in English: a generative lexicon approach », in *Many Morphologies*, Paul Boucher ed. Cascadilla press, pp. 29-48.

BATEMAN J.A., 1992. *Towards Meaning-Based Machine Translation: using abstractions from text generation for preserving meaning*. Kittredge.

BLANCO X., 2001. « Dictionnaires électroniques et traduction automatique espagnol-français ». *Langages n°143, septembre 2001*, « *Lexicologie contrastive espagnol-français* », pp.1-21.

BLUMENTHAL P., 2006. « De la logique des mots à l'analyse de la synonymie », in *Langue Française n°150, juin 2006.*, pp.14-31.

BLUMENTHAL P., BERNARD P., 2005. « Quels corpus pour un dictionnaire des collocations ? », in *La linguistique de corpus*, G. Williams éd. Linguistiques, Presses Universitaires de Rennes, pp.233-243.

BORILLO A., 2001. « Il y a prépositions et prépositions », in « la préposition », *Travaux de Linguistique n°42-43, Revue Internationale de Linguistique Française*. Duculot, pp.141-155.

BOUILLON P., 1998. *Polymorphie et sémantique lexicale : le cas des adjectifs*. Thèse de troisième cycle, Lille. Presses Universitaires du Septentrion.

BOUILLON P., CLAS A., 1993. *La traductique*. Les presses de l'université de Montréal, AUPELF UREF.

BREAL M., 1897. *Essai de sémantique (science des significations)*. Librairie Hachette et Cie : Paris.

BRONCKART J-P., 2000. « Possibles de l'activité langagière et impossibles de la langue », in *Incidences de l'Impossible dans le Langage, revue Sémiotiques, Décembre 2000, n°18/19*. CNRS-ILF, pp.153-172.

CADIOT P., 1999. « Espaces et prépositions », in *Revue de Sémantique et Pragmatique. 1999. Numéro 6*. Presses Universitaires d'Orléans, pp. 43-70.

CADIOT P., HABERT B., 1997. « Aux sources de la polysémie nominale », in *Langue française, n°113*, pp. 3-11.

CERVONI J., 1991. *La préposition. Étude sémantique et pragmatique*. Champs linguistiques, Duculot.

CHAROLLES M., LAMIROY B., 2002. « 'zézayer', 'dire', 'parler'. Les verbes de parole ou qu'est-ce qu'un verbe (in)transitif », *travaux préliminaires au projet ILF « Transitivity »*. Paris.

CHOMSKY N., 1957. *Syntactic structures*. Mouton.

CORBIN D., CORBIN P., TUTIN A., ALIQUOT S., 1996. « Ce que les linguistes peuvent attendre d'un dictionnaire informatisé », in *Lexicographie et informatique. Autour de l'informatisation du Trésor de la Langue Française. Actes du Colloque International de Nancy (29,30 et 31 mai 1995)*, publiés par David PIOTROWSKI. Didier Érudition, Paris, pp.51-77.

CORTIER C., 2001. « Les syntagmes prépositionnels prédicatifs dans les grammaires universitaires : un observatoire de la place accordée aux prépositions », in « la préposition », *Travaux de Linguistique n°42-43, Revue Internationale de Linguistique Française*. Duculot, pp.121-140.

COSÉRIU E., 1980. « Aspect verbal ou aspects verbaux ? Quelques questions de théorie et de méthode », in : *La notion d'aspect, actes du colloque organisé par le Centre d'analyse syntaxique de l'université de Metz, sous la dir. de David J., Martin R.*, Paris : Klincksieck, pp. 13-25.

COULON D., KAYSER D., 1982. « Les sens uniques conduisent à des impasses ». *1^{er} Colloque de l'A.R.C.*, Pont-à-Mousson, Actes pp.1-32.

CROFT W., CRUSE D.A., 2004. *Cognitive linguistics*. Cambridge University Press.

CRUSE D.A., 1986. *Lexical Semantics*. « Cambridge textbooks in linguistics ». Cambridge University Press.

DANCETTE J., 1995. *Parcours de traduction. Étude expérimentale du processus de compréhension*. « Étude de la traduction ». Presses Universitaires de Lille.

DANLOS L., 1999. « Sur la coréférence événementielle ». Actes des VIèmes journées scientifiques du Réseau LTT, Beyrouth, Novembre 1999.

DE VOGÛE S., PAILLARD D., 1997. « Identité lexicale et hétérogénéité de la variation co-textuelle : le cas de *suivre* », in *Co-texte et calcul du sens*, Claude Guimier éd., Presses Universitaires de Caen, pp.41-61

DENIS P., MULLER P., 2004. “A semantics for temporally dependent referring expressions”, in O. Bonami and P. Cabredo Hofherr, editors, *Empirical Issues in Syntax and Semantics 5*, Papers from Cssp 2003, pp.45-62.

DESCLES J-P., 1997. « Systèmes d’exploration contextuelle », in *Co-texte et calcul du sens*, Claude Guimier éd., Presses Universitaires de Caen, pp.215-232.

DOMINICY M., MARTIN F., 2001. “A travers, au travers (de) et le point de vue », in « la préposition », *Travaux de Linguistique n°42-43, Revue Internationale de Linguistique Française*. Duculot, pp.211-227.

DRIX G., 1993. *Langage C norme ANSI, vers une approche orientée objet*. 2^{ème} édition. Masson, Paris.

DUBOIS J., DUBOIS C., 1971. *Introduction à la lexicographie : le dictionnaire*. « Langue et langage ». Larousse.

DUBOIS U., 1997. « Co-texte et calcul du sens dans le cas de quelques valeurs de *do* et du passé composé », in *Co-texte et calcul du sens*, Claude Guimier éd., Presses Universitaires de Caen, pp.157-163.

DUGAS A., 2000. *Le verbe et ses prépositions*. Les éditions LOGIQUES.

DUGAS A., 2001. « Une analyse des constructions transitives indirectes en français », in « La Préposition », *Travaux de Linguistique n°42-43, Revue Internationale de Linguistique Française*. Duculot, pp.111-120.

DUTOIT D., 2000. *Quelques opérations Sens → Texte et Texte → Sens utilisant une Sémantique Linguistique Universaliste a priori*. Thèse soutenue à l'Université de Caen, spécialité Informatique Linguistique.

EINSTEIN A., 1994. « Sur l'électrodynamique des corps en mouvement, et 6 autres textes fondamentaux sur la Théorie de la Relativité restreinte et générale ». Traduction de M.Solovine, J. Gabay.

FANG A.C., 1997. "Verb Forms and Sub-Categorizations". *Literary and Linguistic Computing*, Vol. 12, n°4. Oxford University Press, pp.209-217.

FAUCONNIER G., 1984. *Espaces mentaux. Aspects de la construction du sens dans les langues naturelles*. Propositions. Les Éditions de Minuit.

FINDLER N.V. ed., 1979. *Associative networks. Representation and use of knowledge by computers*. Academic Press Inc. London.

FRADIN B., 1988. « Décrire un verbe », in *Lexique*, 6, *Lexique et Paraphrase*, pp.87-138.

FRANÇOIS J., 1989. *Changement, causation, action. Trois catégories sémantiques fondamentales du lexique verbal français-allemand*. Librairie Droz.

FRANÇOIS J., 1997a. « Le cadrage cognitif des prédications de contact dans un corpus de déclarations d'accidents de la route : effets du contexte et du co-texte », in *Co-texte et calcul du sens*, Claude Guimier éd., Presses Universitaires de Caen, pp.73-88.

FRANÇOIS J., 1999. « Les caractères aspectuels et participatifs des prédications verbales et la transitivité », *Bulletin de la Société Linguistique de Paris*, XCIV, fasc. 1, pp.139-184.

FRANÇOIS J., DENHIÈRE G. eds, 1997. *Sémantique linguistique et psychologie cognitive. Aspects théoriques et expérimentaux*, Presses Universitaires de Grenoble.

FRANÇOIS J., MANGUIN J-L., 2006. « *Dispute théologique, discussion oiseuse et conversation téléphonique* : Les collocations adjectivo-nominales au cœur du débat », in *Langue Française n°150, juin 2006.*, pp.50-65.

Frantext. Base littéraire du français, ATILF-CNRS laboratoire, Nancy, France.

FUCHS C., 1994. *Paraphrase et énonciation*. Paris, Ophrys.

FUCHS C., 1996. *Les ambiguïtés du français*. Paris, Ophrys.

FUCHS C., 1997. « La synonymie en co-texte », in *Co-texte et calcul du sens*, Claude Guimier éd., Presses Universitaires de Caen, pp.31-39.

GARDES-TAMINE J., 1990. *La Grammaire. 2/Syntaxe*. Cursus, Armand Colin.

GEVAUDAN P., 1997. « La polysémie verticale : Hypothèses, analyses et interprétations ». *PhiN. Philologie im Netz*2/1997, pp.1-22.

GODARD D., JAYEZ J., 1993. « Le traitement lexical de la coercion », *Cahiers de linguistique française*, n° 14, pp.123-150.

GODARD D., JAYEZ J., 1995. "Principles as lexical methods". *Proceedings of the AAAI Workshop on Representation and Acquisition of Lexical Knowledge, Stanford*, pp.57-68.

GODARD D., JAYEZ J., 1996. « Types nominaux et anaphores : le cas des objets et des événements ». *Cahiers Chronos 1, Anaphores temporelles et (in-)cohérence*. Textes réunis par W. De Mulder, L. Tasmowski-De Ryck et C. Vettters. Rodopi, pp.41-58.

GOSELIN L., 1996a. « Le traitement de la polysémie contextuelle dans le calcul sémantique ». *Intellectica*, 1996/1, 22, pp.93-117.

GOSELIN L., 1996b. *Sémantique de la temporalité en français. Un modèle calculatoire et cognitif du temps et de l'aspect*, Duculot, Louvain-la-Neuve.

GOSELIN L., 1997. « Les études psycholinguistiques sur la compréhension des expressions ambiguës : une critique linguistique », in *Sémantique linguistique et psychologie cognitive. Aspects théoriques et expérimentaux*, Jacques François et Guy Denhière eds., Presses Universitaires de Grenoble, pp.75-115.

GOSELIN L., 2001. « Temps et aspect en français ». *Programme de l'École thématique CNRS / Réseau de sciences cognitives d'Île-de-France*.

GREVISSE M., GOOSSE A., 1997. *Le bon usage : grammaire française. Treizième édition*. Duculot, Louvain-la-Neuve.

GROSS G., 1996a. *Les expressions figées en français. Noms composés et autres locutions*. Collection l'Essentiel Français. Ophrys.

GROSS G., 1996b. «Rendre les dictionnaires actifs» in *Lexicographie et informatique. Autour de l'informatisation du TLF*. Actes du Colloque International de Nancy (29, 30 et 31mai 1995) publiés par David PIOTROWSKI. Didier Érudition, Paris, pp.195-212.

GROSS M., 1975. *Méthodes en syntaxe. Régime des constructions complétives*. Hermann.

GROSS M., 1994. « Dictionnaires électroniques et traduction automatique », in *Le traducteur et l'ordinateur*. Langages n°116, Décembre 94. Larousse, pp.48-58.

GUILLAUME G., 1964. *Langage et science du langage*. Paris, Nizet.

GUIMIER C. éd, 1997. *Co-texte et calcul du sens*. Presses Universitaires de Caen.

GUIRAUD P., 1962. *Les locutions françaises*. Collection Que sais-je ? Presses Universitaires de France.

HANKS P., 1990. « Evidence and intuition in lexicography », in *Meaning and Lexicography*. LLSEE 28, J. Tomaszczyk & B. Lewandowska-Tomaszczyk eds. John Benjamins, pp.31-41.

HATON S., 1999. « Analyse et modélisation des sens métaphoriques des verbes de mouvement en français ». Mémoire de maîtrise de sciences du langage, option TAL, Université Nancy2.

HATON S., 2000. « Typologie des expressions anaphoriques et/ou coréférentielles mettant en jeu un ou plusieurs syntagmes verbaux », mémoire principal de DEA sciences du langage, Université Nancy2.

HATON S., 2003. « Les Champs Sémantiques Multilingues Unifiés ». Actes du congrès international *Représentation du Sens*, Montréal.

HATON S., 2004. « Sens, polysémie et multilinguisme : comment générer des champs synonymiques à partir de dictionnaires de langues ». Actes du colloque JETOU2003 *autour du sens*, Toulouse.

HATON S., 2005a. « L'intégration des périphrases verbales dans les Champs Sémantiques Multilingues Unifiés : étude autour de la périphrase *se mettre à* ». Actes du congrès international *Les Périphrases Verbales* à Caen. John Benjamins, pp.397-406.

HATON S. 2005b. « Transitivity, constructions prototypical and creativity in language : comment un corpus littéraire réorganise la typologie des verbes de parole en français ». Actes du colloque JETOU2005, *autour du corpus*, Toulouse, pp.151-164.

HATON S., GRANDJEAN M., 1998. « La grammaire TESP8 : construction d'un analyseur syntaxique à partir de l'analyseur CAT2 ». Rapport de licence, cours d'analyse syntaxique automatique par Mounira LOUGHRAÏEB, université de Nancy2.

HATON S., MASCHERIN L., 2005. « A cross-languages study about Space and Time : the french verb *traverser* and its equivalents in english ». Actes du congrès international STLL05, Rijeka, Croatie. Zadar Publications. À paraître.

HAUSMANN F.J., BLUMENTHAL P., 2006. « Présentation : collocations, corpus, dictionnaires », in *Langue Française n°150*, juin 2006., pp.3-13.

HEINZ M., 1993. *Les locutions figées dans le « Petit Robert ». Description critique de leur traitement et propositions de normalisation*. LEXICOGRAPHICA Series Maior 49. Niemeyer.

HELMREICH S., JIN W., WILKS Y., GUILLEN R., 1993. « Questions de traduction automatique au Computing Research Laboratory (CRL) », in *La traductique*, Bouillon et Clas eds. Les presses de l'université de Montréal, AUPELF UREF, pp.334-348.

HIRST G., 1987. *Semantic interpretation and the resolution of ambiguity*. Studies in Natural Language Processing. Cambridge UP.

HÜLLEN W., SCHULZE R. eds, 1988. *Understanding the lexicon. Meaning, sense and world knowledge in lexical semantics*. Linguistische Arbeiten. Niemeyer.

HUTCHINS W.J., 1986. *Machine translation. Past, present, future*. Ellis Horwood.

JACQUES M-P., 2005. « Pourquoi une linguistique de corpus ? », in *La linguistique de corpus*, G. Williams éd. Linguistiques, Presses Universitaires de Rennes, pp.21-30.

JACQUEY E., 2001. *Ambiguïtés lexicales et Traitement Automatique des Langues : Modélisation de la polysémie logique et application aux déverbaux d'action ambigus en français*. Thèse soutenue près l'Université Nancy2 le 13 Décembre 2001.

JAMROZIK E., 1992. *La syntaxe et la sémantique des verbes de parole en français*. Thèse soutenue à l'Université de Warszawa, Pologne.

JI H., PLOUX S. & WEHRLI E., 2003. "Lexical Knowledge Representation with Contextonyms". In *Proceedings of the 9th MT summit*, pp.194-201.

KAROLAK, S. 2003. 'Réflexion sur la catégorie de l'aspect'. *Lingua Posnaniensis* N°45: pp. 151-175.

KAROV Y., EDELMAN S., 1996. "Learning similarity-based word sense disambiguation from sparse data". *Cmp-lg/9605009*, pp.42-55.

KILGARRIFF A., 1997. « I don't believe in Word Senses », *Computers and the Humanities*, vol. 31, n°2. Kluwer Academic Publishers, pp.91-113

KLEIBER G., 1987. *Du côté de la référence verbale. Les phrases habituelles*. Collection sciences pour la communication. Peter Lang.

KLEIBER G., 1990. *La sémantique du prototype. Catégories et sens lexical*. Linguistique Nouvelle. PUF.

KLEIBER G., 1994. « Discours et stéréotypie : le contexte peut-il remettre d'aplomb une anaphore associative mal formée ? », in C. Schnedecker, M. Charolles, G. Kleiber et J. David (éd.), *L'anaphore associative. Aspects linguistiques, psycholinguistiques et automatiques*, Paris, Klincksieck, pp.93-112.

KLEIBER G., 1997. « Quand le contexte va, tout va et ... inversement », in *Co-texte et calcul du sens*, Claude Guimier éd., Presses Universitaires de Caen, pp.11-29.

KLEIBER G., 1999. *Problèmes de sémantique. La polysémie en questions*. Sens et structures. Septentrion, presses universitaires.

KOZLOWSKA, M. 1998a. 'Aspect, modes d'action et classes aspectuelles'. In J. Moeschler (éd), *Le temps des événements*, Paris: Kimé. pp. 101-121.

KOZLOWSKA, M. 1998b. 'Bornage, télicité et ordre temporel'. In J. Moeschler (éd), *Le temps des événements*, Paris: Kimé. pp. 221-244.

KRAIF O., 1997. « Modèles probabilistes pour le traitement automatique de corpus textuel : perspectives et applications », in *Travaux du LILLA N°2*, sous la direction de Henri ZINGLE. Université de Nice-Sophia-Antipolis, pp. 81-100.

KRZESZOWSKI T.P., 1990. « The axiological aspect of idealized cognitive models », in *Meaning and Lexicography*. LLSEE 28, J. Tomaszczyk & B. Lewandowska-Tomaszczyk eds. John Benjamins, pp.135-165.

LABEAU E., 2002. « L'unité de l'imparfait : vues théoriques et perspectives pour les apprenants du français, langue étrangère », in *Travaux de Linguistique n°45*, « La notion d'invariant sémantique », *Revue Internationale de Linguistique Française*. Duculot, pp.157-184.

LACA B., 2004. « Les catégories aspectuelles à expression périphrastique : une interprétation des apparentes 'lacunes' du français », in *Langue Française n°141*, « le français parmi les langues romanes ». Larousse, pp.85-98.

LAMARCHE J., 2002. « Invariance et contingence », in *Travaux de Linguistique n°45*, « la notion d'invariant sémantique », *Revue Internationale de Linguistique Française*. Duculot, pp.15-44.

LARJAVAARA M., 2000. *Présence ou absence de l'objet*. Helsinki: Academica Scientiarum Fennica.

LARRIVEE P., 2002a. « La notion d'invariant sémantique », in *Travaux de Linguistique n°45*, « La notion d'invariant sémantique », *Revue Internationale de Linguistique Française*. Duculot, pp.7-14.

LARRIVEE P., 2002b. « Invariants sémantiques et constructions syntaxiques », in *Travaux de Linguistique n°45*, « La notion d'invariant sémantique », *Revue Internationale de Linguistique Française*. Duculot, pp.67-82.

LE GOFFIC P., 1997. « Formes en *-ant* et contexte », in *Co-texte et calcul du sens*, Claude Guimier éd., Presses Universitaires de Caen, pp.123-133.

LEEMAN D., 1999. « *Dans un juron, il sauta sur ses pistolets*. Aspects de la polysémie de la préposition *dans* », in *Revue de Sémantique et Pragmatique*. 1999. Numéro 6. Presses Universitaires d'Orléans, pp. 71-88

LERAT P., 1999. « L'offre en sémantique lexicale », in *Cahiers de Lexicologie*. 1999-2. Numéro 75. *Revue Internationale de Lexicologie et Lexicographie*. Didier Érudition, pp.5-22.

LEVIN B., 1993. *English Verb Classes and Alternations. A preliminary investigation*. The University of Chicago Press, Chicago and London.

LEWANDOWSKA-TOMASZCZYK B., 1990. « Meaning, synonymy and the dictionary », in *Meaning and Lexicography*. LLSEE 28, J. Tomaszczyk & B. Lewandowska-Tomaszczyk eds. John Benjamins, pp.181-208.

LUDEWIG P., 2001. « LogoTax – un outil exploratoire pour l'étude de collocations en corpus », in *Linguistique de corpus*, n°2/2001, Coll. *Traitement automatique des langues* vol 42. Hermes, pp.623-642.

LYONS J., 1970. *Linguistique générale*. Paris. Larousse.

MAINGUENEAU D., 2003. *Linguistique pour le texte littéraire*, 4^e édition. Collection Lettres sup., Nathan Université. Paris.

MALMBERG B., 1983. *Analyse du langage au XXe siècle. Théorie et méthodes*. PUF, Paris.

MALRIEU D., RASTIER F., 2001. « Genres et variations morphosyntaxiques », in *Linguistique de corpus*, n°2/2001, Coll. *Traitement automatique des langues* vol 42. Hermes, pp.570-579.

MANGUIN J-L., FRANÇOIS J. et alii, 2004. *Le dictionnaire électronique des synonymes du CRISCO. Un mode d'emploi à trois niveaux*. Cahiers du CRISCO n°17, juillet 2004. Université de Caen.

MARTIN E., 1993. *Reconnaissance de contextes thématiques dans un corpus textuel. Éléments de lexico-sémantique*. Collection « Études de sémantique lexicale » CNRS-INALF. Didier Érudition.

MARTIN R., 1988. « Temporalité et "classes de verbes" », *L'information grammaticale* n°39, pp.3-8.

MAYAFFRE D., 2006. « Rôle et place des corpus en linguistique : réflexions introductives ». Actes du colloque *Rôle et place des corpus en linguistique* JETOU'2005. Université de Toulouse-Le Mirail, 1-2 juillet 2005, pp.5-17.

MEL'ÇUK I., 1996. « Paraphrase et lexique dans la théorie linguistique Sens-Texte », in *Lexique et paraphrase*.

MEL'ÇUK I., 1997. « Vers une linguistique Sens-Texte ». *Leçon inaugurale* au Collège de France, Chaire Internationale.

MEL'ÇUK I. et alii, 1984. *Dictionnaire explicatif et combinatoire du français contemporain. Recherches Lexico-Sémantiques I*. Les Presses de l'Université de Montréal.

MEL'ÇUK I., CLAS A., POLGUÈRE A., 1995. *Introduction à la lexicologie explicative et combinatoire*. Champs linguistiques. Éditions Duculot, AUPELF UREF.

MULLER C., 1996. *La subordination en français*. Collection U, série « Linguistique ». Armand Colin.

MULLER P., SARDA L., 1998. « Représentation de la sémantique des verbes de déplacement transitifs du français », *T.A.L.*, vol. 39, n°2, pp.127–147.

MULLET V., DENHIERE G., 1997. « Accès au lexique et ambiguïtés lexicales nominales : effet de la polarité des homographes et de la nature du contexte », in *Sémantique linguistique et psychologie cognitive. Aspects théoriques et expérimentaux*, Jacques François et Guy Denhière eds., Presses Universitaires de Grenoble, pp.51-74.

NOWAKOWSKI M., 1990. “Metaphysics of the dictionary *versus* the lexicon”, in *Meaning and Lexicography*. LLSEE 28, J. Tomaszczyk & B. Lewandowska-Tomaszczyk eds. John Benjamins, pp.3-19.

NUNBERG G., ZAENEN A., 1992. “Systematic polysemy in lexicology and lexicography”. *Euralex 92, Proceedings*, pp.387-396.

OOI V.B.Y., 1998. *Computer corpus lexicography*, Edinburgh Textbooks in Empirical Linguistics.

OTMAN G., 1996. *Les représentations sémantiques en terminologie*. Sciences cognitives, Masson.

PAILLARD D., 1997. « Co-texte, collocations, lexique », in *Co-texte et calcul du sens*, Claude Guimier éd., Presses Universitaires de Caen, pp.63-71.

PAILLARD D., 2002. « Prépositions et rection verbale », in *Travaux de Linguistique n°44*, « La préposition (suite) », *Revue Internationale de Linguistique Française*. Duculot, pp.51-67.

PEETERS B., 2002. « La métalangue sémantique naturelle au service de l'étude du transculturel », in *Travaux de Linguistique n°45*, « La notion d'invariant sémantique », *Revue Internationale de Linguistique Française*. Duculot, pp.83-101.

PEROZ P., 2002. « Régularités de la variation sémantique du verbe *lutter* », in *Travaux de Linguistique n°45*, « La notion d'invariant sémantique », *Revue Internationale de Linguistique Française*. Duculot, pp.45-66.

- PICOCHÉ J., 1994. *Didactique du vocabulaire français*. Nathan, Paris.
- PICOCHÉ J., ?. « Les deux recettes de base de la cuisine lexicale ». *Revue le français dans tous ses états*, n°30 Paul Valéry-enseigner la langue. Montpellier.
- PLOUX S., 1997. « Modélisation et traitement informatique de la synonymie ». *Linguisticae Investigationes, Tome XXI/1997, Fascicule 1*.
- PLOUX S., JI H., 2003. « A Model for Matching Semantic Maps Between Languages (French / English, English / French) ». *Computational Linguistics*. 29(2), pp.155-178.
- PLOUX S., VICTORRI B., 1998. « Construction d'espaces sémantiques à l'aide de dictionnaires informatisés des synonymes ». *TAL*, 39(1), pp.161-182.
- POLGUÈRE A., 1998. « La théorie Sens-Texte », in *Dialangue*, Vol. 8-9, Université du Québec à Chicoutimi, pp.9-30.
- POLGUÈRE A., 2000. « Towards a theoretically-motivated general public dictionary of semantic derivations and collocations for French ». *Proceedings of EURALEX 2000, Approaches to Lexical Combinatorics*, pp.517-529.
- PREVOST S., 2003. « Les compléments spatiaux : du topique au focus en passant par les cadres », in *Travaux de Linguistique n°47, « adverbiaux et topiques », Revue Internationale de Linguistique Française*. Duculot, pp.51-77.
- PUSTEJOVSKY J., 1996. *The generative lexicon*. The MIT Press.
- PUSTEJOVSKY J., BOUILLON P., 1996. « Aspectual Coercion and Logical Polysemy » in : Pustejovsky J. et al. (eds.). *Lexical Semantics*, Oxford, Clarendon Press, pp.133-162.
- QUINE, 1960. *Word and Object*. MIT Press.
- RABATEL A., 2003. « Les verbes de perception en contexte d'effacement énonciatif : du point de vue représenté aux discours représentés », in *Travaux de Linguistique n°46, Revue Internationale de Linguistique Française*. Duculot, pp.49-88.
- RASTIER F., 2005. « Enjeux épistémologiques de la linguistique de corpus », in *La linguistique de corpus*, G. Williams éd. Linguistiques, Presses Universitaires de Rennes, pp.31-45.

REBOUL A., 1999. « Concepts spécifiques, concepts génériques : la référence et le système conceptuel dans les représentations mentales ». *Séminaire interne, LED-LORIA*, 25/11/1999. Document de présentation, pp.17-18.

RECANATI F., 1997. La polysémie contre le fixisme. *Langue Française*, n°113, Mars 1997. *Aux sources de la polysémie nominale*. P. Cadiot, B. Habert (ed), pp.107-123.

RECANATI C., RECANATI F., 1999. « La classification de Vendler revue et corrigée », *Cahiers Chronos 4, La modalité sous tous ses aspects*, textes réunis par S. Vogeleeer, A. Borillo, M. Vuillaume et C. Veters, Amsterdam/Atlanta, pp.167-184

REY A., CHANTREAU S., 1997. *Dictionnaire des expressions et locutions*. Le Robert.

RIEGEL M., PELLAT J-C., RIOUL R. 1994. *Grammaire méthodique du français*. PUF Linguistique Nouvelle.

ROUSSEAU A., 2000. « L'irréel ou le défi de l'impossible dans la langue », in *Incidences de l'Impossible dans le Langage*, revue *Sémiotiques*, Décembre 2000, n°18/19. CNRS-ILF, p ;15-30.

ROUSSELOT F., GAGEAN N., 2005. « Les phénomènes linguistiques et extra-linguistiques qui altèrent l'analyse linguistique automatique de textes », in *La linguistique de corpus*, G. Williams éd. Linguistiques, Presses Universitaires de Rennes, pp.337-356.

SAINT-DIZIER P., VIEGAS E. eds, 1995. *Computational Lexical Semantics*. "Studies in Natural Language Processing". Cambridge University Press.

SANCHEZ PEREZ F.J., 1994. « De la Désambiguïsation Assistée par Ordinateur à l'Élaboration des Grammaires Textuelles ». *Literary and Linguistic Computing*, vol. 9, N°3, 1994. Oxford University Press, pp.195-202.

SCHNEDECKER C., CHAROLLES M., KLEIBER G. et DAVID J. (ed.), 1994. *L'anaphore associative. Aspects linguistiques, psycholinguistiques et automatiques*, Paris, Klincksieck.

SIBLOT P. ed., 1990. *Le langage et le réel*, Cahiers de Praxématique 15.

SMALL S.I., COTTRELL G.W., TANENHAUS M.K. eds., 1988. *Lexical Ambiguity Resolution*. Perspectives from psycholinguistics, neuropsychology & artificial intelligence ; Morgan Kaufmann Publishers.

SNELL-HORNBY M., 1990. "Dynamics in meaning as a problem for bilingual lexicography", in *Meaning and Lexicography*. LLSEE 28, J. Tomaszczyk & B. Lewandowska-Tomaszczyk eds. John Benjamins, p;209-226.

SOWA J., 1983. "Generating language from conceptual graphs", *comp. & Maths with Appls*, volume 9, numéro1, pp.29-43.

SOWA J., 1992. "Logical structures in the Lexicon" in *Lexical Semantics and Commonsense Reasoning*, edited by James Pustejovsky and Sabine Bergler, LNAI 627, Springer-Verlag, Berlin, 1992, pp.39-60.

TODOROV T., 1967. *Littérature et signification*. Paris: Librairie Larousse.

TOMASZCZYK J., LEWANDOWSKA-TOMASZCZYK B. eds, 1990. *Meaning and Lexicography*. LLSEE 28. John Benjamins.

TOURATIER C., 1996. *Le système verbal français*. Collection U, série « Linguistique ». Armand Colin.

TOURY G., 1990. "Translation-specific lexical items and their representation in the dictionary", in *Meaning and Lexicography*. LLSEE 28, J. Tomaszczyk & B. Lewandowska-Tomaszczyk eds. John Benjamins, pp.287-300.

VALETTE M., 2004. « Sémantique interprétative appliquée à la détection automatique de documents racistes et xénophobes sur Internet », in *Approches Sémantiques du Document Numérique, Actes du 7e Colloque International sur le Document Électronique, 22-25 juin 2004*, Patrice Enjalbert et Mauro Gaio, éd., 2004, pp.215-230.

VANDELOISE C., 1999. "Quand dans quitte l'espace pour le temps », in *Revue de Sémantique et Pragmatique. 1999. Numéro 6*. Presses Universitaires d'Orléans, pp.145-162.

VENDLER Z., 1967. *Linguistics in Philosophy*, Ithaka, New York: Cornell University Press.

VERSTIGGEL J-C., 1997. « Le rôle du contexte dans la construction de la signification », in *Sémantique linguistique et psychologie cognitive. Aspects théoriques et expérimentaux*, Jacques François et Guy Denhière eds., Presses Universitaires de Grenoble.

VET C., 1980. *Temps, aspects et adverbes de temps en français contemporain*, Genève. Droz.

VICTORRI B., 1997a. « La place du mouvement dans la structure sémantique des langues », communication présentée à *l'école d'été de l'ARC/CNRS* : « *Le mouvement des boucles sensori-motrices aux représentations cognitives et langagières* ». (article obtenu à l'adresse : www.utc.fr/arco/activites/ecoles/Bonas77/Victorri.html)

VICTORRI B., 1997b. « Modéliser les interactions entre une expression polysémique et son co-texte », in *Co-texte et calcul du sens*, Claude Guimier éd., Presses Universitaires de Caen, pp.233-245.

VICTORRI B., ?. « Comment modéliser la construction dynamique du sens d'un énoncé ? ». *Séminaire Intelligence Artificielle du LIP6*.

VICTORRI B., FUCHS C., 1996. *La polysémie, construction dynamique du sens*. Hermes.

VIOLI P., 2000. Prototypicality, typicality and context. In *Meaning and Cognition*, Liliana Albertazzi ed. CELCR, John Benjamins Publishing Company, pp.103-122.

WALTER H., 1994. *L'aventure des langues en Occident. Leur origine, leur histoire, leur géographie*. Robert Laffont.

WIERZBYCKA A., 1996. *Semantics. Primes and universals*. Oxford University Press.

WILLIAMS G. & alii, 2005. *La linguistique de corpus*. Linguistiques, Presses Universitaires de Rennes.

YOUNG S., BLOOTHOOFT G. eds, 1997. *Corpus-based methods in language and speech processing*, ELSNET, Kluwer Academic Publishers.

Dictionnaires

Dictionnaire bilingue français-anglais Robert & Collins.

Dictionnaire bilingue français-espagnol Robert & Collins, « *les pratiques* ».

Dictionnaire bilingue français-italien Robert & Collins, « *les pratiques* ».

Dictionnaire bilingue français-espagnol Larousse.

Dictionnaire bilingue français-italien Bordas.

TLF et TLFi, CNRS-InalF, Nancy.

Petit Robert français.

Petit Larousse français.

Webster's Universal Dictionary English.

Dictionnaire des synonymes de l'INALF.

Nouveau dictionnaire des synonymes Larousse.

Routledge Dictionary of Language and Linguistics, by H. BUSSMANN, translated and edited by G.P. Trauth & K. Kazzazi. Routledge, 1996.

Quelques Pages Internet

Sur les classes d'objets :

<http://www-lli.univ-paris13.fr/rapport/LG/ClassesObjets.htm>

<http://www-lli.univ-paris13.fr/rapport/LG/humains.htm>