



HAL
open science

Les lexiques-miroirs. Du dictionnaire bilingue au graphe multilingue

Sébastien Haton, Jean-Marie Pierrel

► **To cite this version:**

Sébastien Haton, Jean-Marie Pierrel. Les lexiques-miroirs. Du dictionnaire bilingue au graphe multilingue. Revue TAL : traitement automatique des langues, 2007, pp.1-10. halshs-00141261

HAL Id: halshs-00141261

<https://shs.hal.science/halshs-00141261>

Submitted on 12 Apr 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Les Lexiques-Miroirs. Du dictionnaire bilingue au graphe multilingue.

Sébastien Haton, Jean-Marie Pierrel

Laboratoire ATILF Nancy Université – CNRS
BP 30687 44 avenue de la Libération, 54063 Nancy CEDEX
sebastien.haton@atilf.fr jean-marie.pierrel@atilf.fr

Résumé

On observe dans les dictionnaires bilingues une forte asymétrie entre les deux parties d'un même dictionnaire et l'existence de traductions et d'informations « cachées », i.e. pas directement visibles à l'entrée du mot à traduire. Nous proposons une méthodologie de récupération des données cachées ainsi que la « symétrisation » du dictionnaire grâce à un traitement automatique. L'étude d'un certain nombre de verbes et de leurs traductions en plusieurs langues a conduit à l'intégration de toutes les données, visibles ou cachées, au sein d'une base de données unique et multilingue. L'exploitation de la base de données a été rendue possible par l'écriture d'un algorithme de création de graphe synonymique qui lie dans un même espace les mots de langues différentes. Le programme qui en découle permettra de générer des dictionnaires paramétrables directement à partir du graphe.

Abstract

Lexical asymmetry and hidden data, i.e. not directly visible into one lexical entry, are phenomena peculiar to most of the bilingual dictionaries. Our purpose is to establish a methodology to highlight both phenomena by extracting hidden data from the dictionary and by re-establishing symmetry between its two parts. So we studied a large number of verbs and integrated them into a unique multilingual database. At last, our database is turned into a “multilexical” graph thanks to an algorithm, which is binding together words from different languages into the same semantic space. This will allow us to generate automatically dictionaries straight from the graph.

Mots-clés : dictionnaires bilingues, traduction automatique, graphe multilingue, algorithme, polysémie verbale, dissymétrie lexicographique

Keywords: bilingual dictionaries, machine translation, multilingual graph, algorithm, verbal polysemy, lexicographical asymmetry

1. Introduction

A 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. 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 est à noter que toutes nos études incluent une composante 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 que nous décrivons dans cet article. Tout d'abord, nous discuterons 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 la méthodologie que nous avons développée pour essayer de pallier ces lacunes et pour mettre en valeur les informations cachées ou indirectes que ces ouvrages contiennent. 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.

2. Analyse critique d'une entrée d'un dictionnaire bilingue français-anglais

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 l'extrait suivant 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 par ailleurs dans (Haton 2006), ce qui est beaucoup :

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

VT **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*] abandonner ♦ **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 est donc : 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,

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

VT a (= *go away from*) [+ town] **quitter, partir de** ;

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 *town*. 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 travail, et 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, dans une certaine mesure, deux dictionnaires de synonymes monolingues. En tout état de cause, il faut que ces liens apparaissent clairement dans notre modélisation.

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.

3. Principe des lexiques-miroirs et généralisation de la méthode

Pour bien illustrer le fonctionnement des lexiques miroirs ou inversés (Haton 2003, 2006), 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 prenons comme exemple le verbe *traverser* pour lequel nous

avons extrait du dictionnaire Le Petit Robert toutes les acceptions et nous les avons numérotées :

•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*

Ensuite, nous avons fait le relevé des verbes anglais qui entretiennent un lien de traduction avec *traverser* et nous avons associé à chaque lien l'indice correspondant. La deuxième colonne du tableau correspond au LD 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.

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 jambes, 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	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)
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

Figure 1 : Tableau de comparaison LD/LI pour la traduction du verbe traverser en 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, nous retenons 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 et 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 regroupements 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. Ensuite, nous nous sommes attachés à rechercher une modélisation du problème dans le but d'automatiser intégralement les liens de traduction directs, cachés et asymétriques avec comme objectif avoué d'autoriser la restitution d'un dictionnaire bilingue entièrement symétrique sans omission d'information. Ce souci d'exhaustivité que nous revendiquons rejoint celui qui sous-tend les travaux de Mel'çuk (1984, 1995) mais sans ajout extérieur pour notre part. Sowa (1992) et Dutoit (2000) adoptent la même démarche dans leurs propres modèles en partant du principe que toute la langue est potentiellement formalisable dans un même espace. Bien entendu, nous voyons cette préoccupation comme un objectif stimulant mais pas comme une fin.

Dans l'optique d'une modélisation efficace, nous avons introduit la notion linguistique de lexie au sens que Pottier puis Mel'çuk et son équipe (1995) 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.

4. Modélisation des données et programmation des graphes

4.1 Modélisation du dictionnaire : le traitement des lexies au cas par cas

Le dernier volet de notre travail concerne la modélisation proprement dite des lexies repérées dans les dictionnaires et – éventuellement – dans les corpus littéraires. La première étape de la modélisation est la transformation en bases de données des dictionnaires bilingues. Revenons à notre fragment d'extrait de *to leave*.

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

VT a (= *go away from*) [+ *town*] **quitter, partir de** ;

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).

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 (Haton 2004 & 2006) :

- 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 identiques.

Le tout peut être contenu dans une base de données intermédiaire entre le dictionnaire et le graphe qui peut découler de l'implantation de tous les liens de synonymie entre les lexies. 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.

Dans leur méthodologie d'élaboration, les graphes que nous nous proposons de générer ressemblent aux graphes construits par l'équipe de Sabine Ploux à l'ISC Lyon (1997, 2003) ainsi qu'au dictionnaire des synonymes du CRISCO à Caen (Manguin & François, 2004).

4.2 Algorithme de transformation en graphe multilingue des lexiques bilingues transformés

Nous présentons ci-dessous l'algorithme qui a servi pour l'écriture du programme initial. 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) ; Nil représente l'ensemble vide ou l'absence d'élément ; 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, Collocsuj, 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, Collocsuj, Registre, Info compl)

Dico = {Nil} ; *Liens* = {Nil} ;

Ligne = 1 ;

Motsource = TAB[*ligne*, 1] ;

LS = TAB[*ligne*, 2] ;

Numcons = 0 ;

Tant que motsource .NE. Nil répéter

Econstr = {Nil} ;

Tant que (*Motsource*, *LS*) = (TAB[*ligne*, 1], TAB[*ligne*, 2]) répéter

Numcons = *Numcons* + 1 ;

IndiceMS = *Numcons* ;

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. Nil 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. Nil alors début

IndiceSMC = *Numcons* ;

Numcons = *Numcons* + 1 ;

EltconstrSMC = (*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. Nil et SMC.NE. Nil 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

L'algorithme de création de graphes « multilexicaux », car nous n'osons dire pleinement multilingues, a été programmé tel quel il y a quelques mois par Etienne Petitjean à l'ATILF. Nous n'avons évidemment pas eu le temps de tester toutes ses potentialités mais nous savons déjà que la transformation des bases de données en graphe est acquise (<http://www.atilf.fr/perso/pierrel/telechargement/result.txt>). D'autre part, la perspective de disposer de lexiques 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.

5. Conclusion

Bien sûr, 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, nous pensons 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, naturellement orientés, soient désorientés et le dictionnaire bilingue symétrique est prêt à être généré.

Les applications les plus directes que nous envisageons sont de deux types : nous espérons pouvoir générer au moins partiellement des dictionnaires inédits à partir de ceux qui auront déjà été transformés. Par exemple, admettons que nous ayons saisi un dictionnaire français-anglais et un dictionnaire français-italien. À partir de cela, nous souhaitons réfléchir aux moyens informatiques à mettre en œuvre pour générer un dictionnaire anglais-italien complètement inédit. Ce travail a déjà été entamé et les premiers résultats sont encourageants, nonobstant la surgénération de liens de synonymie entre l'anglais et l'italien via le français. D'autre part, la méthode semble offrir un certain intérêt pour l'aide à l'alignement des textes en plusieurs langues, ce dernier générant un feedback très profitable pour l'enrichissement du

graphe multilingue. En définitive, le dictionnaire « fusionné » et les textes alignés constituent des mémoires de traduction quasiment inépuisables et susceptibles de s'enrichir mutuellement.

Références

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.

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., (2006). *Analyse et modélisation de la polysémie verbale dans une perspective multilingue : le dictionnaire bilingue vu dans un miroir*. Thèse de doctorat soutenue le 25 novembre 2006. Université de Nancy 2.

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.

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.

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.

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.