

La base de données liégeoise du *Robert & Collins*: un outil pour le lexicographe, le linguiste ou le traducteur

Cet article traite de la transformation du dictionnaire bilingue *Robert & Collins* (anglais-français) en une base de données lexico-sémantiques offrant une grande diversité de chemins d'accès. Le modèle théorique sous-jacent est celui du linguiste russe Igor Mel'cuk dont la théorie des fonctions lexicales est utilisée pour modéliser les relations collocationnelles et paradigmatiques découvertes au sein de la microstructure des entrées du dictionnaire bilingue. Le but est de fournir aux lexicographes, aux linguistes ou aux traducteurs un outil souple permettant d'accéder facilement aux collocations par le biais de la base, du collocatif, des catégories syntaxiques ou des relations lexico-sémantiques. Les programmes d'application permettant l'extraction de données linguistiquement pertinentes sont décrits en détail.

Termes-clés :
collocation; lexicographie
computationnelle; dictionnaire
informatisé; bases de données
lexicales; dictionnaire bilingue;
fonction lexicale.

1 Introduction

Depuis une quinzaine d'années environ, la communauté scientifique et, dans une certaine mesure, le monde industriel, s'intéressent à l'exploitation et à la construction de dictionnaires informatisés. Les raisons de cet intérêt sont multiples. Confrontée à une explosion du volume de textes à traduire et à traiter, notre société de l'information se voit contrainte de développer toute une série d'outils destinés à faciliter la tâche du rédacteur, du traducteur ou du documentaliste. La qualité des systèmes de traduction assistée par ordinateurs et des systèmes de recherche documentaire dépend en grande partie de la taille et du raffinement linguistique du module lexical qui les accompagne. La construction de ces lexiques (bilingues ou monolingues) est cependant une tâche coûteuse et de longue haleine, qui réclame des compétences lexicographiques et linguistiques encore trop peu répandues. Il n'est dès lors pas surprenant que les chercheurs se soient penchés sur les possibilités de réutilisation et d'exploitation offertes par certains dictionnaires

commerciaux disponibles sur support informatique. La tradition lexicographique anglo-saxonne a en effet produit une série de dictionnaires d'apprenants dont la richesse potentielle a alléché bon nombre de linguistes désespérément en quête de données lexicales susceptibles de nourrir leurs applications computationnelles (voir entre autres les dictionnaires des maisons Longman, Oxford University Press ou HarperCollins - Procter 1978, Cowie 1989, Sinclair 1987; voir aussi Boguraev & Briscoe 1989 pour un panorama de la lexicographie computationnelle et une description détaillée de recherches sur le *Longman Dictionary of Contemporary English*).

Les dictionnaires bilingues ont généralement moins attiré l'attention des chercheurs. Il est vrai que les maisons d'édition se sont montrées moins enclines à distribuer les bandes magnétiques de leurs ouvrages bilingues, souvent par peur d'une utilisation malhonnête de leurs données. Une raison scientifique de ce faible intérêt réside dans le fait que les dictionnaires bilingues existant sous format informatique se prêtaient moins facilement à une structuration en base de données que leurs équivalents monolingues anglais. La structure de ces derniers rendait plus facile l'identification de divers types d'information d'une entrée lexicale

Communauté française de

Belgique

(partie du discours, codes syntaxiques, définitions, exemples, propriétés morphologiques, etc.) alors que les quelques dictionnaires bilingues disponibles n'étaient rien d'autre que les bandes magnétiques ayant servi à la photocomposition de l'ouvrage. Un énorme travail d'interprétation et de décodage des entrées était dès lors nécessaire afin de transformer un simple dictionnaire lisible par machine en une véritable base de données lexicales permettant un accès rapide, aisé et diversifié aux informations lexicales. C'est exactement ce que nous avons tenté de réaliser dans le cadre d'un doctorat en linguistique anglaise à l'Université de Liège (Fontenelle 1995). Le dictionnaire utilisé dans cette recherche était le dictionnaire anglais-français *Robert & Collins* (Atkins & Duval 1978) dont la version magnétique avait été mise à notre disposition dans le courant des années quatre-vingts, sous contrat de recherche avec les éditeurs (les Dictionnaires Le Robert à Paris et HarperCollins Publishers à Glasgow). Les recherches ont pu être menées à bonne fin grâce au concours précieux de Jacques Jansen et Luc Alexandre, informaticiens à l'Université de Liège. Le premier a conçu la structure de la base de données relationnelle et a analysé les bandes magnétiques du dictionnaire. Le second a participé, à mes côtés, à la création de toute une série de programmes d'application permettant d'interroger la base de données et d'en extraire de l'information.

2 Une base de données lexico-sémantiques

Les lecteurs familiers des dictionnaires *Robert & Collins* savent que les lexicographes ont appliqué un système cohérent permettant de spécifier les restrictions de sélections,

les collocations restreintes et les indicateurs de champs sémantiques dans les entrées. Ce système de codage est basé sur la typographie, ainsi que l'illustrent les entrées suivantes:

grip 4 vi [*wheels*] adhérer, mordre; [*screw, vice, brakes*] mordre; [*anchor*] crocher (sur le fond)

jam 4 b (*become stuck*) [*brake*] se bloquer; [*gun*] s'enrayer; [*door, switch, lever*] se coincer

line vt *clothes* doubler (*with de*); [*bird*] *nest* garnir, tapisser; (*Tech*) revêtir, chemiser; *brakes* garnir

squeal 2 vi a [*person, animal*] pousser un *or* des cri(s) aigu(s) *or* perçant(s); [*brakes*] grincer, hurler; [*tyres*] crisser
squeal 1 n [*person, animal*] cri aigu *or* perçant; [*brakes*] grincement *m*, hurlement *m*; [*tyres*] crissement *m*

Comme on peut le constater, l'appareil métalinguistique en italiques couvre toute une série d'informations cruciales pour la désambiguïsation et la sélection de la traduction correcte dans un contexte donné. Les informations en italiques vont de la spécification des parties du discours (*n*, *vt*, *adj*, *vi*...) aux codes matières (*Bio*, *Tech*, *Scol*...), en passant par les collocations et autres restrictions de sélection. Il est clair que, dans les exemples ci-dessus, les mots *person* et *animal* (s.v. *squeal*) font référence à une restriction d'ordre sémantique (le sujet du verbe doit être +HUMAIN ou +ANIMAL) alors que les diverses occurrences de *brake* (= fr *frein*) ou *tyre* (= *pneu*) dénotent une restriction d'ordre collocationnel. On conçoit aisément l'importance de ce genre d'information pour le traitement automatique de la langue et, plus spécifiquement, pour la traduction assistée par ordinateur où la détermination du sens d'un mot dépend des informations contextuelles dont dispose le système. L'approche utilisée par les lexicographes du *Robert & Collins*

pour le codage de ces contraintes peut être résumée comme suit:

- Les noms sujets typiques d'un verbe apparaissent entre crochets;
- Les noms utilisés comme compléments d'un autre nom apparaissent aussi entre crochets;
- Les objets directs d'un verbe et les noms modifiés par un adjectif apparaissent à côté de la partie du discours (pas de crochets ni de parenthèses);
- Les adjectifs, verbes ou adverbes modifiés par un adverbe apparaissent sans crochets ni parenthèses.

Les entrées ci-dessus appellent plusieurs commentaires. Le verbe *squeal* est traduit comme «pousser un cri *or* des cris aigu(s) *or* perçant(s)». La conjonction *or* est utilisée ici par le lexicographe pour gagner de la place mais il est évident qu'une machine ne pourra pas se servir de ce genre de traduction. Un décompactage est absolument nécessaire afin de permettre à la machine de reconnaître quatre traductions (*pousser un cri perçant*, *pousser des cris perçants*, *pousser un cri aigu*, *pousser des cris aigus*). Ce travail de décompactage est difficilement automatisable, mais nécessaire si l'on veut utiliser le champ des traductions comme clé d'accès (et chercher, par exemple, les verbes anglais correspondant à la traduction *pousser un cri perçant*). Ce travail de bénédictin a été mené à bien par Jacques Jansen grâce à un programme interactif qui émettait des hypothèses sur la reconstitution des traductions.

On peut également remarquer que, pour un utilisateur francophone, la partie anglais-français du *Robert & Collins* est un dictionnaire de décodage puisqu'elle permet d'interpréter le texte anglais et de choisir la traduction française correcte en fonction des éléments contextuels de l'appareil métalinguistique. Dans le même temps, on note que les différentes collocations correspondent au modèle de combinaison polaire

développé par Hausmann (1979) puisqu'elles comportent une base et un collocatif. Dans ce modèle bien connu, la base est responsable de la sélection du collocatif dont le sens est en quelque sorte adapté en fonction du contexte. Le problème de l'accès à ces collocations est cependant crucial, ainsi qu'en témoignent les quelques dictionnaires de collocations influencés par les théories de Hausmann, comme par exemple le *BBI* anglais (Benson *et al.* 1986) ou le *Kontextwörterbuch* français-allemand (Ilgenfritz *et al.* 1989). Ces dictionnaires offrent un accès aux collocations par le biais de leurs bases, ce qui répond à un besoin dans une perspective d'encodage. Il faut malheureusement constater que, dans cette optique, les entrées ci-dessus ne sont que d'une faible utilité si le lecteur souhaite partir d'une base quelconque et retrouver les collocatifs potentiels. L'organisation purement alphabétique du dictionnaire ne permet en effet l'accès que via les collocatifs (les entrées proprement dites) alors que les bases correspondent aux éléments en italiques, inaccessibles en tant que clé d'accès primaire dans la version imprimée.

C'est ici que l'informatique apporte une aide précieuse au linguiste ou au traducteur qui a recours au dictionnaire. Dans sa version développée à Liège, la base de données est organisée de façon à permettre l'accès à n'importe quel type d'information, qu'il s'agisse de l'entrée, de la partie du discours, de la base (à l'origine en italiques), de la traduction française, des registres de langue ou des codes matières. Cela signifie qu'il est par exemple possible d'extraire les verbes prenant le nom *brake* comme objet direct (*jam*, *line*, *operate*, *reline*) ou les verbes intransitifs dont *brake* peut être sujet (*drag*, *fail*, *grip*, *jam*, *scream*, *screech*, *squeal*) ou les traductions françaises respectives de ces verbes (*frotter*,

lâcher, *mordre*, *se bloquer*, *hurler*, *grincer*). On imagine aisément les applications potentielles de ce type de base de données puisqu'il est possible d'accéder en quelques secondes à des données implicites dans la version «papier» du dictionnaire. Pour obtenir le même résultat, il faudrait parcourir les quelque 800 pages du dictionnaire afin d'y repérer, au sein des entrées, les occurrences de *brake* ou de toute autre base en italiques. Les perspectives dans le domaine de l'enseignement de la traduction sont tout aussi évidentes puisqu'il s'agit en fait de développer un dictionnaire en ligne permettant de fournir à l'étudiant ou au traducteur chevronné du matériau linguistique lui permettant d'opérer les choix de traductions les plus judicieux en fonction d'un contexte donné. Or, les recherches sur l'organisation mentale du lexique ont montré que notre cerveau fonctionne par associations d'idées et que nous partons généralement d'un mot ou d'un concept qui, en quelque sorte, active un réseau de mots ou concepts partenaires avec lesquels il entretient une série de relations dites lexico-sémantiques.

3 Enrichissement et accès par des fonctions lexicales

La notion de relation lexico-sémantique évoquée dans le paragraphe précédent est une notion clé dans l'approche relationnelle qui sous-tend nos travaux. Si une manipulation informatique du dictionnaire *Robert & Collins* a permis de générer les réseaux sémantiques d'environ 10 000 bases (à l'origine en italiques), ce qui représente un peu plus de 70 000 paires de «partenaires» potentiels (*brakes drag*, *fail*, *jam*, *scream*; *les freins frottent*, *lâchent*, *se bloquent*, *hurlent...*), il n'en reste pas moins que les liens qui unissent les

diverses composantes de ces collocations sont de nature diverse. Ainsi, des verbes comme *drag-frotter* évoquent une idée d'«obstruction» alors qu'un verbe comme *fail* (*lâcher*), en combinaison avec *brake*, signifie que les freins cessent de fonctionner correctement (fin d'un processus). Des verbes comme *scream*, *screech* ou *squeal*, quant à eux, désignent le son typique émis par les freins (on notera par ailleurs la récurrence des onomatopées à l'initiale en anglais alors que les équivalents français – hurler, grincer – ne présentent pas les mêmes régularités). La diversité des relations dans notre base de données nous a convaincus qu'il serait utile d'explicitier ces liens lexico-sémantiques en les étiquetant pour permettre une modélisation plus fine des réseaux sémantiques ainsi produits. Cette tâche de longue haleine allait ainsi nous permettre d'enrichir le dictionnaire en ajoutant à la base de données une clé d'accès nouvelle, à savoir la nature du lien sémantique unissant l'élément en italiques dans la version papier et l'entrée sous laquelle cet élément apparaît. Cette approche était destinée à permettre à l'utilisateur de la base de données lexicales de poser des questions sémantiquement beaucoup plus fines comme par exemple :

- Y a-t-il des verbes signifiant qu'un frein ne fonctionne pas comme il le devrait ?
- Quels verbes peuvent être utilisés pour désigner le bruit/son typique des freins ?
- Y a-t-il des verbes transitifs/causatifs signifiant que quelque chose rend les freins inopérants ?
- Y a-t-il des substantifs désignant des parties de freins (sabots, patins, garnitures...) ?

Les questions ci-dessus évoquent immanquablement le développement de dictionnaires spécialisés d'un genre nouveau, à savoir des dictionnaires

semi-techniques où l'accent est mis sur la combinatoire lexicale restreinte (voir entre autres Frawley 1988). Il ne s'agit certes pas de terminologie à proprement parler, mais il ne s'agit plus non plus tout à fait de langue générale et même un dictionnaire *a priori* non technique comme le *Robert & Collins* s'avère receler une richesse impressionnante en collocations restreintes pour la langue générale et le langage semi-technique.

La question se posait, au début du projet, de pouvoir déterminer un ensemble d'étiquettes permettant d'identifier tous ces liens lexico-sémantiques. Comme nous n'avions pas affaire à un glossaire terminologique nécessitant la détection et le codage de liens hyper-spécialisés (v. Blampain 1993, Blampain *et al.* 1992), nous nous sommes tournés vers la théorie Sens ↔ Texte d'Igor Mel'čuk et plus particulièrement vers son système complexe de fonctions lexicales. Ce système est suffisamment connu, du moins dans les grandes lignes, pour m'en épargner une nouvelle description (v. Mel'čuk *et al.* 1984, Mel'čuk, Clas et Polguère 1995). Je rappellerai tout simplement que le terme *fonction lexicale* désigne une relation de sens assez abstraite telle que l'expression linguistique de cette fonction dépend du lexème auquel elle vient se joindre. En se basant sur le modèle mathématique traditionnel, on utilise la notation $f(x)=y$, où f est la fonction lexicale (FL), x est le mot clé (la «base» chez Hausmann) et y la valeur de la fonction. Ainsi, si l'on postule qu'il existe une fonction «très» dénotant l'intensité, on aura:

«très» (malade)= très, gravement (*grièvement)

«très» (blessé)= grièvement, gravement (mais *très)

Les fonctions lexicales chez Mel'čuk sont souvent désignées à l'aide de mots d'origine latine. Ainsi, la fonction «très» est mieux connue sous le nom de Magn (Magn (prix)=

haut, élevé...). On distingue une cinquantaine de fonctions, comme par exemple:

- **Oper** (qui codifie les verbes supports, c'est-à-dire presque vides de sens et «supportant» le mot clé): Oper₁ (attention)= faire; Oper₁ (bévue)= commettre;
- **Liqu** (codifie les verbes exprimant la destruction de quelque chose): Liqu (maladie)= éradiquer; Liqu (loi)= abroger, abolir; Liqu (fichier)= détruire, effacer;
- **Son** (codifie le verbe exprimant le son typique): Son (plancher)= craquer; Son (cloche)= sonner; Son (éléphant)= barrire;
- **Fact**₀ (codifie les verbes exprimant la «réalisation» du mot clé): Fact₀ (rêve)= se réaliser;
- **Mult** (désigne un ensemble régulier): Mult (clé)= trousseau; Mult (abeille)= essaim; Mult (poisson)= banc;
- **Sing** (désigne une portion ou unité): Sing (poussière)= grain; Sing (herbe)= brin; Sing (paille)= fêtu.

Les fonctions lexicales peuvent se combiner pour former des fonctions complexes. Ainsi, IncepOper₁ dénote le verbe exprimant le début d'un processus (Incep< inchoatif) pour un verbe support: IncepOper₁ (habitude)= prendre. De la même façon, S₀Son générera le substantif (S₀) exprimant le son typique d'un mot clé: S₀Son (éléphant)= barrissement.

Les listes complètes des fonctions lexicales et de leurs propriétés sont données dans Mel'čuk (1984, 1988, 1992) et Mel'čuk *et al.* (1995), entre autres.

Dans le cas de la base de données du *Robert & Collins*, nous avons décidé de coder de façon systématique la fonction lexicale unissant l'élément en italiques et son entrée. Bien sûr, dans un certain nombre de cas (18000 sur un peu plus de 71000 enregistrements), le champ que nous avons créé dans la base de données pour y loger la FL

est resté vide. L'idée était, en effet, de coder la fonction lexicale pour autant que la relation corresponde à une des fonctions paradigmatiques ou syntagmatiques standard du modèle mel'čukien. Nous n'avons donc pas essayé de «forcer» les données dans un moule théorique. Nous n'avons cependant pas voulu non plus adhérer de façon trop flexible au modèle Sens ↔ Texte et nous avons ajouté une série de relations sémantiques non prévues par la théorie des FL. Par exemple, ayant noté que la relation partie-tout était largement représentée dans les données du *Robert & Collins*, nous avons décidé d'utiliser le modèle des fonctions lexicales en créant la fonction *Part* pour rendre cette relation accessible par le biais des programmes d'application, même si Mel'čuk la relègue au rang d'information encyclopédique, et donc non lexicale (*Part* (brake/frein)= drag/sabot, lining/garniture).

Un programme a été écrit afin de faciliter la tâche de l'encodeur et lui permettre de sélectionner la fonction appropriée à partir de menus déroulants émettant des hypothèses basées sur la partie du discours des composantes de la collocation et la nature typographique de l'élément en italiques (selon qu'il apparaissait entre parenthèses, entre crochets, hors crochets, etc.). L'enrichissement de la base de données et l'encodage proprement dit ont réclamé de nombreux mois de travail. Le résultat en valait cependant la peine puisque les programmes d'application permettent maintenant d'accéder à l'information via l'une des clés suivantes:

- Base (à l'origine, l'élément métalinguistique en italiques);
- «Collocatif» (à l'origine, l'entrée vedette);
- Partie du discours;
- Traduction de l'entrée-vedette;
- Traduction de la base;
- Fonction lexicale;

Les deux dernières clés d'accès correspondent à l'information qui a été ajoutée au dictionnaire. Ainsi, les éléments en italiques (par ex. *brake*) ont été systématiquement traduits afin de créer une véritable base de données bilingue. Il est donc possible de demander quels sont les verbes transitifs anglais se combinant avec le mot *brake* ou, sans passer par l'anglais, poser la même question pour le mot français *frein*. Comme on le verra plus bas, il est possible de combiner les clés d'accès ou de les utiliser séparément. Les programmes d'application permettent également d'extraire l'information, de l'exporter ou de mettre la base de données à jour sans en connaître la structure interne⁽¹⁾.

(1) La base de données est interrogeable sur PC et sur station UNIX (Sun SparcStation). Les programmes d'application ont été écrits en C par Luc Alexandre et en Clipper par moi-même. La structure proprement dite de la base de données relationnelle et les possibilités d'exploitation de celle-ci sont décrites en détail dans Fontenelle (1995).

On remarque que la structure des données imite le format $f(x)=y$. L'item appartenant à la métalangue en italiques apparaît entre parenthèses (*brake*) avec sa traduction, alors que la vedette contenant dans sa macrostructure une référence à *brake* apparaît après le signe =. La fonction lexicale (part, degrad, fact0, causobstr...) permet de regrouper des mots partageant un certain sens en combinaison avec une base de données. On notera également que certaines fonctions sont en fait des relations 1→n puisque les valeurs qui y sont associées ne sont pas nécessairement synonymes (part indique les différentes parties des freins; preparfact0 génère un ensemble de verbes indiquant les opérations à effectuer pour assurer le fonctionnement normal des freins, même si *line* et *reline* ne sont pas synonymes du tout).

Finalement, l'information entre crochets à la fin de chaque ligne fait référence au contenu d'un champ de la base de données qui indique la

nature typographique de l'item en italiques dans la version imprimée du dictionnaire. Ainsi, [C] signifie que ce terme apparaissait à l'origine entre crochets; [P] signifie qu'il apparaissait entre parenthèses et [S] qu'il apparaissait au niveau de «surface», c'est-à-dire sans parenthèses ni crochets. Comme on l'a vu, l'association de ce type d'information à la spécification de la partie du discours permet de déterminer si on a affaire à une collocation Verbe+Objet, Sujet+Verbe, Adjectif+Nom, Nom+Nom, etc.

Les programmes d'application ont été conçus de façon à pouvoir interroger la base de données souplement en offrant la possibilité de combiner les clés d'accès. Ainsi, le programme d'interrogation *Robcol*, conçu par notre collègue Luc Alexandre, permet, par le biais d'une ligne de commande, de spécifier les conditions et contraintes que l'utilisateur souhaite formuler. En utilisant la notation suivante, l'utilisateur peut sélectionner les champs à interroger en faisant précéder leur abréviation d'un trait d'union:

- i : item à l'origine en italiques;
- h : vedette/entrée (headword);
- pos : partie du discours (part of speech);
- lex : fonction lexicale.

La commande suivante, par exemple, permet d'extraire de la base de données les verbes prenant le mot *law* comme objet direct et signifiant 'liquider', 'détruire', 'éradiquer' (fonction lexicale = liqu):
robcol -i law -lex liqu.

La base de données contient les informations suivantes (qui peuvent être présentées sous d'autres formes, en ne sélectionnant que certains champs, par exemple, ce qui nous permet de réutiliser ces données pour interroger automatiquement un corpus de textes et valider les informations du dictionnaire):

L'exemple suivant illustre le type d'informations contenues dans la base de données pour le mot *brake*:

part (<i>brake</i> / frein)	= drag (sabot <m> de frein) [P]
degrad (<i>brake</i> / frein)	= drag (frotter) [C]
finfact0 (<i>brake</i> / frein)	= fail (lâcher) [C]
magnfact0 (<i>brake</i> / frein)	= grip (mordre) [C]
causobstr (<i>brake</i> / frein)	= jam (bloquer) [S]
obstr (<i>brake</i> / frein)	= jam (se bloquer) [C]
preparfact0 (<i>brake</i> / frein)	= line (garnir) [S]
part (<i>brake</i> / frein)	= lining (garniture <f>) [C]
antifact0 (<i>brake</i> / frein)	= to be off (être desserré) [C]
fact0 (<i>brake</i> / frein)	= to be on (être serré) [C]
causfact0 (<i>brake</i> / frein)	= operate (faire marcher) [S]
a1exces° (<i>brake</i> / frein)	= overheated (qui chauffe) [S]
s0antireal1 (<i>brake</i> / frein)	= release (dégagement <m>) [C]
preparfact0 (<i>brake</i> / frein)	= reline (changer la garniture de) [S]
son (<i>brake</i> / frein)	= scream (hurler) [C]
s0son (<i>brake</i> / frein)	= screech (grincement <m>) [C]
son (<i>brake</i> / frein)	= screech (grincer) [C]
s0son (<i>brake</i> / frein)	= sound (bruit <m>) [C]
s0son (<i>brake</i> / frein)	= squeal (grincement <m>) [C]
son (<i>brake</i> / frein)	= squeal (grincer) [C]
(<i>brake</i> / frein)	= to tamper with (toucher à (<sans> <permission>)) [S]

abolish (vt) : ~law~
 → abroger (loi,liqu)
 annul (vt) : ~law~
 → abroger (loi,liqu)
 do away with (vt fus) : ~law~
 → supprimer (loi,liqu)
 repeal (vt) : ~law~
 → abroger (loi,liqu)
 rescind (vt) : ~law~
 → abroger (loi,liqu)
 revoke (vt) : ~law~
 → rapporter (loi,liqu)

La flexibilité de l'outil permet de sélectionner les verbes transitifs (vt) jouant le rôle de verbe support sémantiquement appauvri (fonction lexicale = oper1) pour des items commençant par la lettre a, ce qui se formule comme suit:

rbocol -i «a*» -pos «vt*» -lex oper1

L'utilisation de l'astérisque après «vt» permet d'extraire tous les types de verbes transitifs, même les verbes à particules (phrasal verbs) si fréquents en anglais et codés «vt sep» ou «vt fus» (*bring forward, put forward, put in...*).

4 Conclusions

Il n'est pas possible de détailler ici tous les types d'application de la base de données bilingue que nous avons développée. Il va de soi qu'elle recèle des informations très précieuses pour le traducteur ou l'apprenant à la recherche du mot juste. Les possibilités de structuration sémantique offertes par l'enrichissement du dictionnaire à l'aide de fonctions lexicales permettent à l'utilisateur de poser des questions sémantiquement très fines en courant moins le risque de se voir submergé par des données non pertinentes. Nous avons en fait voulu apporter une dimension thésaurique à notre dictionnaire informatisé dans le but de créer une nouvelle génération d'outils lexicographiques. Plus fondamentalement, nous avons également tenté d'utiliser cette base de données enrichie pour des études de sémantique lexicale où l'accès à

une grande quantité de données est crucial pour pouvoir dégager des généralisations, par exemple en se concentrant sur la réalisation, à travers tout le lexique, d'une fonction lexicale particulière (voir par exemple Fontenelle 1994 sur la relation entre les fonctions lexicales et certains types de métaphores). Dans une autre optique, notre équipe compte exploiter ce dictionnaire dans une perspective de désambiguïsation automatique et de sélection de la traduction appropriée. Nous sommes convaincus que les recherches en linguistique computationnelle et en lexicographie vont, de plus en plus, nécessiter le développement de grandes bases de données lexicales, mono- et multilingues. Le *Robert & Collins* liégeois et les réseaux lexico-sémantiques qu'il contient n'offrent probablement qu'une parcelle de solution au problème complexe de la désambiguïsation. Nous avons cependant la faiblesse de croire qu'associé à d'autres outils, il pourrait jouer un rôle non négligeable.

Thierry Fontenelle,
 Département d'anglais,
 Faculté de Philosophie et lettres,
 Université de Liège.

Échantillon de la liste:

lay (vt) : ~accusation~
 sling (vt) : ~accusation~

fling (vt) : ~accusation~

enjoy (vt) : ~advantage~

give (vt) : ~answer~

give (vt) : ~answer~

proffer (vt) : ~apology~

present (vt) : ~apology~

tender (vt) : ~apology~

offer (vt) : ~apology~

put in (vt sep) : ~application~

bring forward (vt sep) : ~argument~

develop (vt) : ~argument~

put forward (vt sep) : ~argument~

pull out (vt sep) : ~argument~

prefer (vt) : ~argument~

pose (vt) : ~argument~

exert (vt) : ~authority~

exercise (vt) : ~authority~

wield (vt) : ~authority~

→ porter (accusation,oper1)

→ lancer (<at> <sb> à qn)

(accusation,oper1)

→ lancer (<at> <sb> à qn)

(accusation,oper1)

→ jouir de (avantage,oper1)

→ faire (réponse,oper1)

→ donner (réponse,oper1)

→ offrir (excuse,oper1)

→ présenter (<to> à) (excuse,oper1)

→ offrir (excuse,oper1)

→ offrir (excuse,oper1)

→ faire (candidature,oper1)

→ avancer (argument,oper1)

→ développer (argument,oper1)

→ avancer (argument,oper1)

→ sortir [informal] (argument,oper1)

→ présenter (argument,oper1)

→ présenter (argument,oper1)

→ exercer (autorité,oper1)

→ exercer (autorité,oper1)

→ exercer (autorité,oper1)

Remerciements

Les recherches décrites dans cet article s'inscrivent dans le cadre du projet Decide (*Designing and Evaluating Extraction Tools for Collocations in Dictionaries and Corpora - Multilingual Action Plan - MLAP 93/19*) partiellement financé par la Commission européenne et clôturé en janvier 1996. Outre l'Université de Liège (coordinateur du projet), le consortium comprenait l'Institut für maschinelle Sprachverarbeitung (Université de Stuttgart) et le Rank Xerox Research Centre de Grenoble.

Outre les éditeurs (HarperCollins Publishers, Glasgow et les Dictionnaires Le Robert, Paris), qui ont gracieusement mis à notre disposition les versions magnétiques du *Robert & Collins*, il me faut remercier ici Messieurs Jacques Jansen et Luc Alexandre dont les compétences informatiques ont permis l'aboutissement de ce travail.

Bibliographie

Atkins (B.T.S.) & Duval (A.), 1978: *Robert & Collins Dictionnaire anglais-français, français-anglais*, Paris, Le Robert, Glasgow, Collins.

Benson (M.), Benson (E.) & Ilson (R.), 1986: *The BBI Combinatory Dictionary of English*, Amsterdam & Philadelphia, John Benjamins.

Blampain (D.), 1993: «Notions et phraséologie. Une nouvelle alliance?», dans *Terminologies Nouvelles*, 10, décembre 1993, pp.43-49.

Blampain (D.), Petrussa (P.) & Van Campenhoudt (M.), 1992: «À la recherche d'écosystèmes terminologiques», dans *L'Environnement traductionnel - la station de travail du traducteur de l'an 2001; Journées scientifiques du Réseau thématique de recherche Lexicologie, terminologie et traduction*, Actes du colloque (Mons, 25-27 avril 1991), Presses de l'Université du Québec et Aupelf-Uref, pp.273-282.

Boguraev (B.) & Briscoe (T.), 1989: *Computational Lexicography for Natural Language Processing*, Longman & New York, Longman.

Cowie (A.P.), 1989: *Oxford Advanced Learner's Dictionary*, Oxford University Press.

Fontenelle (Th.), 1994: «Towards the construction of a collocational database for translation students», dans *Meta*, vol. 39, n°1, pp.47-56.

Fontenelle (Th.), 1994: «Using lexical functions to discover metaphors», dans Martin *et al.* (éd.) *Euralex'94 Proceedings*, Vrije Universiteit Amsterdam, pp.271-278.

Fontenelle (Th.), 1995: «Réseaux sémantiques et dictionnaires bilingues électroniques», à paraître dans les *Actes des 4^{es} Journées scientifiques du Réseau Lexicologie, terminologie et traduction: dictionnaire et lexicomatique*, Aupelf-Uref (Lyon, septembre 1995).

Fontenelle (Th.), 1995: *Turning a bilingual dictionary into a lexical-semantic database*, thèse de doctorat, Université de Liège.

Frawley (R.), 1988: «New forms of specialized dictionaries», dans *International Journal of Lexicography*, vol. 1, n°3, pp.189-213.

Hausmann (F.), 1979: «Un dictionnaire des collocations est-il possible?», dans *Travaux de Linguistique et de Littérature*, vol. 17, n°1, pp.187-195.

Heid (U.), 1994: «On Ways Words Work Together - Topics in Lexical Combinatorics», dans Martin *et al.* (éd.) *Euralex'94 Proceedings*, Vrije Universiteit Amsterdam, pp.226-257.

Ilgenfritz (P.), Stephan-Gabinel (N.) & Schneider (G.), 1989: *Langenscheidts Kontextwörterbuch Französisch-Deutsch*, Langenscheidt, Berlin-München.

Mel'čuk (I.) *et al.*, 1984/1988/1992: *Dictionnaire explicatif et combinatoire du français contemporain: Recherches lexicosémantiques I, II, III*, Presses de l'Université de Montréal.

Mel'čuk (I.), Clas (A.) & Polguere (A.), 1995: *Introduction à la lexicologie explicative et combinatoire*, Aupelf-Uref & Duculot (Universités francophones).

Procter (P.), 1978: *Longman Dictionary of Contemporary English*, Longman, Harlow.

Sinclair (J.), 1987: *Collins Cobuild English Language Dictionary*, HarperCollins Publishers.