

Pistes de description sémantique : le cas de Biolex, dictionnaire des bio-industries

François GAUDIN et Myriam BOUVERET

URA CNRS 1164¹, Université de Rouen et Praxiling, Université de Montpellier, France

Notre texte sera centré sur un dictionnaire des bio-industries qui nous servira de corpus et de fil conducteur. Ce dictionnaire a été conçu et dirigé par Louis Guespin, jusqu'à sa disparition, au sein de l'URA CNRS 1164, équipe rouennaise de recherche en sociolinguistique. Sa réalisation a été rendue possible grâce aux aides de l'ANVAR et du Ministère de la Recherche.

Ce dictionnaire a été développé sur logiciel Termex. Il recense une nomenclature de 1 200 termes dont il propose une description française en 19 rubriques et des traductions en anglais et en allemand. Mais, au-delà de ces caractéristiques techniques modestes, nous voudrions livrer quelques-unes des pistes que nous avons suivies dans un effort qui visait à décrire un vocabulaire tout en conciliant recherche fondamentale et recherche appliquée ; c'est-à-dire en cherchant à mettre à l'épreuve certaines options théoriques. Rappelons que ces options théoriques, sur lesquelles nous ne reviendrons pas ici, ont fait l'objet de quelques publications collectives rassemblant travaux et réflexions autour de la socioterminologie².

Nous nous intéresserons ici principalement aux rubriques que nous avons choisies pour décrire les relations lexicales et prédicatives en mettant en lumière les choix qui ont guidé notre travail. Parmi ces choix, nous relèverons le refus du recours aux définitions, une préférence accordée aux contextes, une prévention à l'égard de la notion de domaine, ainsi que la tentative de distinguer les niveaux de la signification et du concept.

¹ Unité de recherche associée au Centre national de la recherche scientifique, n° 1164 . « Sociolinguistique, usage et devenir de la langue ».

² Voir Gaudin et Assal, 1991 ; Gambier et Gaudin, 1993 ; Gardin *et al.*, 1994 ; Gaudin, 1995

Les relations lexicales

Concernant la description des relations lexicales, Biolex offre 5 rubriques, inégalement productives : hyperonyme, hyponyme, isonyme, antonyme, synonyme.

Les relations d'**hyperonymie** et d'**hyponymie** se sont avérées très utiles, ce qui n'a rien pour surprendre puisqu'elles sont au cœur de tout discours lexicographique. La relation de **synonymie** autorise le regroupement de relations diverses : du syntagme à sa réduction, de la forme en usage à la forme officielle, de l'emprunt à la forme autochtone. Relevons également quelques cas de synonymies commerciale et historique. En revanche, la relation d'**antonymie** est très peu utilisée, même élargie aux contraires complémentaires et réciproques. Outre le caractère propre de cette relation, peu productive, cela s'explique par le fait que la relation d'antonymie se trouve en concurrence, dans notre description, avec celle d'**isonymie**.

Cette relation lexicale unit ce que l'ISO appelle des *notions coordonnées*, celles-ci étant définies comme des notions qui, dans un système hiérarchique, se situent « au même niveau qu'une ou plusieurs autres notions » (ISO, 1990 : 2). Sans entrer dans un débat terminologique, disons que nous appelons « relation d'isonymie » toute relation unissant deux unités mises en concurrence, le plus souvent de même niveau, sans que l'on puisse poser une hiérarchie valable selon tous les points de vue³.

Cette relation est fructueuse, dans la mesure où elle permet de recenser le plus directement les unités auxquelles le terme étudié s'oppose le plus directement dans un paradigme discursif. Dans notre corpus, les termes *néphélogétrie* (nephelê, nuage) et *turbidimétrie* (turbidus, trouble) sont isonymes et ne pourraient être englobés que par une paraphrase du type « technique de mesure de la turbidité » et pas par une unité lexicale attestée. L'analyse dégage là un archiséme sans archilexème correspondant.

Après le premier repérage, au sein d'une catégorie, que permet la relation d'hyperonymie, les renseignements paradigmatiques aident la construction du signifié. Hyponymes, isonymes, antonymes, synonymes permettent de situer la valeur linguistique du terme, voire de sa base lexicale lorsqu'une famille dérivationnelle existe. C'est là un postulat de travail.

Les relations lexicales peuvent également permettre de procéder à une levée d'homonymie quand l'histoire entérine des distinctions lexicales nouvelles. Notre corpus nous a ainsi conduits à distinguer deux entrées *fermentation* :

- *fermentation* 1 : fermentation anaérobie, sens historiquement premier ;
- *fermentation* 2 : fermentation aérobie, sens utilisé en milieu industriel.

Un tel éclatement de la description pourrait paraître excessif. À tout coup, il serait illégitime dans un dictionnaire de langue. Mais ici, l'histoire a vu évoluer les pratiques des bio-industriels et leurs discours, discours dans lesquels le terme *fermenta-*

3. Pour un développement illustré d'exemples tirés de notre corpus, cf. Assal *et al.*, 1992.

tion a pris deux significations différentes : le sens « aérobic », lié aux fermentations industrielles, est venu s'installer à côté de celui d'« anaérobic ».

L'évolution des techniques a suscité la circulation concurrente des deux emplois. Il s'est d'abord agi de spécifications locales, liées à des textes ; mais l'apparition régulière de tels contextes a fini par modifier le sens en discours de ces deux types d'emplois. Ensuite, cette distinction, liée à la pratique technologique, s'est stabilisée en différence de signification.

Cette signification s'est lexicalisée et, aujourd'hui, l'utilisation d'un syntagme comme *fermentation continue* est licite, mais sa compréhension suppose le recours à *fermentation 2*, car cette sorte de fermentation a toujours lieu en présence d'oxygène. En termes de sémantique structurale, on peut diagnostiquer ici une différence de sèmes inhérents.

Dans cet exemple, la divergence lexicale est attestée également dans le paradigme dérivationnel puisque l'on trouve le nom d'instrument *fermenteur* qui sert à désigner des appareils liés à la fermentation aérobic (fermentation 2). Au sein de cette sphère d'activité, on a bien deux signifiés qui s'opposent. Il y a eu retombée de la distinction des deux signifiés sur la famille dérivationnelle. On voit là une illustration de la pertinence des paradigmes dérivationnels, tels que les a mis en lumière Louis Guilbert (1975 : 173 et suiv.), et dont Pierre Lerat a montré toute la pertinence pour les langues spécialisées (Lerat, 1995).

La distinction des signifiés peut être vue comme une lexicalisation d'une différence conceptuelle, lorsqu'un continuum existe entre deux signifiés. Ainsi, dans le cas du terme *viscosité*, la propriété dénotée a d'abord été spécifiée dans des discours étudiant les propriétés des fluides, en restant encore rattachée au trait « caractère visqueux, épais ». Ensuite, au XIX^e, elle s'est isolée pour désigner une propriété mesurable. C'est alors, quand on put isoler une grandeur quantifiable de façon mathématique par des appareils lentement perfectionnés, que les hommes de l'art usèrent de termes nouveaux et propres en recourant à une dérivation basée sur le radical *viscos-* et n'intéressant que le sens physique de *viscosité* : on vit alors apparaître des termes tels que *viscosimètre* en 1831, *viscosimétrie* en 1933, etc.

C'est dans cette logique de la retombée d'une distinction conceptuelle sur le système de la langue que nous avons été conduits à distinguer sous deux entrées le sens de *viscosité*¹ lié au mot visqueux, « gluant », et synonyme d'*adhésivité*, rendu par *Klebrigkeit* en allemand et celui de la *viscosité*², conçue comme propriété de certains fluides que mesure le viscosimètre. L'existence d'un nom d'appareil de mesure vient en quelque sorte attester au plan dérivationnel de la spécificité conceptuelle du second terme.

D'ailleurs, la distinction que nous avons opérée au plan paradigmatique se retrouve au plan syntagmatique, puisque typiquement la *viscosité*² mesurable est celle des fluides, et non la viscosité du « chapeau de certains bolets », « de la main d'œuvre » ou « du marché des capitaux », comme nous le suggère le *Nouveau Petit Robert*, pour ne rien dire de la redoutable « viscosité mentale » qu'il mentionne.

Le lien entre *viscosité*² et *fluide* est une relation que nous dirons prédicative.

Les relations prédicatives

Biolex présente quatre rubriques descriptives consacrées aux relations prédicatives, ce sont les rubriques « action typique », « objet typique », « agent typique » et « application typique ». Les rubriques prédicatives que nous avons utilisées sont inspirées des fiches proposées par Pierre Lerat pour les travaux du Centre de terminologie et de néologie dont il a été le fondateur. Son idée initiale était d'explorer le rendement des propriétés de :

prédicat privilégié (« action typique ») et d'argument logique contigu (« objet connexe ») dans une relation prédicative de premier ordre. Dans la rubrique « objet », il s'agit de repérer les notions qui sont associées le plus spontanément par les experts (la cause, le prévenu, le tribunal, autour de l'action « procès », pour le juge, par exemple) (Lerat, 1988 : 18).

C'était là permettre d'enregistrer des faits syntagmatiques significatifs et non plus seulement des faits paradigmatiques.

Nous avons conservé ces rubriques d'action et d'objet, dénommées « action typique » et « objet typique », en les élargissant, sous une terminologie commune, aux relations d'« agent typique » et d'« application typique ».

Ces quatre rubriques permettent l'enregistrement de renseignements d'ordre syntagmatique assez variés. De tels éléments ne tombent pas du ciel : ils sont repérés dans des discours et l'on va recenser les plus fréquents, les plus centraux – bref, les plus **typiques**. Après le temps de la sélection, vient le temps de la combinaison.

Ces rubriques permettent donc de recenser les cooccurrents les plus fréquents, et donc les éléments de discours les plus centraux, les plus courants. Si les premières rubriques permettaient de cerner la signification, les relations prédicatives permettent de mettre en lumière le sens le plus fréquemment produit dans une formation discursive.

Pourquoi parler de « formation discursive » ? Pour insister sur le fait qu'il s'agit de rendre compte d'un ensemble de discours tenus au sein d'une sphère d'activité, cet ensemble n'étant pas réductible à des découpages institutionnels (cf. Gaudin, 1993). En fait, comme le souligne Michel Foucault, « on ne peut pas établir de relation bi-univoque entre les disciplines instituées et les formations discursives » (1969 : 233). C'est là un des intérêts que présente la description d'un vocabulaire comme celui que rassemblent les bio-industries. Mais revenons aux relations prédicatives.

En se plaçant sur le plan syntagmatique, on est passé de la signification stable au sens produit qui va permettre la construction du concept.

Prenons les arômes en exemple. Tel qu'utilisé en bio-industries, *arôme* ne présente qu'une famille dérivationnelle. Nous en concluons à l'existence d'un seul signifié. Mais ce signifié unique permet de désigner plusieurs concepts. En effet, les arômes peuvent être distingués selon leurs actions typiques, *aromatisation* ou *fumaison*, l'objet typique restant identique : *denrées alimentaires*. Ici deux rubriques seulement sont intéressantes car l'agent typique n'est autre que le vocable étudié et l'application typique se situe à un niveau peu pertinent puisque cette rubrique renvoie

aux industries alimentaires. On est dans un cas de figure où l'application typique rejoint le domaine.

Pareillement, pour la chymosine les deux rubriques renseignées – l'action typique, *coagulation*, et l'objet typique, *lait* – permettent d'en construire le sens utile : la chymosine est l'enzyme permettant la coagulation du lait.

L'utilisation de ces rubriques n'est pas toujours simple. Tout d'abord, l'**action typique** n'est pas nécessairement unique ; ainsi, l'agitateur-disperseur sert, comme son nom l'indique, à plusieurs actions. Mais la morphologie est trompeuse, ou réductrice, car les actions recensées sont les suivantes : *mélange*, *dissolution*, *mise en suspension*, *homogénéisation*, *aération*. Il y a là une pratique implicite sous le terme qui ne peut se réduire à *agiter* et *dispenser*. Les éléments typiques permettent donc de donner des éléments de construction du concept et de les distinguer du niveau de la motivation morphologique.

De même, le *disperseur* sert à *homogénéiser*, *émulsifier* et *dispenser*, ce dernier verbe ne suffisant pas à lui seul pour rendre compte des utilisations de l'appareil.

La difficulté peut tenir à une lacune de métalangage. Ainsi, l'action typique de l'*antimousse* est de « détruire les bactéries ou à empêcher leur développement ». Ici manque un verbe hyperonyme permettant de neutraliser l'opposition entre « détruire » et « empêcher le développement ».

Il peut exister une corrélation entre la variété des actions typiques et des applications typiques. Ainsi, la *gomme xanthane* est utilisée dans les :

{app. typ} industries chimiques, comme
{act. typ} émulsifiant, lubrifiant

et dans les :

{app. typ} industries alimentaires, comme
{act. typ} additif.

La pluralité qui caractérise certaines actions typiques se retrouve pour les **objets typiques**. Il en est ainsi lorsque les utilisations d'un même microorganisme dans différents secteurs obligent à introduire une polysémie au sein d'une entrée unique. On est alors conduit à tourner le dos à la démarche homonymique qui caractérise souvent les travaux terminologiques. C'est ainsi que pour la bactérie lactique nous avons dû décrire quatre actions et objets typiques :

{act. typ} 1. acidification
{obj. typ} 1. fromages

{act. typ} 2. protéolyse
{obj. typ} 2. précurseurs d'arômes

{act. typ} 3. production de polysaccharides
{obj. typ} 3. laits fermentés

{act. typ} 4. fermentation malo-lactique
{obj. typ} 4. vins et cidres

Et une présentation souple comme celle-ci permet de présenter la diversité des applications en évitant la lourdeur et l'arbitraire qui conduit à multiplier les homonymes. En fait, *bactérie lactique* renvoie à une catégorie qui inclut toute une taxonomie de bactéries lactiques différentes. [Dans un dictionnaire plus ciblé, les bactéries feraient l'objet d'un traitement homonymique. La description sémantique se fait en fonction de l'univers de discours considéré.]

La notion d'**agent typique** pose des difficultés de description liées à la notion même d'agent, « être qui accomplit l'action exprimée par le verbe » (Dubois *et al.*, 1994). En fait, l'agent peut être un animé ou un inanimé, un agent direct ou indirect. L'action n'est que rarement accomplie de façon directe : si la *calmoduline* est bien, simple protéine, l'être qui accomplit l'action de *fixer les ions calcium*, cette rubrique pose souvent des difficultés quasi ontologiques, et de diverses natures, notamment dans la description d'appareils et de techniques.

Ainsi l'*électrophorèse en gel de polyacrylamide* utilise le pouvoir de séparation de ce gel. Notre grille de description nous a conduit à poser comme agent typique le « pouvoir sélectif du gel de polyacrylamide ». Pareillement, l'*électrodialyse* décrite comme suit :

{act. typ} désionisation
{obj. typ} l'eau
{ag. typ} un champ électrique

a donc comme agent typique champ électrique. La notion d'agent se trouve ici pourvue d'une extension large et assez imprévisible, mais c'est bien l'histoire des techniques qui le veut ainsi.

Quant à l'*électroéluion*, qui sert à la récupération de protéines contenues dans un gel, elle utilise l'« effet du champ électrique dans la cuve d'électroéluion ». La frontière entre accomplir l'action, la provoquer ou la permettre est bien mince. Mais les scrupules du descripteur sont peu de choses et tiennent, ici, principalement à un cadre grammatical trop simple. Il faut donc faire jouer les catégories pour y faire entrer le réel.

La notion d'**application typique** nous est apparue utile en cours de travail afin de situer le niveau auquel intervenait l'action décrite. Son choix est lié à une attitude initiale de prévention contre celle de domaine. C'est le fait de ne pas recourir à la description en termes de domaines et sous-domaines qui nous a conduits à ajouter cette rubrique, plus indicatrice de la fonction que du domaine proprement dit.

Cette rubrique s'est avérée utile par sa souplesse dans la mesure où elle permettrait de viser le niveau pertinent, qu'il s'agisse d'une sous-application très locale ou au contraire d'un terme intéressant l'ensemble d'un domaine.

Elle s'est avérée également précieuse notamment quand les utilisations sont diverses : par exemple, la dextrane sert aussi bien dans la fabrication de résines que

dans la préparation de plasma. Il y a là des utilisations de ce polysaccharide très différentes et bien plus précises et limitées que chimie, santé ou hématologie. Et nous évitons encore l'artifice qui nous eût conduits à distinguer deux dextrans, là où il existe simplement deux conceptualisations liées à des pratiques différentes.

Ces quatre rubriques nous permettent donc de dégager des réalisations discursives différentes qui, selon nous, tiennent à des concepts différents. On en trouve illustration dans des divergences de traduction pour un signifié unique en français. Ainsi, *arôme* possède deux actions typiques : la fumaison et l'aromatisation. Et l'on retrouve cette distinction dans les traductions :

arôme de fumée = *smoke flavour*, et
arôme de transformation = *transformation aroma*.

On voit ici que les signifiés sont bien culturels et que les unités de traduction nécessitent des distinctions proches permettant de mettre en évidence le niveau conceptuel. Autre exemple patent, les divers *agitateurs* recueillis dans notre corpus appartiennent à une seule et même classe lexicale, classe qui se scinde en deux au contact de l'anglais pour lequel est pertinente l'opposition entre *stirrer* et *shaker*, selon le mode d'agitation. Mais, bien entendu, le niveau proprement conceptuel se situe à un niveau plus fin que le niveau lexical, celui dont rendent compte les catégories prédicatives.

Les relations que nous venons de décrire ont pour but de faciliter le passage de la description du signifié au concept par un élargissement d'une description purement oppositive, celle des paradigmes, à une description positive, celle des relations syntagmatiques attestées. Les relations prédicatives permettent d'aboutir à un noyau prédictif, de décrire des phrases de base élémentaires, telles que :

- « La bioabsorption permet le stockage de métaux toxiques par des micro-organismes pour la dépollution », ou
- « La chambre d'électroporation permet la fragilisation par un faible courant électrique de membranes cellulaires pour la fusion cellulaire ».

C'est assez élémentaire, certes, mais cela ne se devine pas.

Pour une approche contextuelle

Nos relations prédicatives jouent en fait le rôle de définitions minimales, par l'exploitation des renseignements contenus dans les énoncés dépouillés. On s'est ainsi rapproché de la « définition naturelle », tout en écartant les risques liés aux discours métalinguistiques. Mais le lecteur a besoin de renseignements plus nombreux afin d'esquisser la construction de la notion. Ces renseignements, il les obtient à la lecture des contextes qui constituent le troisième volet de notre description.

Les avantages des contextes sont nombreux et connus. On l'a vu, ils autorisent une description en catégories plus ouvertes. Abordant ce travail en sociolinguistes, nous souhaitons éviter le caractère artificiel de la position métalinguistique, et nos contextes sont des énoncés spontanés. Par ailleurs, la variété des énoncés recueillis

nous permettait de faire place à la variation. C'est, là encore, un argument de type sociolinguistique, et non normatif. Une option reste certes une option ; mais nous pensons que, même en se penchant sur des communautés professionnelles apparemment homogènes, il ne faut pas oublier que la variation est un facteur majeur, présent dans toute communauté de parole. Enfin, on sait l'intérêt des contextes pour permettre de rendre compte de la variété des points de vue exprimés sur la notion, le référent, le processus, etc. Ce sont là des questions qui renvoient à la problématique de la catégorisation.

De façon plus modeste, nous voudrions souligner des contextes en ce qu'ils permettent souvent une construction contrastive des concepts. Pour ce faire, ils jouent fréquemment sur la relation d'isonymie. Ainsi le contexte proposé sous *chémostat* présente ensemble le *chémostat* et le *turbidostat* :

chémostat

{ctx1} [...] **Dans un turbidostat**, la vitesse à laquelle les cellules quittent le réacteur (mesurée par la turbidité ou l'opacité du flux sortant) commande l'introduction des substances nutritives. **Dans un chémostat**, la concentration d'une substance nutritive critique, présente dans le flux entrant, est calculée de façon à contrôler la vitesse de la réaction, en limitant la prolifération des micro-organismes : c'est le contrôle par ajustement préalable.

Pareillement, les contextes de *colorimètre* permettent d'opposer ce terme à *spectrophotomètre* et *photomètre* :

colorimètre

{ctx1} À la différence du spectrophotomètre, de bande passante réduite, dans le colorimètre, la sélection de la longueur d'onde se fait par filtres colorés interchangeables, et non par monochromateur [...].

La relation d'isonymie, utilisée ici, autorise des stratégies cognitives économiques et efficaces pour construire des catégories. En effet, elle permet de partir de ce qui est le plus proche au lieu de recourir au genre prochain. Et, même si cela est préférable, il n'est pas nécessairement besoin que la notion voisine soit beaucoup plus familière, car en fait, on peut arriver à construire deux notions de façon conjointe. Nous avons déjà rencontré cette stratégie dans des dictionnaires de sciences (cf. Gaudin, 1992).

Les éléments qui précèdent s'en tiennent à des problèmes de description. Il sont peu reliés aux travaux, à caractère plus théorique, évoqués en introduction. Se pose en fait à nous la question de la possibilité d'un travail terminographique qui s'inspire de positions socioterminologiques. Cette possibilité est attestée pour les travaux relatifs aux politiques linguistiques : en témoigne l'effort méthodologique développé principalement au Québec et, depuis peu, en France.

En revanche, de récents débats l'ont illustré, la sociolinguistique a été de peu de secours à la lexicographie, concernant « le contexte social et situationnel dans lequel chacun des mots peut apparaître. La sociolinguistique n'a pas de solution à proposer aux lexicographes » sur ce point, note Vachon-L'Heureux (1995 : 2). En fait, les praticiens sont un peu réduits à, par exemple, proposer l'exclusion de la marque populaire, auquel cas « la suite privilégiée serait soutenu, neutre, familier et très familier » (Vachon-L'Heureux, 1995 : 3)

Que le problème ait un niveau de pertinence, cela n'est pas douteux : le mouvement vers une normalisation des produits imposera vraisemblablement des ouvrages « lexicographiquement corrects ». Mais, faute d'assises théoriques, la description des variétés sociolectales est restée rudimentaire eu égard au nombre des études sociolinguistiques menées depuis vingt ans. Les descriptions des variations diachroniques et diatopiques se sont affinées alors que force est de constater que les progrès ont été fort minces concernant les faits diastratiques.

Cette lexicographie particulière qu'est la terminographie pour les agnostiques du concept peut-elle tirer profit de l'application de principes méthodologiques issus de la sociolinguistique ? L'exemple de la lexicographie englobante, ou générale, montre que les faits sociolinguistiques ne se laissent pas décrire à peu de frais. En revanche, les technolectes ou sociolectes professionnels, ou langues spécialisées, permettent d'étudier des communautés linguistiques aux effectifs limités. Cela permet, notamment, de travailler finement les questions d'individuation sociolinguistique, mais aussi d'étudier la circulation des vocables, en recourant à des typologies fonctionnelles assez fines.

Mais plus profondément, il reste que le but d'un dictionnaire est de satisfaire des usagers. Les changements des habitudes dans ce domaine sont difficiles et il n'est pas sûr que les insuffisances mises en évidence par la réflexion soient ressenties par les utilisateurs. Mais la période de formidable mutation technologique que connaît l'art lexicologique devrait favoriser les propositions théoriquement stimulantes.

Le lexique génératif :

Une alternative au traitement de la polysémie

Le cas des adjectifs qui dénotent un état mental

Pierrette BOUILLON

ISSCO (Institut pour les Études Sémantiques et Cognitives), Université de Genève, Suisse

1. Introduction

Récemment, les travaux en sémantique lexicale ont connu des développements intéressants : motivées par des soucis de cohérence du lexique, des théories nouvelles commencent enfin à voir le jour et à exploiter des descriptions sémantiques plus expressives, ainsi que des méthodes de composition plus puissantes (voir notamment Pustejovsky, 1995 et Briscoe et Copestake, 1993).

Cet article exploite l'une de ces théories : celle du Lexique Génératif (*Generative Lexicon*, GL) (Pustejovsky, 1991 et 1995), en l'étendant au traitement de la polysémie des adjectifs français qui dénotent un état mental. L'article commence par cerner les problèmes posés par ce type d'adjectifs (section 2). Il oppose ensuite GL à l'approche lexicographique traditionnelle (section 3). Enfin, il conclut en montrant l'adéquation de GL pour le traitement des adjectifs d'état mental (sections 4 et 5).

2. La polysémie des adjectifs d'émotion et orientés-agent

Les adjectifs d'état mental qui dénotent un état émotionnel (exemples (1)) et une compétence ou un comportement (« orientés-agent », selon la dénomination de Ernst, 1984) (2) présentent un comportement polysémique remarquable, bien mis en évidence dans la littérature (notamment dans Lehrer, 1990 et Croft, 1984).

(1) *triste, irrité, déprimé, gai, etc.*

(2) *intelligent, habile, ingénieux, stupide, doué, grossier, poli, impoli, etc.*

a. Tout d'abord, la plupart d'entre eux peuvent modifier des noms de types sémantiques « humain », « objet » et « événement », sans que ce comportement ne soit généralisable à tous les adjectifs de la classe : « *doué* » et « *irrité* », par exemple, constituent des exceptions à cet égard (4).

- | | | | |
|-----|------------------------------------|--|---|
| (3) | (humain)
(objet)
(événement) | a. <i>un homme triste</i>
b. <i>un livre triste</i>
c. <i>un examen triste</i> | a. <i>un homme ingénieux</i>
b. <i>un livre ingénieux</i>
c. <i>un examen ingénieux</i> |
| (4) | (humain)
(objet)
(événement) | a. <i>un homme irrité</i>
b. * <i>un livre irrité</i>
c. * <i>un examen irrité</i> | a. <i>un homme doué</i>
b. * <i>un livre doué</i>
c. * <i>un examen doué</i> |

b. Ensuite, ils présentent des sens différents en fonction du type sémantique modifié : quand ils sélectionnent un nom de type « humain », ils dénotent normalement l'état dans lequel se trouve l'individu (5) ;

- | | | | |
|-----|-----------------------------|---|------------------------------------|
| (5) | <i>un homme triste</i> | → | qui est dans l'état de tristesse |
| | <i>un homme intelligent</i> | → | qui est dans l'état d'intelligence |

quand, au contraire, ils modifient un événement ou un objet, ils présentent, soit un sens « causatif » (6b) (dans ce cas, l'objet ou l'événement est la cause de l'état), soit « explicatif » (6c) et (7c) (dans ce cas, l'objet ou l'événement dénote la manifestation de l'état). Certains peuvent combiner les deux sens, comme « *triste* » en (6b,c). Enfin, dans des contextes marqués, le sens « causatif » est aussi possible avec des noms de type « humain » (8).

- | | | | |
|-----|------------------------------------|---|---|
| (6) | <i>un livre/voyage triste</i> | → | a. *qui est dans l'état de tristesse
b. qui cause la tristesse de quelqu'un
c. où se manifeste la tristesse de quelqu'un |
| (7) | <i>un livre/voyage intelligent</i> | → | a. *qui est dans l'état d'intelligence
b. *qui cause l'intelligence de quelqu'un
c. où se manifeste l'intelligence de quelqu'un |
| (8) | <i>un homme triste à voir</i> | → | a. *qui est dans l'état de tristesse
b. qui cause la tristesse de celui qui le voit |

3. Approches monomorphiques et semi-polymorphiques

Pour représenter la polysémie de ces adjectifs, deux approches opposées peuvent être envisagées : l'approche monomorphique ou celle, semi-polymorphique, du « Lexique Génératif » de Pustejovsky.

a. L'approche traditionnelle, monomorphique, consiste à créer dans le lexique une entrée différente pour chaque type sémantique dénoté par le mot ; ainsi, elle créera au moins deux entrées différentes pour chaque adjectif de (1) et (2), comme en (9) pour « *triste* » :

- (9) a. *triste* (+ humain) : qui est dans un état de tristesse
b. *triste* (+ humain/objet/événement) : qui cause un état de tristesse
c. *triste* (+ objet/événement) : où se manifeste un état de tristesse

Cette approche pose évidemment un certain nombre de problèmes. Tout d'abord, elle postule qu'il est possible d'énumérer toutes les interprétations des mots, ce qui ne rend pas compte de la création des sens en contexte. Certains adjectifs changent en effet de sens en fonction du nom modifié, comme « *rapide* » qui a un sens différent dans « *une dactylographe rapide* » (qui exécute avec promptitude), « *un pas rapide* » (qui s'accomplit rapidement), « *un cheval rapide* » (qui se meut rapidement), « *un moyen rapide* » (qui conduit vite au but escompté) ou « *une piste rapide* » (qui permet de hautes performances). Comment, dans ce cas, énumérer tous les sens du mot ? Ensuite, elle ne rend pas compte du fait que certaines ambiguïtés sont régulières, comme celles des adjectifs d'état mental. En troisième lieu, elle considère les entrées comme indépendantes l'une de l'autre et n'explique pas le lien qui existe entre certaines d'entre elles, comme par exemple entre les deux sens de « *triste* », l'état de tristesse et sa cause. Enfin, elle crée une ambiguïté, difficile à résoudre dans la suite du traitement. Si on crée, par exemple, deux entrées pour « *cuire* », pour distinguer les deux interprétations de (10), il sera très difficile de faire dans la suite la bonne sélection puisque les deux sens sont très proches et se chevauchent.

- (10) a. *Jean cuit les pommes de terre* (changement d'état, sans création)
b. *Jean nous a cuit un cake* (création)

b. L'approche semi-polymorphique de Pustejovsky (Pustejovsky, 1991 et 1995) s'oppose à une énumération de types sémantiques. Elle propose de donner à chaque mot une riche représentation, qui constitue une réserve de types, en indiquant pour chacun d'eux les différentes extensions de sens possibles. Pour se faire, elle rompt avec l'approche monomorphique de deux manières : tout d'abord, elle incorpore les différents types sémantiques dénotés par le mot dans une méta-entrée, qui indique les relations qu'ils entretiennent entre eux. Ensuite, elle définit des mécanismes génératifs capables d'opérer sur la méta-entrée, pour changer éventuellement sa dénotation en contexte. La suite de ce chapitre se concentre sur la méta-entrée et les mécanismes génératifs, puis la représentation de la polysémie nominale en GL.

3.1. La méta-entrée dans le GL

Dans GL, tous les mots sont représentés dans le lexique avec une structure complexe. Celle-ci peut combiner jusqu'à trois niveaux de représentation, connectés entre eux dans un réseau hiérarchique (Pustejovsky, 1991 et 1995).

a. La structure argumentale définit le nombre et le type d'arguments du mot. « *Construire* », par exemple recevra trois arguments (11) : un sujet de type « animé » (« arg1 ») et un objet « artefact » (« arg2 »), ainsi qu'un troisième argument (« D-arg1 ») de type « matériel » qui diffère des deux autres puisqu'il ne doit pas nécessairement être exprimé au niveau syntaxique (11) (« *default argument* » D-ARG) ;

- (11) argstr = arg1 = animé
 arg2 = artefact
 D-arg1 = matériel

b. La structure événementielle spécifie le type d'événement d'un mot ou d'une phrase (état, procès ou transition), ainsi que sa structure interne ; « *construire* », par exemple, sera défini comme une transition, composée de deux sous-événements : un procès initial (l'acte de construire) (« E1 ») et un état résultatif (12) (celui de l'objet construit) (« E2 »). La tête (« *head* ») indique le sous-événement le plus important dans la structure, « E1 » dans le cas de « *construire* » ; « E2 » dans celui de « *mourir* ».

(12) evenstr = E1 = procès
 E2 = état
 head = E1/E2¹

c. La structure des qualia décrit les caractéristiques sémantiques du mot, en indiquant comment les événements et les arguments sont sémantiquement liés. Pour ce faire, elle utilise quatre rôles (13), définis comme suit : formel (quelle est sa catégorie sémantique ?), constitutif (quels sont ses éléments constitutifs ?), agentif (comment il est créé ?) et télique (quelle est sa fonction ?)

(13) qualia = const = ...
 form = ...
 telic = ...
 agentif = ...

« *Couteau* » (14), par exemple, dénote un artefact. Il a un télique : sa fonction est de « *couper* », ainsi qu'un agentif : il est créé par un être humain.

(14) couteau
 argstr = arg1 = x :artefact
 qualia = artefact-lcp
 form = x
 telic = couper(e1,x,y)
 agentif = créer(e2,z,x)

Dans la méta-entrée, il est important de distinguer les types de base d'un item (qui définissent son « paradigme lexical conceptuel » LCP), des autres types qui ne sont accessibles que par des opérations génératives. « *Couteau* », par exemple, en (14), dénote un « artefact », mais d'autres types, comme les événements « e1 » et « e2 », pourront être accessibles par une des opérations génératives suivantes (Pustejovsky, 1995 : chap. 7).

a. La coercion de types permet de remplacer le type sémantique d'un item lexical par celui qui est requis par le prédicat, pour autant que ce dernier soit présent dans la qualia de l'item. « *Commencer* », par exemple, sélectionne un objet de type « événement ». Dans « *cette usine vient de commencer les couteaux de luxe* », « *couteau* », de type artefact, peut prendre le type « événement » requis par « *commencer* », puisque sa structure des qualia contient des événements : « *cette usine vient de commencer les couteaux de luxe* » signifie entre autres « *cette usine vient de commencer à fabriquer des couteaux de luxe* » (voir Pustejovsky et Bouillon, 1995, pour plus de détails).

¹ La barre oblique «/» indique une disjonction

b. Le liage sélectif permet à un item lexical d'opérer sur un type qui ne se trouve pas dans le paradigme lexical conceptuel de l'item qu'il modifie. Dans « *une voiture rapide* », par exemple, « *rapide* » sous-modifie l'événement défini dans le télique du nom, mais sans changer sa dénotation, comme en (15) : « *une voiture rapide* » signifie « *une voiture qui roule rapidement* » (Pustejovsky et Bouillon, 1995).

(15) $\lambda x [\dots \text{Telic}(x) = \lambda e [\text{rouler}(x)(e) \text{ et } \text{rapide}(e)]$

c. La cocomposition est une fonction bilatérale qui permet à un argument de changer la sémantique du prédicat, en composition. Elle n'intervient pas dans le traitement des adjectifs d'état mental et nous ne nous y attarderons pas ici.

3.3. Traitement de la polysémie nominale en GL

GL distingue au moins deux niveaux polysémiques différents.

a. Les noms non-polysémiques dénotent un type simple, comme « *couteau* » en (14). Ces derniers ont un seul argument (« *arg1* »), qui correspond directement au formel (« *x* ») (16).

(16) alpha
 argstr = arg1 = x:t
 qualia = t-lcp
 form = x
 const = ...
 telic = ...
 agentif = ...

b. Les noms polysémiques font référence à plusieurs types en même temps. Ils dénotent un type complexe (« *pointé* ») qui correspond au produit cartésien des types définis dans la structure argumentale (noté *t1.t2*) (17). Dans ce cas, le mot a deux arguments (« *arg1* » et « *arg2* »), dont la relation est définie au niveau du formel (ici relation P). Il dénote trois types de base différents, qui peuvent être projetés indépendamment : le type pointé « *t1.t2* » et les deux sous-types « *t1* » et « *t2* » (17).

(17) alpha
 argstr = arg1 = x:t1
 arg2 = y:t2
 qualia = t1.t2-lcp
 form = P(x,y)
 const = ...
 telic = ...
 agentif = ...

Le « *livre* » appartient à cette catégorie (18). Il dénote le type complexe « *contenu-imprimé* » (« *info.obj-phys* ») qui est le produit des deux types « *info* » et « *obj-phys* ». Il présente donc deux arguments, dont le formel indique la relation : « *y* » contient « *x* ». Sa fonction est d'être lu et il est créé par l'acte d'écriture. À partir de cette méta-entrée, il sera possible de projeter indépendamment les trois types de base : le type pointé (« *contenu-imprimé* ») et les deux sous-types « *info* » et « *obj-phys* » (19a,b,c).

Les deux types « e1 » et « e2 » (événement) pourront aussi être accessibles par des mécanismes génératifs, comme la coercition ou le liage sélectif (19d,e) (Pustejovsky et Bouillon, 1995).

(18) livre

argstr = arg1 = x:info
 arg2 = y:obj-phys
qualia = contenu-imprimé_lcp
 form = contenir(y,x)
 const = ...
 telic = lire(e1,w,x,y)
 agentif = écrire(e2,v,x,y)

- (19) a. *je lis un livre de Proust* (« contenu-imprimé »)
 b. *je ne m'intéresse pas à la couverture, mais au livre* (« info »)
 c. *je ne m'intéresse pas au contenu, mais au livre* (« obj-phys »)
 d. *je commence un nouveau livre* (« événement »)
 e. *c'est un long livre* (« événement »)

La thèse que nous voudrions défendre dans le chapitre suivant est que les adjectifs d'état mental dénotent, eux aussi, des types complexes. Ces derniers expliquent le comportement polysémique esquissé dans la section 2.

4. Traitement des adjectifs d'émotion et orientés-agent dans GL

4.1. Proposition générale

Nous proposons le traitement global suivant :

- distinguer deux types d'adjectifs d'état mental par leur structure de qualia² : les adjectifs causatifs (d'émotion) (exemples (1)) et orientés-agent (2) ;
- représenter la polysémie de ces adjectifs au niveau de leur structure des qualia, par des types complexes ;
- expliquer les différences de sélection à l'intérieur des classes par la notion de projection et de tête.

Les deux premiers points seront l'objet de la section 4.1.1, le troisième de 4.2.2., 4.2.3 examinera ensuite plus en détail le comportement des adjectifs causatifs.

4.1.1. Deux sortes d'adjectifs, avec type complexe

Les adjectifs de (1) et (2) recevront respectivement les représentations (20) et (21). Ceux qui peuvent appartenir aux deux classes (voir section 2) seront codés avec la structure (22) qui n'est pas spécifiée quand au second rôle : téléique ou agentif.

2 Ces deux types d'adjectifs correspondent aux classes 3 et 1 de Picabia, 1978

(20) adj_caus

eventstr = E1 = e1:état
E2 = e2:AGENTIVE_INTELLECTUAL-événement
head = e1/e2/e1,e2
argstr = arg1 = x:humain
D-arg1 = e2
qualia = e1.e2_lcp
form = Adj(e1,x)
agentive = P(e2,x,...)

(21) adj_agent-orienté

eventstr = E1 = e1:état
E2 = e2:AGENTIVE_INTELLECTUAL-événement
head = e1/e2/e1,e2
argstr = arg1 = x:humain
D-arg1 = e2
qualia = e1.e2_lcp
form = Adj(e1,x)
telic = P(e2,x,...)

(22) adj_causatif-agent-orienté

eventstr = E1 = e1:état
E2 = e2:..._EXPERIENCER-événement/..._INTELLECTUAL-événement
head = e1/e2/e1,e2
argstr = arg1 = x:humain
D-arg1 = e2
qualia = e1.e2_lcp
form = Adj(e1,x)
agentif/telic = P(e2,x,...)

Ces structures indiquent :

a. que ces adjectifs exigent deux arguments : un de type « humain » (« arg1 ») et, un second (« D-arg1 »), facultatif au niveau syntaxique, événementiel, de type « perceptif » actif ou statif³ pour les adjectifs causatifs et qui dénote un « acte intellectuel » pour les adjectifs orientés-agent (voir Croft, 1984 : 21, pour une proposition similaire) ;

b. qu'ils font référence à deux types sémantiques différents (« e1.e2 »), dont la structure des qualia indique la relation (tout comme pour « livre ») :

i. un état résultatif (« e1 ») (encodé dans le formel) (état de tristesse, d'intelligence, d'ingéniosité, etc.) ;

ii. un événement (« e2 »). Encodé dans le rôle agentif (20), il dénote alors la cause de l'état émotionnel et doit être de type perceptif, statif ou actif : comme Croft (1990), nous considérons en effet qu'il y a deux processus impliqués dans un état émotionnel : l'expérimenteur porte son attention, intentionnellement ou non, sur un stimulus, qui suscite ensuite un état mental ; dans le rôle télique au contraire (21), l'événement dénote la manifestation de l'état.

3 Sur cette distinction, voir Lehrer, 1990.

En d'autres termes, (20) signifie que quelqu'un (« x ») est dans un certain état (« e1 ») à cause d'un événement (« e2 »); (21) que l'individu (« x ») est dans un état (« e1 ») dont l'événement (« e2 ») est la manifestation.

4.1.2. La notion de tête (« head »)

Tous les adjectifs ne vont cependant pas pouvoir projeter indépendamment les deux types qu'ils dénotent, en fonction de leur tête (voir chapitre 3.1.), qui va jouer le rôle de filtre au niveau de l'interface syntaxe/sémantique (Pustejovsky, 1995 : chap. 6). En (20) et (21), il apparaît que trois types de configuration sont possibles, qui déterminent autant de projections différentes.

a) Tête sur « e1 » (« head = e1 ») : l'adjectif est projeté via le formel et il en résulte un sens « statif » (exemples (23) et (24)) ; dans ce cas, l'adjectif a un seul argument obligatoire « x » de type « humain ». Un second, de type « événementiel », est cependant possible s'il fait référence à l'agentif (23) ou au télique (24). La différence au niveau du type des événements suffit à expliquer des comportements spécifiques au niveau de la complémentation des adjectifs (1) et (2) :

i. d'une part, les adjectifs causatifs partagent, avec les autres prédicats « émotifs » (« regretter », etc.), la particularité de pouvoir apparaître avec un complément subjonctif ou infinitif, en distribution complémentaire (23) (« obviative phenomenon ») (Kiparsky et Kiparsky, 1979 : 169 et Rochette, 1988 : 251) : dans ce cas, l'infinitive sature directement le prédicat « P(e2,x,...) » de l'agentif, tandis que la complétive au subjonctif est interprétée comme l'objet (« y ») d'un événement perceptif sous-entendu (en (23b), je suis triste suite à « la perception de ton départ » « Exp(e,je,que-tu partes) »).

- (23) a. *je suis triste de partir*
 b. *je suis triste que tu partes*
 c. **je suis triste que je parte*

ii. D'autre part, la plupart des adjectifs orientés-agents permettent, comme second argument, une infinitive introduite par « à » (A-VP) (24b) ou « de » (24a) (DE-VP), le A-VP saturant directement « e2 », tandis que le DE-VP étant considéré comme l'objet (« y ») de l'acte intellectuel (en (24b), « je suis ingénieux de prendre l'initiative de partir »).

- (24) a. *je suis ingénieux de partir*
 b. *je suis ingénieux à jouer aux échecs*
 c. **je suis ingénieux que tu partes*
 d. **je suis ingénieux que je parte*

b) Tête sur « e2 » (« head = e2 ») : l'adjectif est projeté via l'agentif ou le télique et il en résulte un sens « causatif » ou « explicatif » (25) ; dans ce cas, l'argument du nom est nécessairement de type événementiel (25a) ou bien est interprété comme tel au niveau sémantique (25b,c).

- (25) a. *une lecture triste* → une lecture qui cause la tristesse du lecteur

- b. *un livre triste* → un livre dont la lecture cause la tristesse du lecteur
- c. *un examen intelligent* → un examen dont la conception témoigne de l'intelligence de la personne qui l'a conçu

c) Pas de spécification de la tête (« head = e1, e2 ») : l'adjectif est projeté via le formel ou le télique/agentif, ce qui explique son ambiguïté.

La suite de cet article donne quelques exemples pour chacune de ces configurations.

4.1.3. De quelques exemples

Voyons d'abord plus en détail le cas de l'adjectif « *triste* », dans son sens causatif uniquement. Comme il n'est pas spécifié quant à la tête (26), il peut être projeté via le formel ou l'agentif.

- (26) triste
- eventstr = E1 = e1:état
 - E2 = e2:STATIF_ACTIF_EXPERIENCER-évén
 - head = e1, e2
 - argstr = arg1 = x:humain
 - D-arg1 = e2
 - qualia = exp-causative_lcp
 - form = triste(e1,x)
 - agentive = P(e2,x,...)

a. Quand le formel (« e1 ») reçoit la tête, la phrase exprime l'état de tristesse d'un individu, comme en (27a). Dans cet exemple, la cause de l'état n'est pas spécifiée : Jean est triste pour une raison non déterminée. Elle peut cependant être rendue explicite, comme en (27b,c) où le complément sature l'agentif.

- (27) a. *Jean est triste*
 b. *Jean est triste de lire ce livre*
 c. *Jean est triste que je lise ce livre*

b. Quand l'agentif (« e2 ») reçoit la tête, la phrase exprime la cause de l'état de tristesse d'un individu, comme en (28). Dans ce cas, le nom peut être :

- i. un nom événementiel, comme « *lecture* » en (28a) ;
- ii. d'autres types de noms, dont la structure des qualia indique une relation directe avec un événement de perception, comme « *livre* », en (28b) (dont la qualia contient les événements « *lire* » et « *écrire* » (18)). Dans ce cas, la perception est intentionnelle et contrôlée ;
- iii. d'autres types de noms, dont le type permet la mise en relation avec un événement perceptif statif générique, comme « *percevoir par ses sens* » (28c). Ici, au contraire, la perception n'est ni intentionnelle, ni contrôlée.

- (28) a. *une lecture triste*
 b. *un livre triste* → à lire ou à écrire
 c. *un sapin triste* → à percevoir pas ses sens

En cas d'ambiguïté entre les deux lectures (stative et causative), la syntaxe permet de distinguer les deux sens : en (29a), il s'agit de la place de l'adjectif et, en (29b), du choix de la préposition « à » (et non « de », comme en (27b)).

- (29) a. *de tristes enfants* → à percevoir par ses sens
 b. *un homme triste à voir*

Enfin, (30a,b) est impossible, puisque deux événements essayent de saturer le rôle agentif (« *lecture du livre* » et « *voir le film* » en (30a)). (30c), au contraire, est possible puisque « *enfant* » et « *mère* » saturent des variables différentes (« x » et « y »), comme explicité en (31).

- (30) a. **un livre triste à cause du film qu'il a vu*
 b. **de tristes enfants de la mort de leur mère*
 c. *des enfants tristes de la mort de leur mère*

- (31) qualia = exp-causative_lcp
 form = triste(e1,enfants)
 agentive = Exp(e2,enfants,mort_de_leur-mère)

D'autres adjectifs causatifs vont être spécifiés quant à la tête, ce qui explique un comportement différent de « *triste* » : les participes présents et la plupart des adjectifs en « *-able* » et « *-ible* », qui dérivent d'un verbe psychologique, auront la tête sur l'agentif ; les participes passés l'auront sur le formel. Cette restriction explique pourquoi certains adjectifs causatifs (1) ne peuvent modifier des noms de type « contenu-imprimé » (32), ainsi que le sens exclusivement causatif des adjectifs en « *-ant* » et « *-able* » (33).

- (32) **un livre ennuyé*

- (33) a. *un homme ennuyant* → qui cause de l'ennui
 b. *un livre agréable* → qui cause du plaisir

5. Conclusion

Dans cet article, nous avons étendu la théorie du Lexique Génératif au traitement des adjectifs d'état mental français. Nous avons montré ainsi l'élégance de l'approche semi-polymorphique, pour :

- éviter la prolifération des entrées : les différents sens de (1) et (2) sont intégrés dans une même méta-entrée ;
- expliquer les liens entre les différents sens : la structure des qualia rend explicites les liens entre les différents types sémantiques de (1) et (2) – en (1), l'événement cause l'état ; en (2), il en est la manifestation ;
- faire découler le comportement polysémique de (1) et (2) de leur représentation sémantique : les différences explicitées dans le chapitre (2) s'expliquent par la structure des qualia et la manière de la projeter ;
- distinguer deux sortes d'adjectifs, en fonction de leur dénotation : type simple ou complexe. Cette distinction reprend l'opposition traditionnelle entre adjectifs

« statiques » et « dynamiques » (dénotant des qualités sujettes au contrôle du possesseur) (Quirk *et al.*, 1994 : 434, par exemple), pour la caractériser et la représenter sémantiquement : un adjectif est dynamique s'il fait référence soit à ce qui crée l'état ou à sa manifestation ultérieure.

Remerciements

Nous remercions vivement James Pustejovsky et Laurence Danlos pour leurs commentaires sur une version antérieure de cet article.

Quand l'informatique tutoie le dictionnaire des difficultés de la langue française...

Daniel BLAMPAIN

Institut supérieur des traducteurs et interprètes (ISTI), Bruxelles, Belgique

Quand on consacre une partie importante de son temps à analyser les difficultés de la langue française et à les présenter au plus large public sous la forme d'un dictionnaire (Hanse et Blampain, 1994), s'il est bien une expression qui, appliquée à la langue française, fascine, c'est *se jouer des difficultés*, expression où le sens ludique du verbe l'emporte sur le sens d'effort et de peine inscrit dans la mémoire sémantique du substantif.

Jouant constamment des deux points de vue, celui du linguiste et celui de l'utilisateur, je m'efforce d'isoler et de décrire les difficultés afin de répondre au mieux à l'attente de l'utilisateur... qui se trouve en difficulté.

Les difficultés sont nombreuses. Un choix pertinent de celles-ci, un traitement qui conjugue l'analyse rigoureuse et la communication agréable déterminent la qualité et l'accueil de l'ouvrage. Difficultés liées à l'oral ou à l'écrit, difficultés phonétiques et orthographiques, grammaticales et lexicales, difficultés liées aux registres de langue ou aux français régionaux qui sont les joies de la francophonie, toutes sont traitées au sein d'articles qui occupent des espaces variables, proportionnels à la complexité des cas posés et qui déclinent tout sur le mode d'exemples multiples, au point que le critère de qualité est l'utilisation satisfaisante de l'ouvrage aussi bien par des francophones que par des allophones.

Et quand l'informatique, au service de l'utilisateur de la langue française, propose son aide en la matière... qu'en est-il de ce trait d'union placé entre deux outils ? Qu'en est-il de cette familiarité naissante entre le dictionnaire des difficultés et ce que, dans le cadre des industries de la langue, on appelle de manière austère les « correcteurs grammaticaux » ? Qu'en est-il de cette dictionnaire qui tente d'unir le lexique à la syntaxe ? À une époque où plus de 70 % du travail porte sur l'écrit, sur le traitement et l'échange de messages d'information, où, dans les secteurs spécialisés de la presse et de l'édition, il revient de plus en plus à chacun d'assumer la correction de son texte,

le métier de correcteur disparaissant en raison du coût de la révision (25 % de la saisie), l'utilisateur de l'informatique accueille avec beaucoup d'intérêt les outils d'assistance mis à sa disposition pour assurer la qualité de ses textes.

Je voudrais d'emblée limiter mon propos au correcteur dit de la 3^e génération, disponible sur ordinateur personnel courant. On désigne par là le correcteur qui repose sur un analyseur syntaxique, à la différence des correcteurs locaux qui travaillent sur des listes de mots et des accords grammaticaux localisés. Ma réflexion et mon expérimentation, qui ne se confond pas avec les tests traditionnels portant sur la vitesse d'exécution ou sur l'étendue du dictionnaire, ont principalement porté sur le *Correcteur 101* (logiciel Machina Sapiens - Copyright 1992-1994. Montréal), reconnu par l'ensemble des revues spécialisées pour ses performances exceptionnelles. Il ne s'agira pas non plus de comparer l'imprimé et l'informatisé. Les partages sont évidents. Le dictionnaire des difficultés traite de la langue orale, décrit la langue sémantiquement et diachroniquement, réduit l'arbitraire par l'ampleur de l'analyse et par la présentation de la diversité des usages. Il est en outre un dictionnaire d'auteur, c'est-à-dire qui a son style.

Notre objectif est plutôt de cerner la problématique déployée par le correcteur grammatical dans son évolution, sur le terrain commun des difficultés grammaticales, tant il est vrai que le « grammairien sous Windows », comme on l'a parfois appelé, peut nous aider à améliorer les algorithmes de reconnaissance syntaxique et la description de la combinatoire. Dans le domaine des industries de la langue, on a aujourd'hui compris qu'il est préférable de penser en termes de petits progrès répondant à des besoins précis plutôt qu'en termes de compétence linguistique générale. De son côté, le linguiste a bien perçu l'intérêt qu'il y a à mener le plus loin possible la description des phénomènes linguistiques de manière à rendre celle-ci opératoire pour la machine.

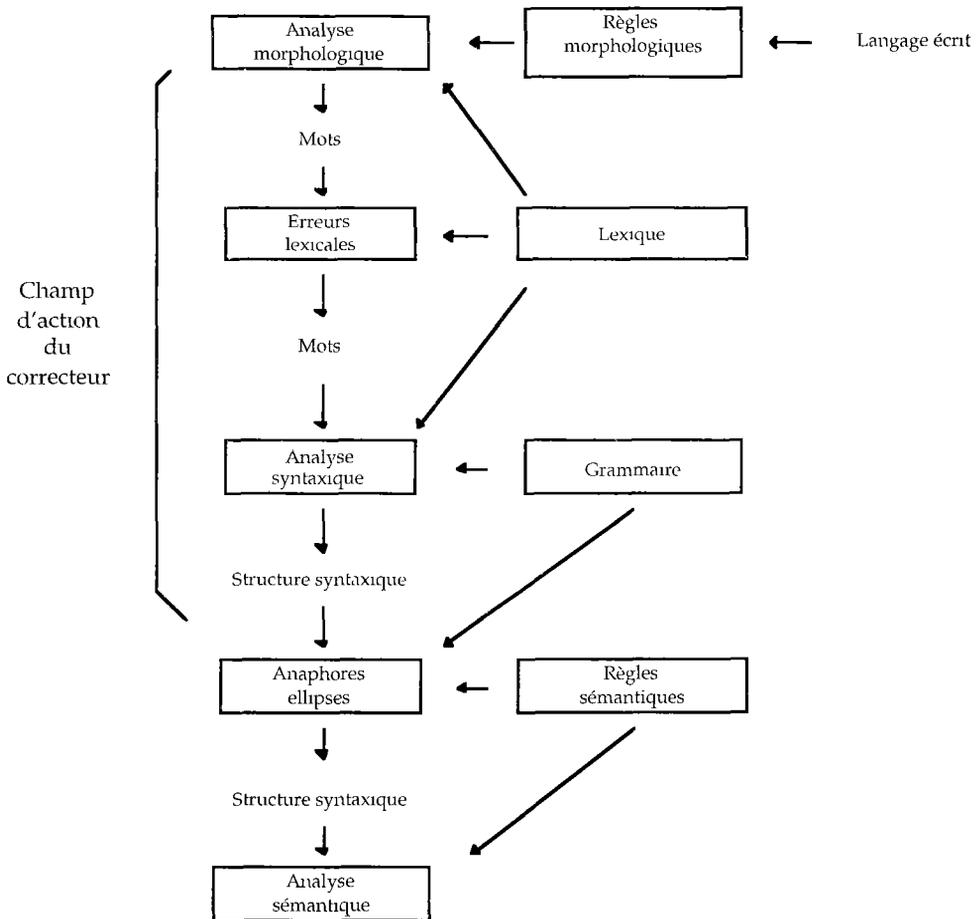
Si je me réfère au 101 (Nouvelle version 2.0), les caractéristiques suivantes sont à prendre en considération. Né de l'intention de développer un outil d'apprentissage de la grammaire par ordinateur, il se présente aujourd'hui comme un outil destiné aux « professionnels de la langue » disposant d'un ordinateur compatible IBM de type 386 ou 486. Indépendant des traitements de texte, il fonctionne sous Windows et reconnaît le format d'un nombre important de documents issus des traitements de texte les plus connus. Il occupe 2 Mo sur disque rigide. Le dictionnaire issu du *Petit Robert*, du *Petit Larousse* et du *Multidictionnaire* compte 66 000 entrées ou lemmes, soit plus de 500 000 formes déclinées. La grammaire repose sur l'actualisation de 3 500 « règles ». Plus de 1 000 règles ont été ajoutées dans la deuxième version. Deux modes de correction sont proposés : autonome et interactif, ce dernier permettant de mener l'analyse phrase par phrase. Par un lien hypertexte on accède à l'explication de l'erreur et de la correction suggérée.

Nous adopterons successivement le point de vue du linguiste et de l'utilisateur, conformément à l'attitude définie plus haut. Le linguiste s'intéressera au dictionnaire et à la syntaxe.

Le dictionnaire

La mission du correcteur, qui met en œuvre étiqueteur et lemmatiseur, demeure avant

tout orthographique. À ce titre, il occulte le joyeux flottement de l'orthographe française, qui du reste n'apparaît que lorsqu'on ouvre côte à côte plusieurs dictionnaires. Le problème de l'« orthographe d'usage », c'est-à-dire qui concerne les mots en eux-mêmes se trouve d'autant mieux résolu qu'il est permis d'ajouter des mots, de créer ses propres dictionnaires, hiérarchisés s'il s'agit d'une mise en réseau. L'autre orthographe, appelée traditionnellement « orthographe de règle », a été l'objet de toutes les préoccupations au cours de ces dernières années. Il s'agit de maîtriser le problème du transfert des catégories morphologiques (genre, nombre, personne) associées à des classes grammaticales (nom, pronom) sur d'autres classes grammaticales telles que le déterminant, l'adjectif, le verbe. Les mécanismes de l'accord se répartissent en trois types : à l'intérieur du syntagme nominal, dans le cadre de la phrase et au-delà des limites de la phrase. Nous sommes aujourd'hui au niveau 2.



Pour en arriver là, il a fallu bien sûr répondre à l'impératif d'occuper un minimum d'espace sur le disque tout en visant l'exhaustivité du dictionnaire. On subdivise donc les dictionnaires, on utilise des codes indiquant classes et catégories grammaticales pour éviter de placer toutes les formes dérivées dans le dictionnaire lui-

même, le programme de correction se chargeant de les retrouver. Si l'homonymie se trouve aujourd'hui signalée dans un grand nombre de cas, la phrase « *Un chèque sans provision, c'est rare* » est déclarée tout aussi correcte qu'« *Un cheik sans provision, c'est rare* ». Le syntagme « *les sports divers* » est tout aussi acceptable que « *les sports d'hiver* ». Dès que le sens régit l'orthographe de l'unité syntagmatique, le travail du correcteur s'arrête. « *Une chène de TV* » sera d'abord rectifié par rapport au genre de « chène », masculin. Mais faut-il prendre en considération le fait lui-même ou sa relativité par rapport à la fréquence d'une telle proposition ?

On l'a compris, le *dictionnaire* reste la base de l'analyseur. La procédure d'étiquetage des mots dans la phrase correspond d'abord à la consultation du dictionnaire qui permet d'associer aux mots de la phrase analysée des informations contenues dans le lexique. Si un mot est inconnu, l'analyse se bloque. La technique de l'intégration dans le dictionnaire qui est à mettre en œuvre répondra aux demandes les plus fondamentales de l'analyseur. On précisera donc si le mot ajouté est un nom, un verbe, un adjectif en indiquant le genre et le pluriel du nom, la conjugaison du verbe ou le féminin et le pluriel de l'adjectif, c'est-à-dire que les critères morphologiques sont exclusivement d'ordre flexionnel. Cette fenêtre permet par un système de notes d'indiquer une définition, une variante orthographique, un paronyme, un homonyme, une abréviation...

Trois remarques sont à faire. La classe des affixes dérivationnels, c'est-à-dire la morphologie lexicale ou « syntaxe interne » des mots, n'est pas prise en considération, alors que l'on pourrait tirer beaucoup de l'étude de la suffixation par rapport au genre ou aux changements de classe grammaticale pour l'analyse de la trame formelle et plus largement pour explorer les voies de la néologisation. On observera en outre que les listes de mots de liaison sont considérées comme closes. Enfin, la constitution de lexiques du français régional ou d'anglicismes est aisée à condition de se limiter à l'unité lexicale simple, les syntagmes n'ayant pas encore le droit d'entrée. Ceci constitue un problème majeur par rapport aux langues de spécialité.

La précision de l'identification dans la phrase est ainsi liée à la précision du dictionnaire. On doit se prononcer sur le fonctionnement syntaxique du mot intégré, sans passer par la sémantique. De là des options du type suivant pour le nom : « *Aquaculturiste* » - *Cette personne est — ; Cette personne est un — ; Je veux de l'—*. Les différents types de transitivité d'un verbe à intégrer sont posés selon les schémas de fonctionnement : *Elle — bien ; Elle — quelqu'un ; Elle — à quelqu'un ; Elle — de quelqu'un ; Elle — heureuse (verbe attributif)*.

On est évidemment loin des descriptions des relations lexique-syntaxe établies par M. Gross ou I. Mel'čuk. Non seulement la sémantique n'est pas convoquée ici, mais la complexité des constructions verbales transitives indirectes n'est pas encore prise en considération. « *Les enfants jouent tennis* » est admis comme « *Les enfants jouent leur avenir* » ; « *Il prend soin à lui* » est admis comme « *Il prend tout à son frère* ». L'ensemble des restrictions combinatoires est loin d'être en jeu. Le grammaticalement correct correspond au grammaticalement conforme aux règles choisies, qui ne constituent pas à elles seules l'*acceptable* tel que l'entend le locuteur dont le français est la langue maternelle.

Du dictionnaire à la syntaxe

Le découpage syntaxique se fait donc sur la base de l'unité mot, mot orthographique identifié par rapport à sa classe grammaticale, à ses catégories et à sa fonction, c'est-à-dire à sa définition en termes relationnels selon des désignations conformes au *Bon Usage* de Grévisse-Goosse.

L'unité mot apparaît donc comme l'unité formellement saisissable et analysable par rapport à l'unité phrase. C'est dire que la procédure distributionnelle qui consiste à déterminer la classe grammaticale par le type de fonction remplie n'est pas d'application ici. C'est dire aussi que deux voies sont encore à explorer : celle de l'organisation de l'unité morphématique, qui mettrait en évidence le pouvoir de néologisation de la langue, et celle de l'unité syntagmatique qui, à mon avis, permettrait d'aller plus loin dans la maîtrise des phrases complexes, tant il est vrai que le modèle syntaxique ne génère pas des phrases mais des suites de syntagmes.

Puisqu'aucune homologie directe n'existe entre le niveau morpho-syntaxique et le niveau « sémantico-référentiel », pour reprendre les termes de C. Hagège, on exploite ici fondamentalement la solidarité morphologique : un verbe conjugué ne peut se concevoir sans une marque de temps et de personne ; tout déterminant du nom sera au masculin ou au féminin, au singulier ou au pluriel ; l'ordre d'apparition des morphèmes obéira à une... fixité précise : *grand + e* et *non *e + grand*, *mang + era* et *non *era + mang*.

Mais le système grammatical actualisé par l'analyseur n'est pas sans poser quelques problèmes. Dans le cas du 101, il est précisé que le logiciel est en constante amélioration et que certaines constructions n'ont pas encore été analysées.

1. On sait que la fonction d'un élément n'est pas directement déterminée par sa classe grammaticale. Là se trouve toute la différence entre la physiologie et l'anatomie, disait Todorov. De là, toutes les anomalies entre le dictionnaire et l'analyse syntaxique.

Mais on sait aussi combien il est difficile d'articuler les facteurs qui président à la fonction : le critère positionnel (p. ex. l'adj. épithète), le critère morphologique (accord avec l'élément régisseur ou la forme du mot : les pronoms *je*, *me*, *moi* indiquant eux-mêmes la fonction) ou encore les critères de classe grammaticale.

2. Une ponctuation correcte et courante est nécessaire au bon fonctionnement de l'analyseur. Le point détermine chaque étape de l'analyse et limite celle-ci (unité phrase). Le problème de l'anaphore (ou référence à un élément présenté antérieurement) n'est pas résolu, à l'intérieur de la phrase et *a fortiori* dans la relation transphrastique. « *Pierre connaît ma maison, mais pas le tien* » sera accepté. Particulièrement fréquente dans le cas des pronoms, elle ne pose plus de problèmes que dans le cas du pronom relatif. L'explication se trouve dans la proximité immédiate de l'antécédent. Toutefois, l'antécédent complexe remet le sémantique à l'avant-plan et nécessite le recours à des options.

Il est vrai que l'on retrouve ici la vieille question : l'anaphore, qui joue un rôle important dans les accords fait-elle partie des phénomènes syntaxiques ou sémantiques ?

tiques ? Dans le même sens, d'autres analyses... ne se font pas. Le problème de la coréférentialité en est un exemple. « *Je veux que je fasse ceci* » est-elle une phrase grammaticale ? Aucune erreur ne sera détectée par l'analyseur. Il s'agit pourtant là de progrès réalisables. Certes rares seront les locuteurs qui avanceront ce type de phrase. Le cas de l'ordre des pronoms personnels compléments représente un autre exemple d'amélioration souhaitable et faisable. « **On lui la donne* » est déclaré de structure inconnue, mais *lui* est analysé comme pronom sujet. Or *on* est toujours sujet. Les problèmes d'analyse ne sont pas dans ce cas liés à l'orthographe. L'articulation du grammaticalement correct et de l'orthographiquement correct doit encore être approfondie : « *Lui être intelligent beaucoup* », aucune erreur détectée.

3. La phrase complexe.

Le degré de complexité de la phrase, lié à la subordination et l'insertion, a toujours limité les performances des analyseurs. De nets progrès ont été réalisés. Une phrase extraite du programme de ce colloque (« *On s'interrogera enfin sur l'avenir du dictionnaire, sous quelque forme que ce soit, en s'intéressant à l'apparition de nouveaux outils d'aide à la compréhension (et non à la traduction) avec recours à des interfaces dans la langue maternelle de l'utilisateur* ») peut amener le message pudique suivant : « Je manque de temps ». Rien ne change, si l'on attend. L'analyse fonctionne si l'on supprime le complément entre virgules « *sous quelque forme que ce soit* » et le syntagme entre parenthèses « *et non à la traduction* ». Toutefois dans les cas complexes, des analyses partielles et des options peuvent être proposées. Il apparaît d'ailleurs que les langues de spécialité, où les possibilités combinatoires sont plus restreintes dans les syntagmes et où l'acte illocutoire détermine les structures phrastiques, offrent un terrain privilégié à l'expérimentation d'analyseurs. La condition d'amélioration initiale est de pouvoir intégrer les syntagmes générés par les domaines.

L'utilisateur

S'il est vrai que dans ce genre de logiciel, l'interface est très importante, il faut indépendamment de cette convivialité prendre en considération les caractéristiques de l'habitus (au sens de Bourdieu) que le développement d'un correcteur grammatical peut installer chez l'utilisateur.

1. Le logiciel n'intervient pas dans l'opération d'encodage, à la différence du dictionnaire classique. Son utilisation suppose la systématisation de la production du texte brut, tel qu'on peut le voir circuler sur Internet aujourd'hui.

2. Le logiciel ne corrige pas tout et n'analyse pas tout. Ses limites sont parfois avouées, parfois dissimulées sous la convivialité de l'interface (*Je manque de temps*). Les indices de lisibilité, qui accompagnent certains correcteurs, font partie de ces écrans qui mettent en avant une pseudo-pragmatique occultant l'analyse réelle. Négligeant la complexité grammaticale et sémantique, ils se réfèrent à la longueur moyenne des phrases et du texte et proposent des résultats aberrants sur la base de l'indice de lisibilité de Flesch, sévèrement critiqué il y a déjà 25 ans par G. De Landheere de l'Université de Liège.

En réalité, le maintien de 4 ou 5 % d'erreurs doit-il être analysé dans l'absolu du

linguiste ou selon le type de document, voire selon le type de destinataire ? Plus de 50 % des erreurs corrigées ne sont jamais faites par des étudiants universitaires francophones, mais 68 % d'entre eux ne corrigent pas **tout ceux*, **bien que je cours* (40 %), **les corrections qu'effectuent ce logiciel* (40 %), erreurs que l'ordinateur corrige. Il est peut-être moins important de soumettre les phrases compliquées que de procéder au préalable à un étalonnage des erreurs dans la pratique sociale ! Et de ce point de vue, on ne peut ignorer la chance que peut représenter un correcteur grammatical pour défendre le français à l'étranger, en réduisant sa difficulté. Sur 804 mots d'un texte composé par des étudiants allophones universitaires, 78 types de fautes ont été recensés (soit près de 10 % des mots concernés). Un correcteur grammatical en corrige 40, présente 15 options, ignore 13 erreurs et se bloque dans 10 cas.

3. Ce type de logiciel fait appel à une compétence plus que minimale de l'utilisateur, qui doit parfois choisir parmi deux solutions et qui, surtout, doit faire face – chose nouvelle – à l'invention de la faute, appelée aussi « fausse détection », parfois supérieure à la vraie détection.

4. L'écriture se trouve ramenée à un outil technique et efficace. La règle se trouve rappelée dans sa simplicité scolaire, dans les limites imposées par l'écran, c'est-à-dire bien sûr dans son arbitraire. Est-ce une étape dans la réduction de la sacralisation de l'orthographe ?

Conclusion

L'évolution des études linguistiques s'est toujours faite par séparation des champs de recherche : synchronie-diachronie avec Saussure, forme-sens avec Sapir, Bloomfield, Harris, Chomsky. La formalisation d'une grammaire orthographique de la langue représente-t-elle une étape isolable dans le traitement de la langue par la machine ? L'articulation lexicale-syntaxe ainsi réétudiée module-t-elle une nouvelle approche dictionnaire ?

Un nouveau type d'écriture est-il à envisager ? Il serait conforme à une langue grammaticalement analysable par la machine. La diversité de la langue serait d'autant plus réduite que le correcteur par ses dictionnaires spécifiques signalerait les anglicismes, québécoisismes, belgicismes...

Le dépassement de la vérification locale pour embrasser la phrase représente un grand progrès et une étape importante dans l'établissement d'une grammaire cohérente qui rend compte de l'aspect mécanique du comportement syntaxique. Mais ni l'addition de règles, ni l'ajout de mégas ne suffisent. Plus les descriptions seront exhaustives, plus elles seront opératoires, plus elles permettront de faire coïncider le grammaticalement possible avec le grammaticalement acceptable. La frontière sémantique sera probablement redéfinie grâce aux recherches importantes dans la gestion du documentaire. Si l'orthographe pouvait ainsi se retrouver complètement désacralisée, le champ serait libre pour se jouer des difficultés autres qu'orthographiques et autrement plus passionnantes, pour tutoyer sans réserves le dictionnaire des difficultés.

Choix de grammaire et organisation du lexique*

Philippe BARBAUD

Université du Québec à Montréal, Canada

1. Les faits

Trois séries de faits au moins permettent de soutenir, sur le plan empirique, que les entrées lexicales dénominales ne sont pas toutes des atomes syntaxiques qui « projettent » leur catégorie au niveau maximal de leur structure syntagmatique symbolisée par SN ou N" ou encore X^{max}. Ce phénomène caractérise un phénomène d'« EXOCENTRICITÉ CATÉGORIELLE » et il s'avère particulièrement productif en composition lexicale grossièrement définie comme un processus de formation des mots composés. Nous montrerons dans ce travail qu'une approche syntaxique de ce phénomène sur le plan théorique fournit une base d'implantation motivée dans certaines grammaires d'unification.

1.1. Faux-endocentriques et composés déverbaux

On mentionnera d'abord le cas de certains composés nominaux que l'on qualifiera de « faux-endocentriques », c'est-à-dire des SN dont le genre n'est pas déductible de leur pivot nominal :

- (1) a. *UN (mille-pattes, mille-feuilles, rouge-gorge, deux-temps, trois-pièces, quatre roues motrices, tête-à-queue, du porte-à-porte, etc.)*
b. *UNE (deux-chevaux, six-cylindres, etc.)*

Dans cet échantillon de faux-endocentriques, le SN est d'un genre différent du ou des lexèmes nominaux qu'il contient.

* Ce travail a bénéficié du concours financier de l'AGENCE FRANCOPHONE POUR L'ENSEIGNEMENT SUPÉRIEUR ET LA RECHERCHE (AUPELF-UREF) et du CONSEIL DE RECHERCHE EN SCIENCES HUMAINES (CRSH) du Canada. Nous adressons nos remerciements à Fernande Dupuis, Benoît Habert et Christian Jacquemin pour leurs commentaires relatifs à certaines sections figurant dans une version préliminaire de cet article.

On mentionnera ensuite le cas des Noms Composés à Pivot Verbal (NCPV) du type *tire-bouchon* qui sont toujours de genre masculin (Surrige, 1985) :

(2) *UN (porte-parole, baise-main, brise-glace, casse-tête, gagne-pain, etc.)*

Dans cet échantillon, le genre masculin du SN ne provient ni du verbe ni du lexème nominal qui tient lieu de « COD pro-forma » du composé déverbal (Barbaud, 1994 : 9).

On mentionnera enfin le cas des dérivés nominaux du type [*mise* Prep N] dont le pivot est un participe passé et non pas un nom commun dérivé par conversion, comme nous le soutenons dans Barbaud (à paraître) :

(3) *UNE (mise en cause, à feu, en quarantaine, aux fers, sous vide, etc.)*

Dans ce paradigme particulier, il n'est pas évident que le nœud SN soit une extension catégorielle du V_{pp} *mise*.

À cette dernière observation il est loisible d'ajouter celle qui concerne les SN composés dérivés de verbes à l'infinitif et exceptionnellement à temps fini :

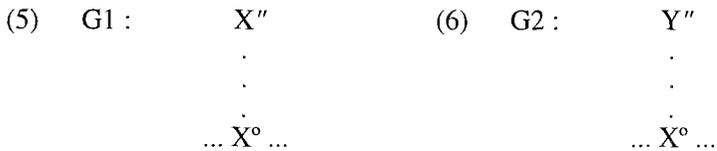
(4) *LE (laissez-passer, rendez-vous, faitout, laisser-faire, laisser-aller, etc.)*

Voilà un état de fait très succinctement décrit à partir duquel, néanmoins, il est loisible d'évaluer l'adéquation de certains modèles de grammaire ou de certains programmes de Traitement Automatique des Langues Naturelles (TALN).

Précisons toutefois que, bien que nous ne disposions d'aucunes données quantifiées pour appuyer notre position, nous croyons que le phénomène de l'exocentricité catégorielle est loin d'être marginal en français et dans l'ensemble des langues romanes en raison de leur morphologie particulière. Si le phénomène de l'exocentricité catégorielle semble effectivement absent des corpus de « langues spécialisées », au sens de Lerat (1995), il est en revanche dûment consigné dans les dictionnaires d'usage élaborés à partir de la langue courante. De ce fait, il est presque toujours associé, comme on peut s'en rendre compte, à un phénomène d'exocentricité sémantique ou d'interprétation « idiomatique ». Autrement dit, on note une très forte corrélation entre l'exocentricité formelle et la création d'entrées lexicales par composition, ce qui permet d'envisager le problème sous l'angle de la formation de mots et, par voie de conséquence, de la terminologie.

1.2. Deux modèles de grammaire en compétition

De manière générale, on peut définir l'exocentricité catégorielle comme une rupture ou une disjonction entre les propriétés formelles du nœud syntagmatique dominant et celles du terme lexical dominé qui lui sert de pivot (« tête » = "head"), peu importe la nature et le nombre des nœuds intermédiaires. Si l'on se réfère aux représentations arborescentes des grammaires génératives, l'exocentricité s'illustre grossièrement au moyen du graphe donné en (6), lequel s'oppose radicalement au graphe endocentrique donné en (5) :



Une grammaire G2 qui est compatible avec le graphe donné en (6) est soutenue par des propositions théoriques et descriptives aussi substantielles que divergentes par rapport à une grammaire G1 qui n'admet que des descriptions conformes au graphe donné en (5). Mais une telle perspective déborde évidemment du cadre limité de cet article.

Revenons à nos graphes pour y apporter une précision importante. Bien que le graphe exocentrique donné en (6) fournisse une représentation adéquate des cas patents de rupture catégorielle illustrés en (2-4), il s'avère incapable de rendre compte du cas des faux-exocentriques illustrés en (1) puisqu'à proprement parler, il n'existe aucune différence catégorielle entre SN = X'' et son pivot X° = N. La rupture catégorielle n'est pas le fait d'une disjonction entre X et Y'' mais plutôt le fait d'une disjonction entre certains traits de X par rapport à ceux de X''. Ce type d'exocentricité se laisse donc mieux caractériser par le graphe suivant, lequel fait état d'une rupture entre les traits α de X° et les traits γ de X''. Aussi le graphe (7) n'est-il possible, malgré les apparences, que dans une grammaire de type G2.



Par conséquent, on constate que la manipulation de traits acquiert autant d'importance vis-à-vis de la structure que la manipulation transformationnelle de nœuds atomiques ou syntagmatiques, ce qui favorise au départ un choix en faveur des grammaires dites « d'unification » par opposition aux grammaires de projection.

Quel que soit le moyen choisi pour assurer l'unification des deux niveaux de la représentation catégorielle d'un syntagme nominal (règles de projection, règles d'unification, conventions de percolation, mécanisme d'héritage ou de transfert de traits, etc.), on saisit d'emblée la difficulté majeure que doit affronter n'importe quel programme sensible aux traits morphologiques ou aux catégories de niveau atomique symbolisées par X°.

1.3. Problèmes de passage

Un parseur ou un analyseur dont l'ingénierie de base consiste à catégoriser des suites de lexèmes pour les regrouper en syntagmes cohérents de catégorie SN se heurte inévitablement à un problème majeur d'analyse interne puisque non seulement le genre du DET peut ne pas correspondre au genre du N mais aussi la catégorie du SN peut ne pas correspondre à la catégorie du lexème-pivot, ce que nous illustrons par les dif-

férentes analyses données en (8) :

- (8) a. [_{SN} DET_{<+masc>} [... N_{<+fém>}...]]
 b. [_{SN} DET_{<+fém>} [... N_{<+masc>}...]]
 c. [_{SN} DET [_{SV}... V ...]]

Tout algorithme strictement conforme à une grammaire de type G1 non seulement exclut radicalement l'exocentricité sur le plan théorique mais échouera aussi sur l'unification de ces trois descriptions qui seront déclarées mal formées bien que le même programme, peut-on présumer, fournisse un parsage et une catégorisation rigoureusement exacts des données. Par contre, dans une grammaire de type G2 qui n'exclut pas l'exocentricité, il est loisible en théorie de faire correspondre les graphes (6) et (7) aux trois descriptions données en (8) à la condition de modifier substantiellement l'ingénierie de base des modèles inspirés de G1 en ce qui concerne l'interface entre le lexique et la syntaxe. Il s'ensuit, du point de vue applicatif, que les cas d'exocentricité catégorielle n'ont rien d'insignifiant en TALN, en acquisition et en mise à jour terminologique parce qu'ils sont inévitablement source de bruit. Quant à la théorie linguistique, elle ignore purement et simplement le problème. Il n'en demeure pas moins que ni la catégorie N de ces composés ni leur sens dénominatif ne sont directement prédictibles sur la base des lexèmes qui les composent mais qui sont pourtant reconnus individuellement par n'importe quel parseur le moins performant. Non seulement le programme doit-il alors neutraliser la reconnaissance des lexèmes constitutifs pour éviter tout blocage d'une lecture contradictoire de leurs traits mais encore force-t-il à recourir à la lecture experte à des fins de repérage et d'encodage dans le lexique.

1.4. La solution de l'atomicité

La lecture experte est évidemment coûteuse mais c'est le moyen le plus efficace pour contourner ce problème d'application sans le résoudre au plan théorique pour autant. Le but de la lecture experte est donc d'encoder ce type d'expressions en tant qu'entrées lexicales particulières, ce qui est correct quant au résultat mais implique le recours à un dispositif de CONVERSION passablement arbitraire et aux incidences théoriques peu désirables. En effet, toute opération de conversion implique que la constituance syntagmatique verbale des NCPV par exemple et des expressions apparentées se voit réduite à une constituance atomique nominale afin de pouvoir assurer l'insertion de l'entrée lexicale dans le domaine du SN et non pas dans un domaine différent catégoriellement. Bref, l'ingénierie de conversion d'une grammaire de type G1 doit permettre d'assigner la catégorie N à des formes préalablement reconnues comme étant de catégorie non-N selon un schéma de règles donné en (9) et ce, pour finalement contourner les analyses données en (8) et plus particulièrement (8c)¹ :

¹ Voir J. Thiele (1987) pour une illustration typique de cette pratique. La nominalisation par conversion se distingue de la nominalisation par affixation du type illustré en (1), cette dernière ayant fait l'objet d'un grand nombre d'études, comme on sait.

(1) a. *La destruction (de Rome par Alaric)*

b. *L'invasion (de la Gaule par Attila)*

Dans ce type de nominalisation, il est loisible d'assimiler le suffixe à un prédicat lexical auquel on assigne une variable argumentale – le morphème radical – en position de spécifieur, *i.e.* celle qui précède le prédicat, ce qui permet d'établir un rapport formel avec la structure syntaxique. Pour une discussion plus approfondie de la conversion, voir Barbaud (1994).

- (9) a. $A \rightarrow N$ (*blanc, possible, important, etc.*)
 b. $V_{inf} \rightarrow N$ (*sourire, déjeuner, lancer, être, avoirs, etc.*)
 c. $V_{ppé} \rightarrow N$ (*attaché, émigré, blessé, préposé, etc.*)
 d. $V_{ppt} \rightarrow N$ (*dirigeant, protestant, militant, dégivrant, etc.*)

Convertir en atome nominal toute expression reconnue par un parseur robuste comme un constituant non-N est une opération catégorielle qu'il faut quand même accomplir manuellement afin de respecter l'architecture de la grammaire G1, architecture que l'on peut représenter au moyen du schéma suivant :

(10)

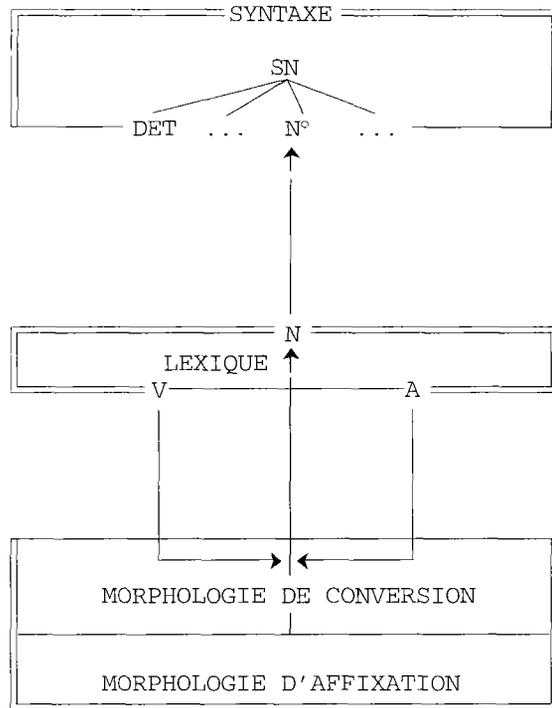
Fonctionnement d'une grammaire de type G1

Algorithme de
constituance
non atomique

Algorithme
d'insertion

Algorithme de
transfert

Algorithme de
constituance
atomique



Dans une telle grammaire, l'endocentricité obligatoire de la constituance non atomique en syntaxe oblige à respecter la condition d'atomicité des entrées lexicales afin de permettre au mécanisme de l'insertion lexicale de s'appliquer. C'est donc à la morphologie que revient la tâche de réduire les composés nominaux déverbaux sous forme d'atomes N au moyen de règles quelconques de conversion. On obtient alors un modèle essentiellement « projectif », c'est-à-dire orienté de la morphologie vers le lexique et du lexique vers la syntaxe. Il fonctionne nécessairement « du bas vers le haut » (*bottom to top*). Dans ce type de grammaire, le lexique se conçoit entre autres comme une base de données MONOSTRATIFIÉE dont chaque élément discret est uniformément encodé sous une seule forme disponible, la forme atomique des X^0 de la grammaire.

1.5. L'opération d'insertion lexicale

On peut ici préciser que c'est l'opération d'INSERTION lexicale telle que Chomsky l'a conçue dans les années 60 qui force à adopter la contrainte de l'atomicité des entrées lexicales d'où découle la conception actuellement dominante de l'endocentricité projective des structures syntagmatiques. Mais l'insertion lexicale est une opération pseudo-grammaticale qui n'a aucun fondement cognitif. Le cerveau ne fonctionne pas à sens unique, comme Changeux (1983) le dit explicitement. L'opération d'insertion lexicale si familière aux générativistes est en réalité une servitude des grammaires transformationnelles primitives des années 50 conçues en fonction des limites inhérentes aux machines de Turing. Cette conception est toujours prédominante. Il n'est pas dit que cet héritage ne doive pas être remis en cause de nos jours et ce, pour trois raisons.

La première est que personne à ma connaissance n'a fourni la moindre preuve démontrant que l'exocentricité des syntagmes est une propriété formelle hors de portée d'un cerveau humain et qu'elle doit être absente des modèles de représentation des propriétés formelles des langues naturelles.

La seconde, plus technique, est que l'insertion lexicale ayant été conçue à l'origine comme une opération de SUBSTITUTION de chaîne phonétique (Chomsky, 1965), rien n'empêche d'envisager que d'autres types d'opérations formelles puissent assurer le couplage du lexique et de la syntaxe de manière plus efficace et plus « explicative ». Il devient évident qu'une grammaire conçue en termes d'insertion d'entrées lexicales sous chaque nœud atomique d'une structure syntaxique exclut radicalement l'insertion de matériel sous les autres nœuds, que ceux-ci soient de niveau intermédiaire ou de niveau syntagmatique (maximal).

La troisième raison est que la contrainte de l'atomicité des entrées lexicales est incompatible avec le phénomène de la variation qui affecte les entrées lexicales polylexémiques et ce, malgré l'incidence du phénomène de figement qui les caractérise significativement². Par exemple, si on déclare que l'expression *basse pression* est une entrée lexicale d'un corpus médical, comment la traiter dans l'expression *une alternance de basses et hautes pressions* ?

En définitive, l'opération d'insertion restreint considérablement l'accès direct au lexique de l'information syntaxique.

2. Le concept de la « syntaxe dérivationnelle »

Le concept de « syntaxe dérivationnelle » que nous allons maintenant esquisser justifie notre choix en faveur d'une grammaire de type G2 que nous estimons plus appropriée au traitement des Entrées Lexicales Polylexémiques du français (ELP). Comme l'indique cette étiquette, la syntaxe est une composante pourvoyeuse de mots,

2. Pour un état de la question en français dans le cadre des grammaires applicatives, voir la **Présentation** de Habert & Jacquemin (1993) au numéro de *Traitement automatique des langues* portant sur le thème des **Traitements automatiques de la composition nominale**, vol. 34 (2).

à l'instar de la morphologie dérivationnelle. Les règles qui en composent l'algorithme engendrent les structures adaptées à la formation des locutions verbales et des noms composés, c'est-à-dire des structures susceptibles d'être interprétées lexicalement, d'où l'importance que nous accordons à la manipulation de traits sémantiques. Nous articulons ainsi, dans un cadre de grammaire moderne, la substantielle proposition de Benveniste (1974) relative à la « synapsie », reprise notamment par Guilbert (1975), mais sans la limiter au seul domaine du SN, ce qui n'a aucune raison d'être aux plans théorique et empirique. La formation d'ELP répond à un besoin de néologie qu'on pourrait qualifier de « transcategoriel », d'où l'enjeu d'une théorie X-barre correctement formulée.

Dans son architecture générale, le modèle que nous préconisons fait appel à quelques principes de base et à des opérations compatibles seulement avec un nombre restreint de modèles de grammaires d'unification. Commençons par les fondements.

2.1. Fondements théoriques

Nous posons les deux postulats suivants. Le premier se formule ainsi :

(11) POSTULAT DE L'EXOCENTRICITÉ NOMINALE

L'exocentricité du syntagme nominal est aussi une propriété des langues naturelles.

De ce premier principe, qui se veut une alternative à la contrainte de l'endocentricité des syntagmes des grammaires génératives, on tire la conclusion que la catégorie SN est aussi une catégorie primitive de la grammaire universelle, au même titre que la catégorie initiale Σ = phrase, puisqu'elle n'est pas toujours dérivée d'un pivot de catégorie N. Un SN n'est donc pas la nécessaire projection de N. Il s'agit là d'un débat dont l'issue ne saurait être validée que sur le plan empirique. À cet égard, les faits de composition lexicale dans les langues romanes soutiennent fortement notre position théorique. En anglais, ce sont les nominaux gérondifs qui supportent l'énoncé (11)³.

Le second postulat s'énonce comme suit :

(12) POSTULAT DE LA NEUTRALITÉ CATÉGORIELLE DES ENTRÉES LEXICALES

Toute catégorie syntaxique est susceptible de définir la forme d'une entrée lexicale.

Ce second principe est une alternative à la servitude de l'atomicité des entrées lexicales inhérente aux grammaires génératives de type chomskyen. Ce principe requiert l'élaboration d'un dispositif d'interface bidirectionnel entre le lexique et la syntaxe et il doit être apte à être implanté dans un modèle applicatif. Un tel dispositif tient dans la proposition suivante :

3. Comme Weibelhuth (1995 : 89 n. 25) l'a fait récemment remarqué dans son manuel, les nominaux gérondifs font encore problème aujourd'hui puisque leur structure viole la contrainte d'endocentricité de la théorie X-barre.

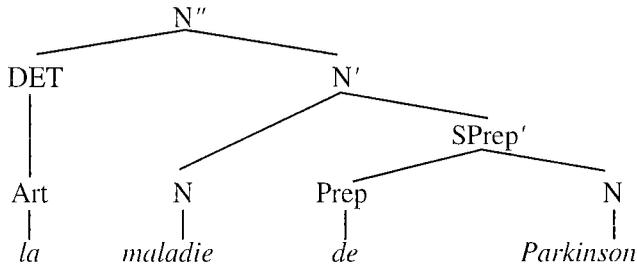
(13) **PRINCIPE DE RÉTROPROJECTION**

Dans un modèle de syntaxe dérivationnelle, chaque SITE d'une structure syntaxique est susceptible de CORRESPONDRE à une entrée lexicale.

2.2. Sites et positions

Précisons brièvement de quoi il retourne. D'abord nous nous inspirons ici d'une distinction féconde de Milner (1987) entre « sites » et « positions » syntaxiques. Dans la configuration (14), on dénombre seulement quatre positions (nœuds terminaux encadrés) mais huit sites. En vertu du principe de fonctionnement (13), ce n'est plus un potentiel de seulement quatre entrées lexicales qui concourt à l'interprétation de cette structure mais un potentiel de huit, ce qui inclut non seulement les catégories pré-terminales (ou atomiques) mais aussi les catégories de niveau nodal X' et X''. C'est certainement là un gain appréciable d'efficacité pour une même structure.

(14)



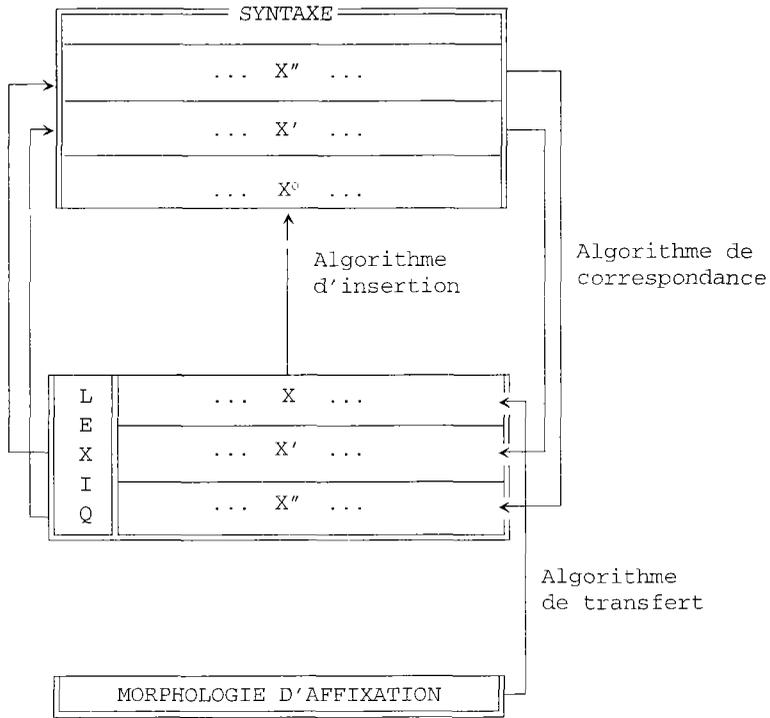
Ensuite, nous basons l'échange lexique \ syntaxe sur une opération de CORRESPONDANCE entre nœud syntaxique et signification lexicale et non de substitution entre atome et terme lexical. La distinction est cruciale parce que la correspondance n'implique pas l'insertion mécanique des entrées lexicales dans l'arbre mais elle ne l'exclut pas pour autant. La correspondance actualise alors la « rétroprojection » des catégories majeures dans le lexique qui se structure en conséquence. Par ailleurs, dans notre modèle comme dans les autres, toute position qui contient un arbre quelconque définit le domaine d'application des règles d'insertion. En outre, une position ne peut être occupée que par un atome parce que le seul objet qu'une règle d'insertion puisse mettre en jeu, c'est l'atome défini comme catégorie lexicale. En revanche, toute position est un nœud qui se définit aussi comme un site mais l'inverse n'est pas vrai. Le principe (13) opérera donc à chaque niveau nodal d'une configuration syntaxique quelconque, ce qui revient à dire que les nœuds supérieurs peuvent aussi tenir lieu de catégories lexicales grâce à une opération de correspondance mais non pas d'insertion.

La bidirectionnalité du modèle est donc assurée à la fois par une opération d'insertion au niveau nodal X° dans le sens lexique → syntaxe et par une opération de rétroprojection aux autres niveaux nodaux dans le sens syntaxe → lexique, ce qui peut être illustré par le schéma suivant :

(15)

Fonctionnement d'une grammaire de type G2

Algorithme de
constituance
non atomique



Algorithme de
constituance
atomique

L'organisation du lexique n'est plus la même : il est maintenant polystratifié en fonction de trois dictionnaires, celui des atomes, des « synapses » et des syntagmes, ce qui garantit un accès direct aux entrées lexicales complexes de forme synaptique, majoritairement composées de mots composés et de locutions, et de forme syntagmatique, moins nombreuses mais tout aussi idiosyncrasiques, telles les dictons, proverbes et sentences du type *La montagne accouche d'une souris*. Celles-ci n'ont plus à être récupérées par l'intermédiaire d'un de leurs éléments constitutifs, comme dans les dictionnaires d'usage. Tout site fait-il ainsi l'objet d'une procédure de vérification constante (*checking device*) dans le lexique. C'est ce que nous voulons dire par « correspondance » sites \ lexique.

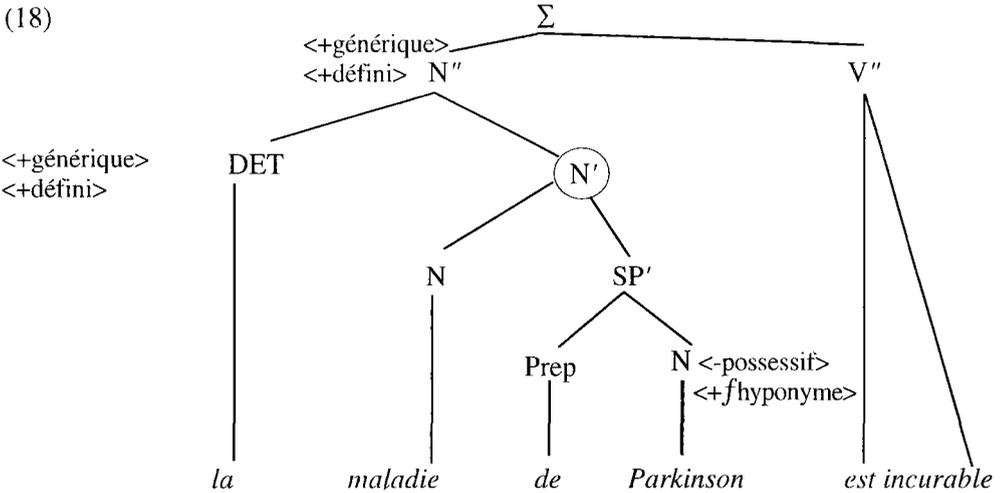
2.3. Traits lexicaux et isomorphie structurale

Notre approche est empiriquement justifiée par le phénomène de l'ambiguïté de nombreux mots composés selon qu'ils sont interprétés soit littéralement, c'est-à-dire par compositionnalité, ou soit non littéralement, c'est-à-dire par idiomaticité. Dans les deux cas, c'est la même structure qui tient lieu de domaine d'application des règles d'interprétation. C'est ce qu'illustre le contraste suivant :

(16) *La maladie de Parkinson est toujours incurable*

(17) *La maladie de Parkinson perturbe notre service des ventes*

Dans le premier exemple, la suite *maladie de Parkinson* s'interprète comme une entrée lexicale qui réfère à une maladie spécifique. De par sa position syntaxique dans la structure, le nom propre est un complément déterminatif qui s'interprète en tant que « foncteur hyponymique » du vocable générique *maladie* agissant comme pivot de la structure dans les langues romanes. Comme de fait, la *maladie de Parkinson* est une {SORTE DE} *MALADIE*, au sens de Levrat & Sabah (1990). Cette interprétation est propre à la sémantique lexicale et elle justifie le recours aux traits pertinents dans une description linguistique⁴. En outre, le mot *Parkinson* est dépourvu de son contenu référentiel en raison de l'absence significative de DET interne, suivant en cela les idées de Cadiot (1991). Cette propriété structurale se reflète directement, soutenons-nous, dans une version adéquate de la syntaxe X-barre des grammaires chomskyennes, (Barbaud, 1992 : 202 ; 1994 : 15), ce qui permet alors d'engendrer la structure grossièrement décrite en (18) :



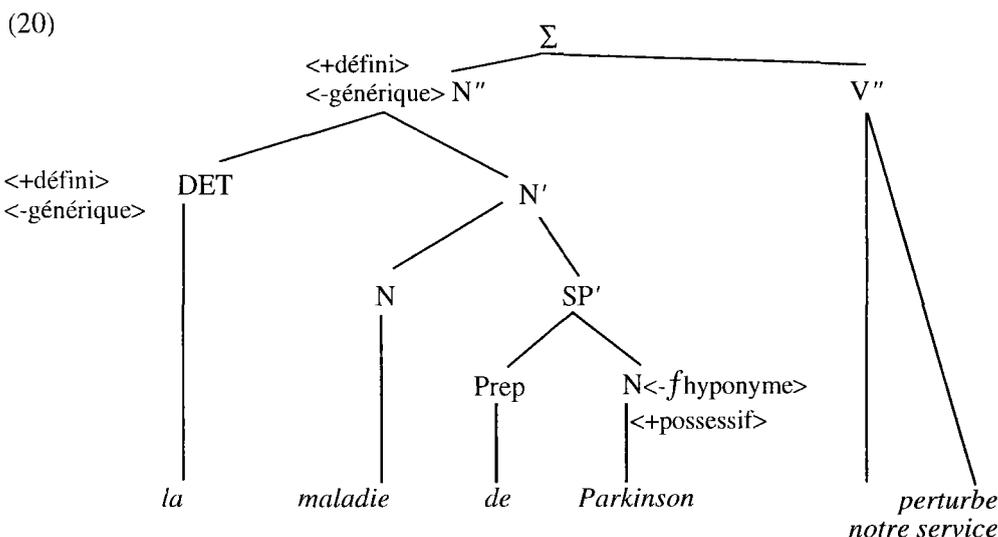
Grâce au mécanisme de rétroprojection constante, il s'établira une correspondance entre le nœud N' encerclé, qui est l'un des sites que nous qualifions de « synaptiques », et un objet du lexique préalablement encodé dans le lexique au moyen du trait <+N'> et dont la matrice contient une description structurale identique à celle du nœud concerné, ce qui inclut la suite des termes [*maladie de Parkinson*]. L'interprétation correcte est conférée par le trait <+hyperonyme> dont la valeur sémantique renvoie à un vocable approprié de la langue (lorsque disponible). Dans les composés endocentriques, il y a coïncidence par unification de traits entre ce vocable et le nom-pivot, ce qui rend compte de la variabilité en genre. À titre d'essai, le prototype d'une telle entrée est donné en (17)⁵ :

4 Pour être plus précis, nous mettons en parallèle les dénominations de catégorie SN et les propositions de catégorie Σ. Chaque domaine possède une interprétation fonctionnelle qui lui est particulière (voir le plaidoyer de Lerat (1990) pour une « hyperonymie fonctionnelle »). Alors que les phrases ont systématiquement recours au canevas des rôles (valeurs) thématiques pour être interprétées, les SN dénominatifs de leur côté ont systématiquement recours aux diverses valeurs d'hyper et d'hyponymie potentiellement associées aux mots. Le statut théorique de ces différentes valeurs interprétatives est le même et leur importance est déterminante dans l'architecture du modèle adopté.

5. Le traitement des locutions verbales tronquées du type *prendre parti, tirer parti, tenir tête*, etc., est similaire à la différence que ces entrées lexicales sont munies du trait <+V'>.

- (19) [maladie de Parkinson] <+N>
- | |
|--|
| N1 = /maladie/; N2 = /parkinson/ |
| N1 = <+hyperonyme> = {MALADIE} |
| N2 = <+fhyponyme> = {avec tremblements constants chez les personnes âgées} |

À défaut de la présence d'une entrée lexicale existante, la compositionnalité des nœuds poursuit son cours. Cependant, chaque fois qu'un nœud synaptique s'apparie à une entrée lexicale, c'est la lecture idiomatique qui prévaut sur la lecture compositionnelle, laquelle est autrement assignée par défaut nonobstant la réelle faisabilité de la chose comme Kayser et Lerat (1990 : 120 *sqq.*) en discutent. C'est ce que nous obtenons dans le cas de l'exemple (17) qui est entièrement compositionnel. La structure est la même qu'en (18) mais l'interprétation diffère à cause des différences importantes de traits :

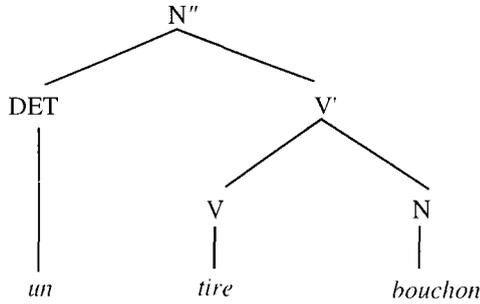


Ce traitement montre que l'interprétation dénomminative, toute lexicale qu'elle est, se laisse appréhender en termes de traits proprement lexicaux et d'isomorphie structurale plutôt qu'en termes de catégories lexicales et de conversion. Une telle approche de l'ambiguïté lexicale en composition favorise les grammaires d'unification.

2.4. Grammaires d'unification et exocentricité

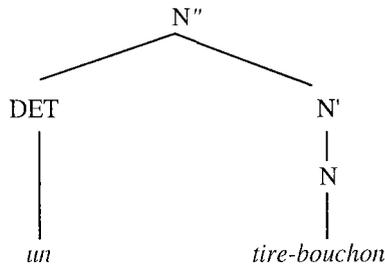
En prenant les NCPV en exemple, nous allons maintenant procéder à une analyse applicative de ces objets lexicaux. La rétroprojection et l'isomorphie structurale vont déterminer notre ligne de conduite. Linguistiquement parlant, nous disons que les NCPV sont des SV basiques nominalisés, en réalité des quasi-SV de catégorie V' (récursive) dominés par le nœud N'' en structure profonde. C'est donc la configuration typique, bien que fort incomplète, que nous donnons en (21), à mettre en parallèle avec la description (8c), qui est l'enjeu de notre analyse :

(21)



L'exocentricité de la configuration (21) contraste avec l'endocentricité de la configuration suivante, la seule actuellement permise en théorie syntaxique générative :

(22)



La validation de la structure (21) est assurée par le fait que la rétroprojection du V' dans le lexique se voit ratifiée par l'existence d'une entrée lexicale de catégorie <+V'> dont la matrice contient le sens {INSTRUMENT} associé à la suite polylexémique *tire bouchon*!. En cas d'échec de la rétroprojection, la structure est rejetée parce que son exocentricité n'est pas interprétable par compositionnalité.

En ce qui concerne l'application, il faut convenir que l'implantation de ces idées dans un modèle de linguistique computationnelle ne peut pas être accomplie par n'importe quelle grammaire d'unification. À vrai dire, il n'y a que les **Grammaires d'Arbres Adjoints** (TAG = *Tree Adjoined Grammar*) qui autorisent un tel transfert en raison du formalisme particulier qui est le leur. En effet, parmi les quatre principaux modèles de grammaire d'unification exposés en détail dans Abeillé (1993), seules les TAG disposent de l'appareillage requis pour véritablement rendre compte de l'exocentricité nominale⁶. Dans ce modèle, la notion de « tête » (pivot) y est uniquement lexicale tandis que dans les modèles concurrents comme la GPSG de Gazdar, Klein, Pullum & Sag (1985) ou comme dans la HPSG de Pollard & Sag (1994), la tête identifie la structure (Abeillé, 1993 : 214). Toujours selon cet auteur, dans le

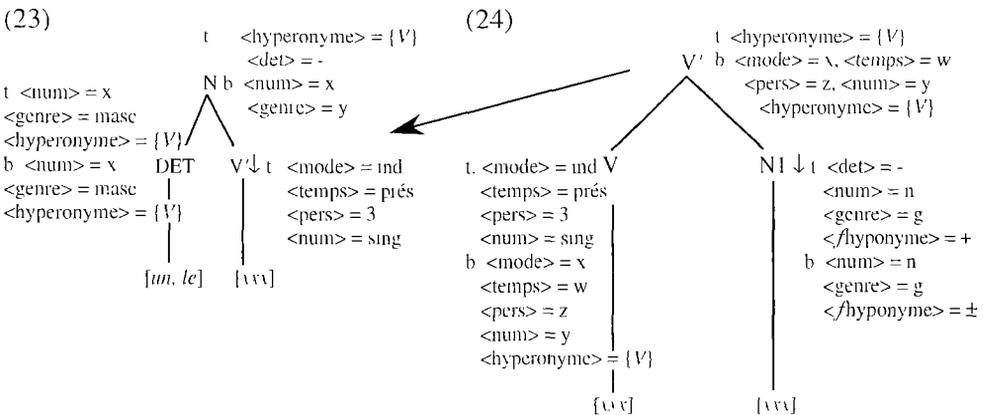
6 Ajoutons toutefois que de l'avis de Philippe Blache et Lorne Bouchard (communication personnelle), le modèle de la GPSG offre les ressources nécessaires pour traiter l'exocentricité formellement. Si c'est le cas, il convient de déterminer si la cohérence interne de la théorie linguistique qui le sous-tend peut s'accommoder du postulat énoncé en (11)

modèle basé sur la LFG de Bresnan (1982), les règles de réécriture hors contexte ne font jamais appel à l'exocentrique catégorielle des syntagmes sauf en ce qui concerne le symbole initial P (phrase). Enfin, précisons qu'aucune grammaire générative de type chomskien comme celle par exemple du *Gouvernement et du Liage* (Chomsky, 1981) n'est dotée du pouvoir de générer des structures syntaxiques exocentriques (cf. Webelhuth, 1995 : 33).

En tout état de cause, les TAG offrent deux dispositifs qui permettent d'implanter cette propriété formelle des langues naturelles. D'abord, la manipulation d'unités de base sous forme d'arbres élémentaires (initiaux ou auxiliaires) définissant le domaine d'application d'opérations d'adjonction ou de substitution. Cela permet d'adjoindre directement un nœud de racine V' à un nœud de racine N pour former un nom. Ensuite, les TAG permettent de dissocier les traits « traits aval » et « traits amont » lors du processus d'unification⁷. Grâce à cette dissociation, il devient possible d'exprimer un grand nombre de contraintes formelles sous forme de traits amont d'un nœud quelconque, ces derniers agissant notamment comme filtres vis-à-vis des lexèmes candidats.

2.5. Arbres initiaux des NCPV

Par conséquent, nous allons postuler deux arbres élémentaires initiaux, l'un de racine N donné en (23) et l'autre de racine V' donné en (24)⁸. C'est donc l'adjacence DETV' qui instancie l'exocentricité des NCPV en TAG :



7 Rappelons que les traits amont d'un nœud donné permettent de spécifier les diverses relations de ce nœud avec ceux qui le dominent tandis que les traits aval permettent de spécifier les relations qu'entretient ce nœud avec ceux qu'il domine, plus particulièrement ceux qui sont associés aux termes lexicaux.

8 Sur ce dernier point, nous divergeons d'opinion avec Abeillé (1993 : 207 n. 100). À l'instar de M. Gross, cet auteur écarte l'hypothèse que les arbres phrastiques élémentaires de catégorie P sont dépourvus de nœud SV intermédiaire pour le français. Nous invoquons au contraire la grande productivité des nœuds SN et SV « tronqués » en composition lexicale, c'est-à-dire dépourvus de position SPEC interne notamment dans quelques 200 locutions verbales et plusieurs milliers de noms composés, pour justifier le caractère récursif des nœuds intermédiaires V' et N', un argument qui interpelle aussi Corbin (1987).

Ce traitement appelle évidemment plusieurs commentaires.

1. L'arbre initial (24) de racine V' se **substitue** (\neq adjoint) au nœud V' de l'arbre initial (23) de racine N. Ce choix est justement dicté par la nécessité de rendre compte de l'exocentricité catégorielle dans le domaine de la racine nominale et non dans celui d'une autre racine. Grâce à la substitution (notée \downarrow) requise par le nœud V' de l'arbre **initial** (23), la configuration du DET devient réceptrice de l'arbre (24) selon des conditions très strictes. Cela n'aurait pas été le cas si on avait choisi un arbre **auxiliaire** étant donné que ce type d'arbre commande une opération d'adjonction qu'on aurait notée V*, ce qui aurait eu pour effet d'insérer l'arbre (23) dans un autre arbre de racine éventuellement différente, par exemple de racine P, d'où l'impossibilité de réinsérer le tout sous un nœud N ;
2. Le trait amont de **genre <+masc>** du nœud DET de l'arbre (23) prédit que les NCPV sont toujours masculins⁹ ;
3. L'invariabilité de la **flexion verbale** des CNPV est prédite par les traits amont du nœud V' de l'arbre (23) ;
4. Le **caractère tronqué** du COD pro forma des NCPV est prédit par la valeur négative du trait amont <det> de l'arbre (24) ; la même valeur en amont du nœud racine de l'arbre (23) évite que l'arbre dérivé puisse être adjoint à un arbre auxiliaire DET ;
5. L'interprétation du COD en tant que **foncteur hyponymique** est requise par la présence du trait amont </hyponyme> spécifié positivement dans l'arbre (24) ;
6. La **contrainte d'hyperonymie** nécessaire à l'interprétation des synapses nominaux exocentriques se traduit par la présence du trait <hyperonyme> en amont des nœuds DET et N de l'arbre (23). Il s'agit d'un trait d'accord destiné à s'unifier avec un trait identique dans le nœud V' adjacent ;
7. Si le nœud V' s'interprète effectivement en fonction d'un hyperonyme quelconque, sa variable notée {V} (pour « vocable hyperonymique ») sera remplacée par la valeur lexicale inscrite dans l'entrée lexicale synaptique, *i.e.* {INSTRUMENT} pour [tire-bouchon]. L'unification se fait par accord DETV'.

Par conséquent, le partage des traits qui résulte de la substitution de l'arbre (24) au nœud V' de l'arbre (23) donne en principe non seulement une structure bien formée dans ce type de grammaire mais aussi rend compte simultanément des propriétés prédictibles des NCPV tant sur le plan morphologique que syntaxique.

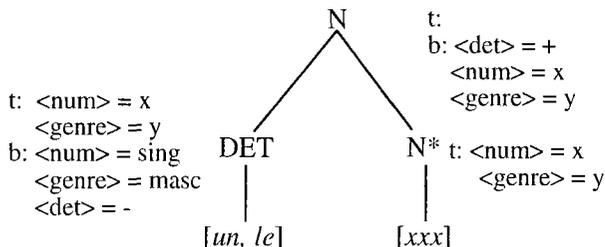
2.6. Canonicité

Toutefois, bien que ce traitement soit à première vue conforme aux quatre principes de bonne formation des arbres élémentaires en TAG, certaines difficultés doivent être résolues. Nous en décelons deux. La première a trait à l'arbre (23) qui n'est pas canonique parce que la substitution est censée être réservée aux noms et aux verbes à

9. Il est possible cependant de se débarrasser de ce trait dans la mesure où le trait <genre> est un « trait d'accord », cf. Abeillé (1993 : 219), dans le domaine du SN et que la grammaire comprend une règle d'accord entre catégories adjacentes selon le genre non marqué en français. Puisque le nœud substitué V' ne contient aucune mention de trait de genre, le trait <genre> du DET prendra automatiquement la valeur non marquée <+masc>. Un tel traitement n'est pas sans rappeler l'accord SPECtête (flexionnelle) ou encore l'accord sujet-verbe dans le domaine de Σ .

compléments nominaux tandis que l'adjonction l'est aux adjectifs, adverbes et autres lexèmes modifieurs, y compris les déterminants, cf. Abeillé (1993 : 206). Ici, l'arbre (23) est un arbre à substitution assigné à DET, ce qui n'est pas permis. La seconde est que si l'on a affaire à un arbre à substitution, ce dernier est censé afficher un nœud racine de même catégorie que le nœud feuille où a lieu la substitution. Ce n'est pas le cas de l'arbre (23) qui affiche trois nœuds différents par comparaison avec la représentation canonique des DET donnée en (25), cf. Abeillé (1993 : 226) :

(25)



La résolution de ces deux difficultés ne pose guère de problème majeur, nous semble-t-il. D'une part, la restriction relative à l'identité du nœud-feuille et du nœud-racine dans les arbres à substitution n'est pas absolue : le nœud P en TAG ne domine pas de nœud-feuille de catégorie P. D'autre part, un modèle de TAG doit pouvoir incorporer des structures à substitution sans égard à la nature des éléments impliqués. On ne voit pas pourquoi la substitution serait interdite aux déterminants. En outre, il n'y a aucune raison de principe à diviser le travail dérivationnel en attribuant l'exclusivité de l'une ou l'autre des deux opérations à un ensemble de catégories plutôt qu'à un autre. Ce n'est pas la catégorie d'un nœud qui commande une opération d'adjonction ou de substitution mais la dynamique de la récursivité dans une langue donnée. C'est ce qu'Abeillé (1993 : 209) a tenté d'illustrer en distinguant structurellement les composés du type *verre de vin* par opposition à ceux du type *verre à vin*. Toutefois, bien que nous souscrivions à son analyse, nous n'entérinons pas le traitement qu'elle propose et qui consiste en fin de compte à traiter l'hyponymie par le biais de différences structurales et dérivationnelles, c'est-à-dire par une structure à adjonction dans le premier cas (arbre auxiliaire) et à substitution dans le second (arbre initial). Il y a un danger réel à attribuer aux opérations qui manipulent des formes catégorielles certaines propriétés de nature interprétative.

La différence de distribution entre ces deux types de composés sur laquelle s'appuie Abeillé pour justifier une différence de traitement syntaxique est attribuable bien plus à une différence lexicale qu'à une différence dérivationnelle. En réalité, il y a une différence d'hyponymie entre les deux expressions nominales bien que leur structure soit la même. Celle des SP de forme [N Prép N] ou [N Prép DET N] du français exhibe toujours la même configuration dans le domaine du SN parce que la récursivité du syntagme prépositionnel ne varie pas en fonction des différences lexicales du pivot : un SP se branche toujours de la même manière par rapport à son pivot. Dans *verre à vin*, nous n'avons qu'un seul hyperonyme référentiel, à savoir une sorte de {VERRE}, autrement dit un contenant sans contenu référentiel, et cet hyperonyme coïncide lexicalement avec le pivot *verre*, comme c'est le cas dans tous les composés endocentriques. Comme de fait, on peut remplir son verre à vin d'eau Perrier par exemple. Dans *verre de vin* par contre, nous avons deux hyperonymes référentiels, le

{*VERRE*} et le {*VIN*}, c'est-à-dire le contenant et le contenu. Comme de fait, lorsqu'on ajoute de l'eau Perrier à son verre de vin, on obtient un contenu dilué, ce qui n'implique pas que l'on boive dans un verre à vin. La distinction relève non pas d'une différence de structure résultant d'une différence dérivationnelle mais bien d'une différence d'interprétation lexicale du SP. Incidemment, une telle différence de structure s'avère redondante par rapport à celle que marque déjà le jeu des deux prépositions. Dans l'expression *verre à vin*, le SP tronqué s'interprète en tant que **foncteur** hyponymique exerçant sa rection sur le pivot qui le précède tandis que dans l'expression *verre de vin*, le SP tronqué n'est pas un foncteur hyponymique car il ne contribue pas à faire de *vin* une {*SORTE DE*} *VERRE*. Cela se reflète dans l'alternance des prépositions : dans le domaine du SN, la préposition *à* est presque toujours un **marqueur** d'hyponymie mais pas la préposition *de*. Il est donc préférable de marquer cette différence interprétative au moyen de traits fonctionnels plutôt que de structures dérivées différemment en syntaxe, d'où le « commutateur » </hyponyme>+.

C'est aussi la raison pour laquelle nous avons opté pour la spécification d'un trait d'hyponymie muni d'une variable dans les arbres (23) et (24), trait qui sollicite l'unification des deux structures par accord. Encore une fois, il s'agit d'un trait de sémantique lexicale amplement justifié par les faits de composition¹⁰. Un tel trait fait partie de l'encodage des entrées lexicales synaptiques. Comme on a pu le remarquer avec la description (17), le vocable qui sert d'hyponyme est encodé dans les entrées lexicales synaptiques. L'encodage du NCPV *tire-bouchon*, par exemple, se présentera grossièrement comme suit :

- (26) [**tire-bouchon**] <+V'>
 {
 V = /*tirel*/
 V = <hyponyme> = {INSTRUMENT}
 N = /*bouchonl*/

Nous appliquons alors aux composés exocentriques du type NCPV la même règle qu'aux composés endocentriques, à savoir que c'est le pivot de la structure qui véhicule le sens de l'hyponyme. C'est donc le nœud V de la description (26) qui se verra attribuer le trait « d'accord » <hyponyme>, lequel partagera sa valeur lexicale explicite avec le nœud DET de l'arbre dérivé après substitution. Ce partage de traits fera que la variable hyperonymique du DET prendra la valeur {INSTRUMENT} du nœud V', ce qui assure l'interprétation correcte des NCPV.

3. Conclusion

Dans notre analyse de certains faits de composition lexicale, nous avons rigoureusement appliqué la logique grammaticale que dicte une saisie directe de certaines données telles qu'un parseur ou un analyseur pourrait les traiter. En prenant appui sur certains cas patents d'exocentricité catégorielle, cette logique conduit à prendre acte de la nature syntaxique des suites polylexémiques que l'on interprète comme entrées lexi-

¹⁰ Précisons que la loi de l'hyponymie en composition lexicale est implicitement postulée dans Boiesdon & Tamba (1991) et Cadriot (1991). Elle l'est explicitement dans Corbin (1992).

cales. Les implications théoriques de cette approche n'ont rien de trivial parce qu'elles sollicitent un réaménagement profond de l'architecture d'une grammaire générative. Le concept de syntaxe dérivationnelle que nous venons de mettre de l'avant est une avenue prometteuse, croyons-nous, pour le traitement informatisé des mots composés du français. L'idée de base est que l'isomorphie des structures syntaxiques fait reporter sur la manipulation de certains traits sémantiques la tâche de générer la partie prédictible (sémantique) de l'interprétation lexicale. En définitive, la néologie par composition ne procède pas au coup par coup au gré des mots de la langue. La néologie par composition est au contraire fortement dépendante de la structure parce qu'elle procède en utilisant l'information syntaxique pertinente aux règles d'interprétation sémantiques qui sous-tendent la formation des entrées lexicales polylexémiques.

Outil d'intégration de bases de connaissances lexicales aux analyseurs syntaxiques

Philippe BLACHE et Mireille DELPUY

2LC - CNRS, Sophia-Antipolis, France

1. Introduction

Les lexiques informatisés constituent le cœur de tout système de traitement automatique du langage naturel. La qualité de leur conception, mais aussi leur exhaustivité est un prérequis indispensable en particulier pour les outils nécessitant large couverture et robustesse : étiqueteurs de corpus, analyseurs couvrants, etc.

Par ailleurs, nous assistons ces dernières années à un vaste courant de lexicalisation des théories linguistiques qui consiste à représenter au niveau lexical un grand nombre d'informations syntaxiques (par exemple, la structure argumentale ou les relations de spécification) ou sémantiques (s'appuyant évidemment sur la sémantique lexicale). Cette propriété est caractéristique des théories basées sur les contraintes. Nous nous intéressons plus particulièrement ici à la théorie HPSG qui entre toutes est celle utilisant les lexiques parmi les plus complexes.

Nous avons donc besoin non seulement d'importantes ressources lexicales à proprement parler, mais aussi d'outils facilitant l'utilisation de ces ressources en termes de conception, d'accès, d'intégration à des systèmes ainsi que de maintenance et d'évolution du lexique lui-même.

On remarque cependant qu'il n'existe qu'une très faible réutilisation des outils lexicaux. En d'autres termes, le développement de systèmes de traitement du langage naturel (notamment d'analyseurs syntaxiques) conduit généralement à la mise au point de nouveaux outils, en particulier concernant l'accès lexical. Le premier objectif du gestionnaire décrit ici est très concret et porte sur la mise à disposition d'un système autonome et facilement intégrable à un analyseur. Mais l'intérêt essentiel de cette approche repose sur le fait que le lexique source utilisé est considéré comme étant en quelque sorte un lexique générique. Il ne s'agit donc pas d'adapter un lexique au formalisme requis par un analyseur donné, mais au contraire de proposer un environnement capable de filtrer ce lexique en fonction de spécifications particulières.

Pratiquement, cet environnement permet la gestion du lexique lui-même, et propose de plus un utilitaire d'accès efficace. Par ailleurs, le module de génération et de maintenance du format lexical HPSG permet de créer, à partir d'une entrée lexicale générique, le schéma de l'entrée lexicale HPSG correspondante. Ce type d'outil présente un double intérêt : il permet, d'une part, de faciliter considérablement la tâche de la constitution d'un lexique HPSG et constitue, d'autre part, un véritable module de filtrage à partir d'un lexique générique. Nous nous situons ainsi dans la perspective de nombreux projets nationaux et internationaux qui ont pour objectif la mise au point de ressources génériques en fournissant un outil de mise en œuvre pour un cadre formel donné¹.

Cet article comporte trois parties : (i) description des problèmes posés par les lexiques HPSG, (ii) présentation du gestionnaire et (iii) intégration à un analyseur.

2. Le lexique en HPSG

HPSG représente toutes les informations à l'aide de traits. Ces données sont hiérarchisées à l'intérieur de structures de traits permettant la description de relations complexes entre les éléments. Le mécanisme explicatif de base en HPSG repose en effet sur le partage de structure (exprimé sous forme de coindexation de sous-structures). La complexité des entrées lexicales HPSG (quelle que soit la catégorie syntaxique) provient donc à la fois de la richesse des informations, mais aussi de la représentation du partage de valeur.

La figure 1 correspond à l'entrée lexicale du déterminant *un*. Les informations principales contenues dans cette structure concernent les aspects syntaxiques et sémantiques et sont de plusieurs types. Les informations *endocentriques* portent sur les caractéristiques internes à l'item. Il s'agit, par exemple, de la partie du discours ou de la structure sémantique (dans ce cas, la quantification). On remarquera au passage que la valeur du trait QMEM gérant la portée de la quantification est partagée avec celle de CONTENU. Mais on trouve également des informations *exocentriques* portant sur les relations de ce déterminant avec un autre constituant, en l'occurrence le nom. Ceci se fait par l'intermédiaire du trait SPEC indiquant que le déterminant spécifie un nom. La valeur du trait CONTENU de ce nom (représentée par l'indice [1]) est partagée avec la valeur du trait sémantique RESTIND. On remarque de plus qu'une restriction est appliquée au nom sélectionné : celui-ci doit être au masculin singulier. En d'autres termes, le déterminant ne porte pas directement de traits d'accord, mais stipule simplement des propriétés de l'objet sélectionné.

1. Il existe en particulier sur ces questions un groupe de travail *Lexique* dans le cadre du GDR-PRC Communication Homme-Machine. Contact : Damien.Genthial@imag.fr

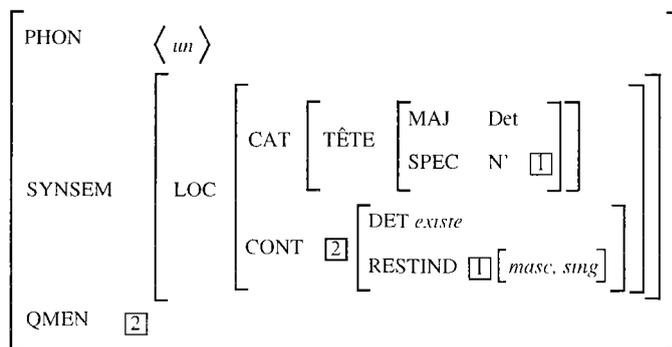


FIGURE 1 : Entrée lexicale de *un*

L'entrée lexicale d'un nom comme celle décrite dans la figure 2 comporte plus d'informations directes. On y remarque en effet des traits sémantiques plus détaillés ainsi que des traits d'accord (faisant partie en HPSG des traits sémantiques). La structure argumentale de cette tête est quant à elle décrite dans le trait VALENCE qui spécifie la sélection d'un déterminant (trait SPR).

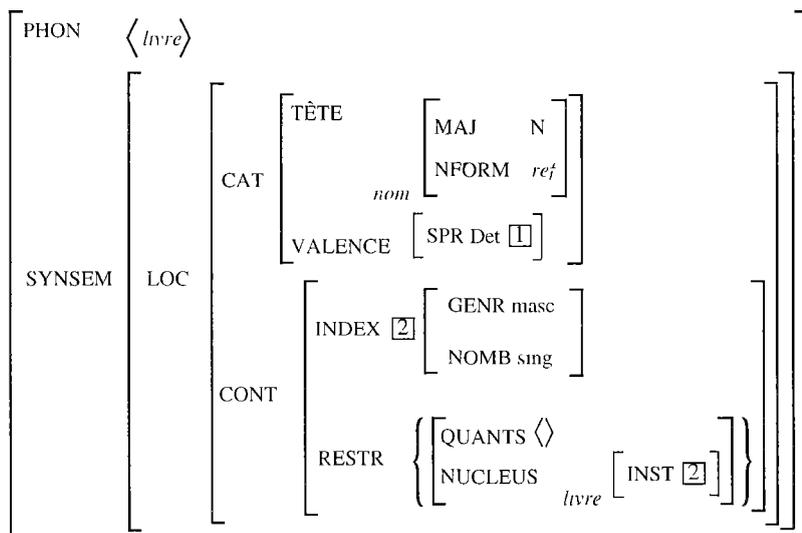


FIGURE 2 : Entrée lexicale de *livre*.

D'une façon générale, ces exemples révèlent deux types de difficultés rencontrées lors de la création d'un lexique HPSG :

- la structuration des données ;
- la représentation du partage des structures.

Il n'est donc pas envisageable de créer « manuellement » un véritable lexique HPSG. Il est en revanche intéressant de tirer partie de certaines propriétés et notamment d'utiliser l'héritage comme solution (désormais classique, cf. Briscoe *et al.*,

1993) de représentation générique. Il est en effet possible (nous le verrons dans la dernière section) de définir un certain nombre de schémas lexicaux constituant la base du lexique. Le système que nous décrivons propose ainsi le formatage d'entrées HPSG à partir d'un lexique général.

3. Présentation du système

L'objectif principal de ce système est de fournir aux analyseurs syntaxiques la possibilité de réutiliser d'importantes bases lexicales. Ceci nécessite une gestion optimale du lexique permettant sa maintenance, son évolution, un accès rapide à ses données ainsi qu'une adaptation efficace des formats du lexique de l'analyseur. Cette adaptation se présente, nous y reviendrons dans la section suivante, comme un module de génération et de maintenance des entrées utilisées par l'analyseur à partir d'entrées lexicales génériques.

D'une façon générale, le système propose :

- la recherche dans le lexique des mots à analyser. Cette recherche s'effectue via le gestionnaire lui-même ;
- une heuristique proposant en fonction de critères de fréquence et de contexte la meilleure probabilité de catégorisation ;
- l'adaptation des formats du lexique générique au formalisme utilisé.

L'outil que nous présentons utilise comme source le lexique LGE (Lexique Général Étendu), développé dans le cadre du projet BDLEX (bases de données et de connaissances lexicales du français écrit et parlé, cf. Perennou, 1988). Chaque entrée lexicale contient des informations phonologiques et syntaxiques enregistrées dans plusieurs champs. Elles se présentent sous la forme suivante :

GRAPH_ACC	HG	PHON_SYLL	FPH	HP	CL_PHON	NS	FREQ	CS	GN	CFLB	DEFLIEN	CONST
<i>amer</i>	11	&/m&	r"	11	E/NE	2	C2B0TR	V	01	060	**	Ta
<i>ambe</i>	11	A/mib	e	11	A/NID	2	-	N	Fn	81	-	• -
<i>elle</i>	11	&/l	-	32	EL	1	COB0	P	FS	s3	elles	-

Sans entrer dans les détails, ces champs représentent respectivement la représentation graphique de l'entrée, son numéro d'homographe, la représentation phonologique, le fonctionnement de la finale, le numéro d'homophone, la représentation phonologique en classes majeures, le nombre de syllabes, la fréquence, la catégorie, la variation genre/nombre, la catégorie flexionnelle, la défection temps/personne, la structure argumentale.

Le gestionnaire décrit ici a été intégré notamment à un analyseur syntaxique HPSG. L'adaptation consiste donc à générer des structures de traits à partir des informations lexicales contenues dans les entrées. L'outil a été développé en Prolog III (cf. Delpui, 1993), il est basé sur des règles de réécriture qui effectuent l'adaptation. Il pourrait être étendu afin de permettre à l'utilisateur de définir les champs utiles dans les entrées lexicales, et de spécifier les renseignements nécessaires au codage des structures de traits. Ce dernier point est décrit dans la section suivante.

3.1. Gestionnaire du lexique

La nécessité pour les systèmes de traitement du langage naturel de disposer d'une bonne couverture impose l'utilisation de lexiques très complets. Un des problèmes posés provient du fait qu'il s'agit de ressources volumineuses. Leur gestion doit être optimale, tant pour l'accès aux données que pour leur maintenance et l'évolution du lexique lui-même.

La structure interne du lexique est importante, car de ce choix va dépendre la rapidité d'accès, de modification, et d'ajout d'entrées lexicales. Nous avons choisi d'utiliser une représentation arborescente particulière appelée *Arbres 2-3*. Ceux-ci sont des cas particuliers des B-arbres :

- chaque nœud intérieur possède 2 ou 3 fils ;
- chaque chemin de la racine aux feuilles est de même longueur ;
- seules les feuilles contiennent les informations des mots ;
- si un élément A est à gauche d'un élément B dans l'arbre, alors la relation clé (A) < clé(B) est vraie. La clé est un champ particulier qui permet de définir l'ordre de classement dans l'arbre.

Ces propriétés font de cet arbre, un arbre équilibré possédant à chaque nœud au plus 3 fils. Ceci autorise un accès rapide aux données pour les consultations et les modifications. L'ajout d'un nouvel élément peut nécessiter un équilibrage de l'arbre. Ce travail s'effectue rapidement en « remontant » les données sur les nœuds internes.

L'outil que nous présentons est basé sur le lexique LGE contenant 22 930 mots. Son chargement dans l'arbre s'effectue en 9 secondes². Les mots sont ordonnés suivant leur représentation graphique. Le gestionnaire a été développé en langage C. Une interface conviviale³ permet de manipuler aisément ce lexique. Tous les accès fournis (consultation, ajout, modification) utilisent directement l'arbre.

4. Intégration à un analyseur

L'intégration du gestionnaire lexical à un analyseur repose à la fois sur la généralité des outils développés (leur portabilité, le format des données, etc.) et sur la capacité d'adaptation au format de l'analyseur lui-même. Nous avons vu dans la section précédente les aspects concernant l'accès et la gestion du lexique. Cette partie est consacrée à la présentation du filtrage nécessaire à la constitution d'un lexique dans un format donné.

4.1. Le filtrage

Le format du lexique source est considéré comme générique. Seules certaines informations seront utiles à la construction d'une entrée lexicale dans un format donné.

2. Ce résultat est obtenu sur une station DEC 5000

3 L'interface est écrite sous X11-XTToolkit et Athena Widget

Pour ce qui concerne l'analyseur HPSG évoqué ici, les informations nécessaires, comme nous l'avons vu dans la première section, sont de deux types et concernent soit des informations que nous qualifierons de *discriminantes* concernant les propriétés inhérentes à la forme rencontrée, soit des informations *génériques* portant notamment sur les relations entre les structures de traits représentées par la coindexation.

L'exemple de la figure 3 indique les champs de l'entrée du lexique source pour le verbe *aimer* utilisés pour la construction de l'entrée HPSG. Nous remarquons ainsi que seuls quelques champs du source sont effectivement récupérés. Il s'agit de PHON_SYLL, CS et CONST utilisés respectivement en tant que valeurs des traits PHON, MAJ et COMPS de la structure HPSG. Le champ GRAPH_ACC du source permet quant à lui d'informer le type de la relation sémantique du trait NUCLEUS.

Ce type d'opération de filtrage est similaire pour les autres catégories syntaxiques. Il est donc possible de définir de véritables *schémas de filtrage* qui indiqueront les champs du lexique source à utiliser, leur destination dans le lexique cible et le partage des structures de l'entrée construite. Ainsi, le schéma de filtrage pour le verbe se présentera sous la forme de la figure 4.

La structure de traits décrite dans la figure 4 indique à l'aide d'un astérisque les champs du lexique utilisés en tant que valeur de trait. On remarquera de plus l'utilisation de variables pour la représentation de la coindexation.

Le sous-ensemble de champs utilisés par le filtrage est bien entendu extensible. En particulier, toutes les informations phonologiques doivent être également utilisées pour la construction d'entrées plus complètes. Ces champs sont sujets à modification en fonction du lexique source utilisé. Mais le principe du filtrage reste le même.

GRAPH_ACC	HG	PHON_SYLL	FPH	HP	CL_PHON	NS	FREQ	CS	GN	CFLB	DEFLIEN	CONST
<i>aimer</i>	11	&/m&	i"	11	E/NE	2	C2B0TR	V	01	060	1+	Ta

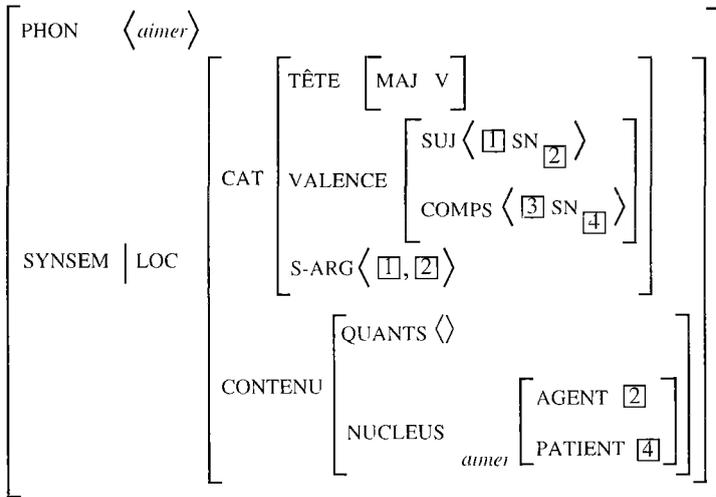


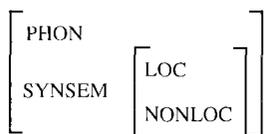
FIGURE 3 : Filtrage de l'entrée lexicale *aimer*

4.2. Implantation

L'implantation de ces schémas de filtrage se fait à l'aide d'un utilitaire permettant de gérer de façon autonome ces informations. Ce module est en effet distinct de l'analyseur lui-même dans la mesure où les schémas sont stockés sous forme de fichier textes composés de la description des structures en question. L'analyseur accède à ces informations via un ensemble de fonctions permettant leur interrogation, mais également leur manipulation. Elles sont représentées par un arbre binaire qui contient à chaque nœud :

- le trait concerné ;
- un arbre fils qui représente les sous-traités du trait ;
- un arbre frère qui représente les traits qui ont même trait père que le trait du nœud.

Ainsi, pour la structure suivante :

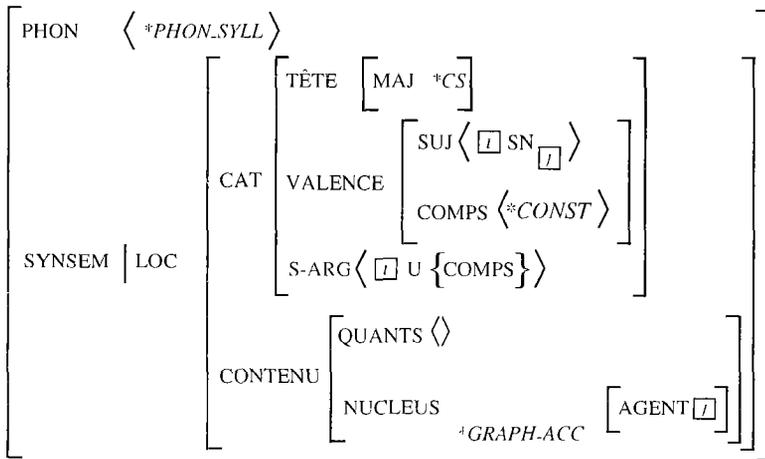


le nœud correspondant à SYNSEM aura :

- pour fils l'arbre composé des nœuds LOC et NONLOC ;
- pour frère l'arbre composé du nœud PHON.

Le gestionnaire permet de générer ces structures à partir des structures syntaxiques du modèle. Il fournit également une interface pour modifier, créer, visualiser les règles.

Cette méthode présente plusieurs avantages. Tout d'abord, nous avons ainsi une séparation claire des données et de l'analyseur. De plus, cet outil permet la maintenance du format des structures. On sait en effet que celles-ci, en fonction des évolutions de la théorie, peuvent être sujettes à des remaniements plus ou moins importants.

FIGURE 4 Filtrage de l'entrée lexicale *aimer*.

5. Conclusion

L'intégration du gestionnaire a été expérimentée sur plusieurs systèmes. Il a, dans sa première version, été utilisé par un correcteur grammatical reposant sur le cadre formel des GPSG (cf. Meyer, 1993). D'un point de vue pratique, le gestionnaire est écrit en C et le système en Object Pascal. Sa seconde utilisation a été expérimentée sur un analyseur HPSG écrit en Prolog III. Son intégration à un nouvel analyseur écrit en **LIFE** est en cours de développement. L'avantage dans ce dernier cas vient du fait que **LIFE** permet de manipuler directement les structures de traits et propose un mécanisme d'héritage sous forme de contraintes. Les schémas d'instanciation décrits plus haut sont donc implantés sous forme de propriétés de catégories syntaxiques héritées par les entrées lexicales elles-mêmes. L'accès au lexique se contentera de retourner les valeurs utiles.

Cette approche permet d'envisager l'utilisation d'un lexique véritablement générique. Il s'agit dans ce cas de proposer un lexique comportant un très grand nombre d'informations ainsi qu'un ensemble de modules de filtrages permettant l'adaptation des entrées à des formats donnés. Il serait ainsi possible soit de construire directement des lexiques adaptés à un formalisme donné (*i.e.* de compiler le lexique générique source) soit de maintenir un filtrage dynamique permettant une utilisation dans des plate-formes plus complexes mettant en jeu par exemple plusieurs formalismes.

Un réseau lexico-sémantique des verbes construit à partir du dictionnaire pour le traitement automatique du français

Karim CHIBOUT et Nicolas MASSON

Groupe Langage et Cognition, LIMSI-CNRS, Orsay, France

*Survoler, c'est « voler au-dessus de » ;
donc survoler est en dessous de voler !*

• Abstract •

Our work turns on computational modelling of verbal polysemy based on a hierarchical representation of verbs (in French) The final goal is to integrate into a parser a specific module which detect verbs used in a figurative sense and interpret them in order to remove semantic incoherences.

We propose a systematic and comprehensive representation of verbs in a hierarchical structure. Our rationale for a classification seen as the keystone of an interpretation system is the following : the different senses of a verb are variations semantically related ; they share a common basis, that is, a common thematic relation structure, and a common basic action. Different senses derive from inflexions in the thematic relation structure, and/or from the sense of the hyperonym, as it is recursively defined in terms of its hyperonym and its own thematic relation structure. The hierarchy of verbs depends on the thematic relations associated with them.

From this representation we have determined three major types of heuristics to help interpreting the different meanings conveyed by a verb and its hyponyms.

The verb network and the specific module which detect verbs used in a figurative sense are used in a scientific texts structuring system. In this application, structuring consists in determining rhetorical relations between sentences or sets of sentences. Some other applications are also discussed.

Introduction

Notre travail porte sur la résolution des incohérences sémantiques liées à la polysémie des verbes. Le principal but est de réaliser un module informatique « expert » dans la

détection et le traitement de sens figurés au niveau de la phrase. On se limitera à l'étude des verbes d'action.

La modélisation de la polysémie est menée selon deux axes complémentaires :

- la représentation des connaissances lexicales ;
- les processus d'interprétation des différents sens susceptibles d'être associés à un prédicat et les modalités de discrimination entre ces significations dans un énoncé donné.

Après un aperçu de quelques études sur les métaphores qui justifient notre approche, nous présenterons les liens entre les verbes au sein d'un réseau lexical et la structure sémantique associée à chacun d'eux. Nous définirons ensuite le modèle de la polysémie proposé à partir de ce formalisme de représentation. Nous terminerons par les autres applications possibles du formalisme élaboré, et en particulier l'aide à la structuration de textes scientifiques.

1. D'une figure de sens aux sens figurés

Nous partons de l'étude d'énoncés mettant en jeu des verbes employés métaphoriquement. La métaphore verbale a été étudiée en particulier par Wilks (1978), qui a réalisé un programme destiné à résoudre les incohérences sémantiques propres aux métaphores verbales via une analogie pertinente entre les termes. La phrase suivante montre le principe général de son système :

My car drinks gasoline
(Ma voiture « boit » de l'essence)

nous sommes en présence d'une anomalie de sens dans la mesure où le verbe *drink* attend un *animé* comme cas agent et un cas objet de type *liquide*. Une métaphore est mise en évidence par la comparaison du verbe *drink* et d'un sens de *use* qui prend comme objet *gasoline*.

Une primitive sémantique commune aux deux verbes (Expend) permet le remplacement de *drink* par *use* dans la phrase. [Expend] désigne une action de base signifiant **dépenser, consommer (énergie, argent...)**, qui s'applique tout aussi bien à *boire* pour un animé qu'à *consommer* pour une voiture. Le verbe *boire* peut prendre le sens d'*ingurgiter* ou de *consommer*. C'est ce deuxième sens qui convient pour cet exemple, l'acte d'*ingurgiter* étant propre aux animés. Par conséquent, le verbe *use* est retenu pour l'interprétation :

My car uses (consomme) gasoline.

D'autres programmes sont fondés sur un principe similaire ; en particulier le système Met* (Metstar) de D. Fass (1991) qui résout, dans sa forme actuelle, les métaphores verbales fondées sur un lien d'analogie entre deux termes et un nombre réduit de métonymies.

Ce type d'analyse permet d'accepter certains énoncés métaphoriques dans lesquels les préférences de sens sont violées.

Nous constatons que dans une organisation hiérarchique de verbes, plus on descend dans la hiérarchie plus les relations casuelles vont être spécifiques et imposer une signification précise (p. ex. *drink*), et *a contrario*, plus on s'élève dans la hiérarchie, moins ces contraintes casuelles vont jouer (p. ex. *use*). Un des modes de résolution des métaphores va donc consister à relâcher certaines des contraintes imposées sur les relations casuelles d'un verbe en remontant dans l'arbre.

Notre propre étude, qui a donné lieu à une implantation sous forme d'un module d'interprétation d'énoncés métaphoriques (Chibout, 1994), montre que ce processus ne concerne qu'un nombre restreint de métaphores verbales.

Les verbes métaphorisés contenant dans leur description **la manière** particulière dont une action est effectuée (exprimée généralement par un adverbe) imposent une interprétation différenciée. En plus de la relaxation de contraintes par la récupération d'un hyperonyme, il est nécessaire, pour une interprétation correcte de la figure, de rattacher l'adverbe de manière à l'hyperonyme. L'exemple suivant répond à ce mécanisme :

*La mer, devenue calme, **berçait** l'embarcation de fortune.*

Le verbe *bercer* a pour description grossière :

balancer (hyperonyme immédiat) *lentement* (manière) *pour endormir* (but).

Berçer a pour cas agent un *humain* et comme cas objet un *humain*. Le système détecte une incohérence sémantique sur les deux liens casuels.

La résolution consiste à rattacher à l'hyperonyme immédiat (que nous appellerons **action de base**) le trait manière afin d'obtenir un énoncé plus sensé :

*La mer, devenue calme, **balançait lentement** l'embarcation de fortune*

Cette idée de « glissements » de sens à partir d'une description sémantique nous a amené à ne plus modéliser une figure particulière isolée du reste des faits de langue mais à aborder la **polysémie** dans son ensemble.

Wilks (1978) et Fass (1991) se sont attachés davantage à l'interprétation des métaphores, qu'à la représentation des connaissances nécessaires à leur résolution. Or il nous semble, au regard des exemples cités, qu'une représentation lexicale suffisamment élaborée est le déterminant majeur pour l'interprétation, c'est-à-dire que les mécanismes à mettre en œuvre en découlent.

Par conséquent, dans notre démarche, nous avons recherché, pour la représentation des connaissances lexicales, un formalisme suffisamment souple pour rendre compte de la polysémie, et suffisamment puissant en termes d'opérations sur les représentations, pour pouvoir déduire aisément la dynamique d'interprétation de la polysémie verbale.

En effet, dans quelle mesure la description sémantique d'un verbe ne pourrait-elle pas être utilisée pour traiter les sens figurés portés par celui-ci ?

Dans les exemples donnés, la description d'un verbe qui semble pertinente pour le traitement des métaphores implique son action de base et des composants de sens qui vont spécifier ce verbe.

Nous reprenons donc ces éléments pour l'étude de la polysémie en créant, d'une part, un réseau lexical hiérarchique pour rendre compte des relations d'hyponymie/hyperonymie entre verbes et, d'autre part, une représentation détaillée associée à chaque verbe sous la forme d'une structure de cas sémantiques.

2. Représentations des verbes

2.1. Quelques travaux

Les classifications du lexique fondées sur des bases sémantiques sont relativement nombreuses en intelligence artificielle. Les réseaux sémantiques sont un des principaux modes de représentation des connaissances lexicales ; mais ils concernent essentiellement les substantifs.

Il existe quelques tentatives d'organisation hiérarchique fine des verbes pour l'anglais (Talmy, 1985), (Miller et Fellbaum, 1991 ; Miller *et al.*, 1989). Les auteurs soulignent par ailleurs la complexité de la tâche, en raison notamment des différents champs sémantiques impliqués dans les relations entre les verbes étant apparentés au niveau du sens.

Talmy (1985) fournit certains critères sémantiques utiles à la constitution d'une taxonomie. Son analyse des verbes de mouvements est la plus significative de ce point de vue. Il met en évidence une primitive de mouvement commune à tous les verbes et d'autres composants sémantiques comme la *manière* et la *cause*, à l'instar de *slide* (glisser) et *pull* (tirer) respectivement. À ces composants peuvent s'ajouter la *rapidité* (*run, stroll*) ou le *moyen de transport* utilisé (*bus, truck, bike, etc.*).

D'une manière similaire, les verbes qui dénotent l'action de frapper (*hit*) peuvent exprimer le *degré de force* utilisé par l'agent (*chop, slam, whack, swat, etc.*).

D'autres verbes font référence au *degré d'intensité* d'une action ou d'un état (*drowse, doze, sleep,...*).

Les travaux en psycholinguistique de Miller et ses collaborateurs (*op. cit.*) portent sur les relations sémantiques liant les mots de trois catégories lexicales : les noms, les adjectifs et les verbes. Les mots de chaque catégorie sont reliés entre eux par un ensemble de relations telles que synonymie, antonymie, hyperonymie/hyponymie, etc.

Concernant les verbes, les auteurs reprennent les travaux de Talmy et définissent une relation de particularisation permettant de réunir les différents composants sémantiques qui distinguent un verbe de son superordonné (hyperonyme). Cette relation

entre deux verbes (V1 hyponyme de V2) est appelée **troponymie** (du grec *tropos* : manière, façon) et est exprimé par la formule « accomplir l'action V1, c'est accomplir l'action V2 d'une manière particulière ».

Le verbe *fight* a pour troponymes *battle*, *war*, *tourney*, *duel*,... Les troponymes des verbes de communication impliquent l'intention du locuteur ou sa motivation à communiquer, comme dans *examine*, *confess*, *preach*, ou le moyen de communication utilisé : *fax*, *email*, *phone*, *telex*,...

Leur étude a donné lieu à une mise en œuvre informatique d'un réseau lexical (Wordnet) qui organise les mots des différentes catégories lexicales en termes de signifiés.

Pour la catégorie qui nous intéresse, ce travail constitue, à notre connaissance, une des rares tentatives de hiérarchisation systématique de verbes dans un réseau sémantique.

2.2. Notre classification

Partant des travaux exposés ci-dessus, un dictionnaire de type terminologique nous semble le seul outil qui puisse expliciter les composants de sens nécessaires à la constitution d'un réseau lexico-sémantique des verbes d'action.

Les définitions sont en effet des descriptions assez précises qui incluent en général l'hyperonyme et font des renvois aux verbes qui ont une parenté de sens avec celui défini. On y retrouve également les différents composants sémantiques (moyen, manière, but, etc.) qui vont spécifier le verbe par rapport à son superordonné.

Les définitions ne sont cependant pas toujours homogènes dans leur structure et leur contenu de sorte qu'on ne peut pas s'appuyer totalement sur le dictionnaire pour réaliser notre réseau. Des dissimilarités se retrouvent dans les définitions selon le dictionnaire utilisé.

N. Wurbel (1995 : 50), qui a étudié un grand nombre de dictionnaires pour définir un automate de génération de définitions pour les substantifs, montre en particulier que les « dictionnaires courants [...] ont des niveaux de description d'informations extrêmement variés ».

Ce phénomène se retrouve pour les verbes. Ainsi le sens habituel de *couper* est formulé de différentes façons selon le dictionnaire consulté :

- **Collins Cobuild English Dictionary** (Collins, 1995) :
Cut (something) : ...use a knife or a similar tool to divide it into pieces,...
- **Hachette, dictionnaire de notre temps** (Hachette, 1993) :
diviser avec un instrument tranchant.
- **Le Petit Littré** (Beaujean, 1990) donne une définition proche de la précédente :
diviser un corps avec un instrument tranchant.
- **Le Petit Robert** (Robert, 1994) :
Diviser (un corps solide) avec un instrument tranchant.

Au travers de ces définitions, on constate principalement des différences dans la précision des descriptions du prédicat. Chacune donne l'action générique (*diviser*) et le moyen propre à l'acte de couper (*instrument tranchant*). Mais en ce qui concerne l'objet de l'action, seule la définition donnée par *Le Petit Robert* comporte une caractéristique importante de l'objet (*corps solide*).

L'exemple suivant montre de manière encore plus significative les disparités qui peuvent exister entre les dictionnaires.

Le sens propre de *élaguer* est défini comme suit :

- **Le Petit Robert :**
dépouiller (un arbre) des branches superflues sur une certaine hauteur.
- **Le Petit Littré :**
couper les branches, principalement les branches inférieures d'un arbre.
- **Collins Cobuild English Dictionary :**
prune a tree or bush : cut off some of the branches so that it will grow better the next year.
- **Hachette, dictionnaire de notre temps :**
débarrasser (un arbre) des branches nuisibles à son développement, à sa fructification, etc.

Les dissemblances entre les dictionnaires sont ici plus marquées : ils ne font pas référence à la même action (*couper, débarrasser, dépouiller*), ni au même qualificatif pour l'objet de l'action (*branches superflues, inférieures, nuisibles,...*)

L'hétérogénéité dans la caractérisation des termes nous a conduit à ne nous appuyer que partiellement sur cet outil terminologique et à proposer une réorganisation des définitions.

Nous suggérons une systématisation des définitions en s'imposant des règles explicites : les verbes sont décrits par l'intermédiaire d'une structure canonique de traits sémantiques. Plus précisément, tout verbe est exprimé par son **action de base** (hyperonyme immédiat) et un certain nombre de cas sémantiques qui la spécifient. Outre les liens casuels habituels (agent, objet, moyen,...), 4 cas apparaissent essentiels pour la description fine d'un verbe d'action :

- la **manière** dont l'action est réalisée (exprimée par un adverbe) ;
- la **méthode** par laquelle elle s'effectue (exprimée par une action) ;
- son **résultat** (représenté par un état) ;
- le **but** intrinsèque à cette action (représenté par un verbe d'action).

La procédure consiste à récupérer les différentes définitions (sens propre) d'un verbe dans différents dictionnaires avec l'hypothèse que le sens maximal est disséminé dans ceux-ci.

Quand un composant vient à manquer dans les définitions mais qu'il est présent sous la forme de sens figuré et qu'il a une pertinence par rapport à la définition, nous l'intégrons dans notre représentation.

C'est le cas du verbe *couper* (ci-après), pour lequel la **méthode** et le **but** n'apparaissent dans aucun des dictionnaires consultés :

couper : *diviser* (action de base) *objet solide* (objet) *en plusieurs parties* (résultat) *à l'aide d'un instrument tranchant* (moyen) *en traversant l'objet* (méthode) *pour en enlever une partie ou en réduire la taille* (but).

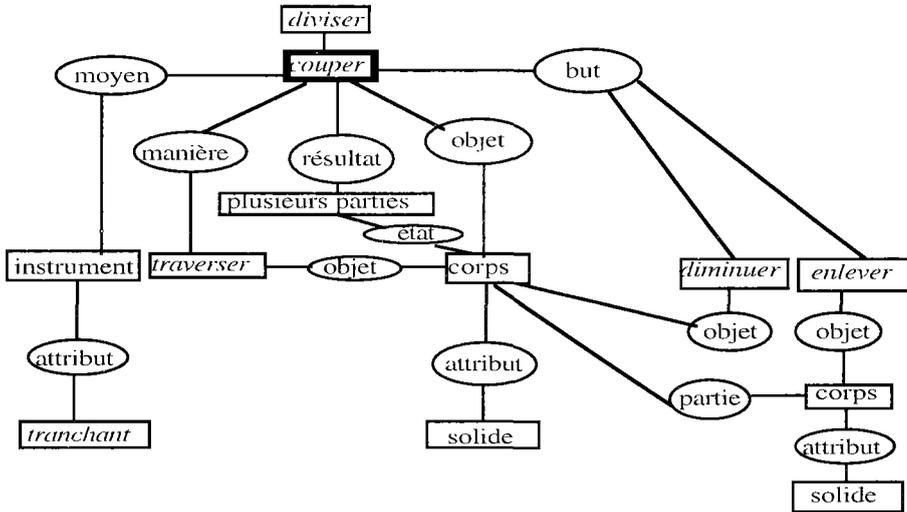


FIGURE 1 Représentation (action de base + structure casuelle) du verbe *couper*.

De plus, ces cas sémantiques permettent, nous allons le voir, de définir plus précisément des critères de hiérarchisation des verbes.

La hiérarchisation des verbes est faite en fonction des relations casuelles qui leur sont associées. Un verbe est hyperonyme (respectivement hyponyme) d'un autre s'ils ont une action de base commune et s'il y a dans la structure casuelle de ce verbe (cf. figure 2) :

1. **absence** (respectivement **présence**) d'une valeur définie pour un cas particulier ;
2. présence d'un cas à **valeur multiple** (respectivement à **valeur unique**) ;
3. présence d'un cas à **valeur générique** (respectivement à **valeur spécifique**).

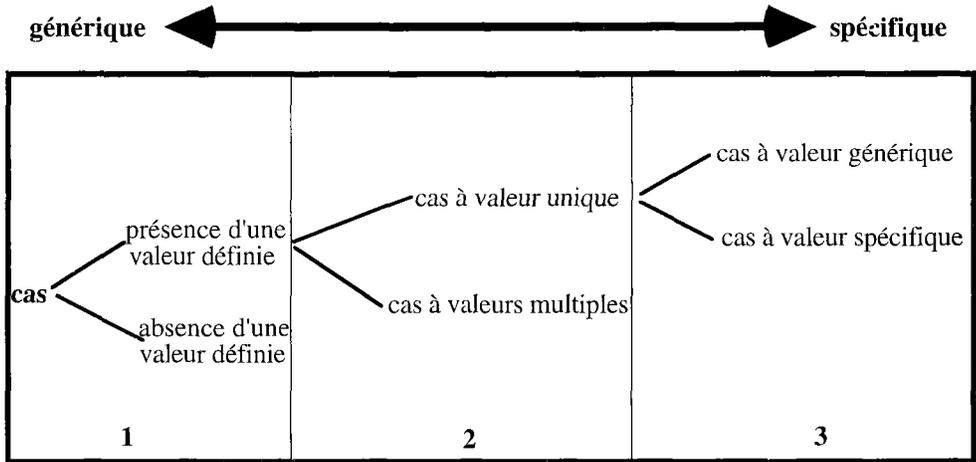


FIGURE 2 : Les valeurs portées par les relations casuelles comme critères de classification des verbes.

1. Exemple : pour la paire **diviser/couper** ; couper, c'est diviser avec un instrument tranchant (cas moyen défini), donc *diviser* est hyperonyme de *couper* (noté *diviser* > *couper*) ;
2. Exemple : pour **couper/élaguer** ; *couper*>*élaguer* parce que le cas but du premier peut être *enlever une partie* (couper les bords de pages), *réduire la taille* (couper du bois) ; alors que le but de *élaguer* est unique : *enlever* (les branches inutiles d'un arbre) ;
3. Exemple : **couper/décapiter**, couper a un cas objet à valeur générique : *objet solide* / alors que l'objet de décapiter est relativement spécifique (*tête*), donc *couper*>*décapiter*.

Le réseau apparaît comme une structure en arbre (cf. figure 3 ci-dessous). Les verbes décrivant des situations du monde, nous plaçons en haut de la hiérarchie l'**événement** qui définit les deux grands types de verbes (d'action et d'état).

Pour la catégorie verbes d'**action**, **faire** est hyperonyme de toutes les primitives sémantiques (p. ex. **faire cesser**) à partir desquelles est déterminée la hiérarchie des concepts verbaux, des plus généraux (p. ex. **séparer**) aux plus spécifiques (p. ex. **guillotiner**).

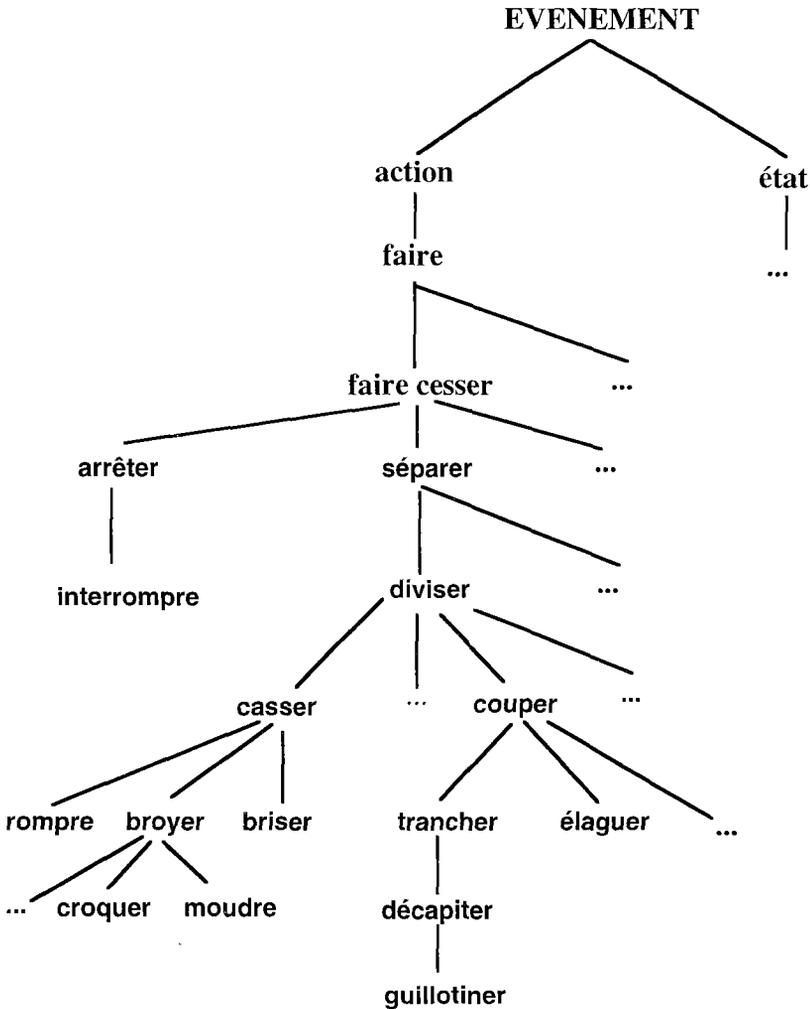


FIGURE 3 . Vue locale du réseau.

Le réseau et les schémas descriptifs des verbes sont principalement destinés au traitement de la polysémie verbale. La partie qui suit explicite le modèle de la polysémie envisagé et notamment les heuristiques mises en évidence pour l'automatisation de l'interprétation des sens figurés.

3. Application à la polysémie des verbes

3.1. Point de vue sur la polysémie

Nous formulons un certain nombre d'hypothèses à la base de notre étude

- Il existe des règles explicites (ou explicitables) sous-jacentes à la construction de

- la polysémie des verbes.
- Ces règles sont en nombre limité.
 - Les règles déterminant la polysémie des verbes prennent appui sur la description de l'action et les différentes caractéristiques (modalités de réalisation) propres à cette action. Ces modalités correspondent aux relations casuelles que nous avons définies auparavant.

Les sens figurés sont principalement contenus dans la définition du sens propre, telle peut être en résumé notre conception de la polysémie. De ce point de vue, une description sémantique fine d'un verbe permettra de retrouver l'ensemble, ou tout au moins une grande partie de ses sens.

Les différentes significations d'un verbe sont vues comme des variations sémantiquement apparentées ; elles partagent un support commun : la structure casuelle et hiérarchique du verbe telle que nous l'avons définie. Ces significations dérivent de l'ossature casuelle du verbe et/ou de la signification de son (ou d'un de ses) hyperonyme(s). De par l'organisation hiérarchique du réseau, un hyperonyme est lui-même défini récursivement par son hyperonyme et sa propre structure casuelle.

Or, si les différents sens d'un mot dérivent d'une même représentation, quels sont les **modes d'interprétation** à l'origine de cette dérivation de sens ?

3.2. Heuristiques de résolution

À partir de cette représentation, nous avons déterminé trois catégories d'heuristiques pour l'interprétation (sans toutefois prétendre à l'exhaustivité) des sens multiples véhiculés par un verbe et ses sous-ordonnés :

- **récupération d'un hyperonyme**

- l'action de base
Les coulées de lave marchaient vers le village (SE DÉPLACER)
- un hyperonyme plus général
Jean coupa l'eau (ARRÊTER)

Arrêter est récupéré à partir de la primitive sémantique **faire cesser** commune à *couper* et *arrêter* (cf. le réseau local au verbe *couper* figure 3). Notons que dans l'exemple la substitution de *couper* par *arrêter* n'est pas suffisante du fait de la présence d'une figure elliptique (*arrêter l'écoulement de l'eau*).

- **attachement d'un cas à l'hyperonyme récupéré**

- l'action de base
Les vagues couraient jusqu'aux rochers (se déplacer RAPIDEMENT)
- un hyperonyme plus général
Il brisa leur conversation (interrompre BRUSQUEMENT)

De la même manière que pour couper l'eau, interrompre est accessible via la primitive **faire cesser** qu'ils partagent.

- **transfert d'un cas comme action effective en remplacement du verbe polysème traité.**
 - Le cas méthode

Ce mode de résolution est illustré par l'exemple suivant :

Le paysan coupa par le champs (TRAVERSER)

dans lequel *traverser* se substitue à *couper*, parce que ce dernier contient dans sa description détaillée le cas méthode « en traversant l'objet » (cf. figure 1).

- Le cas manière

Le vieil arbre tranche avec le paysage

Trancher a, en particulier, pour action de base *couper* et pour cas manière *distinctement*. En redonnant une forme verbale au cas manière, on obtient après substitution :

Le vieil arbre se distingue du paysage

- Le cas but

Il a élagué son exposé (ENLEVER LES PARTIES INUTILES)

élaguer : *couper* (action de base) enlever les branches inutiles (but).

Ces heuristiques représentent autant de « glissements » de sens à partir de la structure unique qui définit le verbe.

Polysémie locale récurrente

Il est apparu que certains sens portés par un verbe sont également véhiculés par des verbes localement proches (*i.e.* ayant un hyperonyme en partage) dans le réseau.

Par exemple, les verbes *couper*, *entrecouper*, *briser*, *hacher* qui relèvent tous de la primitive sémantique **faire cesser** ont un sens figuré commun :

couper la parole (INTERROMPRE)

briser une conversation (INTERROMPRE BRUSQUEMENT)

entrecouper ses phrases de sanglots (INTERROMPRE FRÉQUEMMENT)

hacher un discours (INTERROMPRE DE FAÇON RÉPÉTÉE)

Ces heuristiques, qui répondent à la catégorie **recupération d'un hyperonyme**

avec en sus pour les trois derniers exemples, l'attachement du cas manière spécifique à chacun des verbes, valident partiellement le principe d'une hiérarchisation du lexique.

Cette récurrence de sens locale nous conforte donc dans l'adoption d'une structure arborescente pour représenter les concepts verbaux.

3.3. Les limites du modèle

Les heuristiques proposées ne permettent pas de résoudre la totalité des modes polysémiques des verbes. Mais si nous ne prétendons pas à l'exhaustivité, compte tenu de la complexité du phénomène, nous souhaitons parvenir à une meilleure couverture de la polysémie.

En outre, il reste à déterminer précisément les conditions de choix entre les différentes heuristiques. En effet dans quelle(s) condition(s) appliquer telle ou telle des heuristiques définies ?

Nous savons d'ores et déjà que les catégories sémantiques des valeurs portées par les cas agent et objet (quand ils existent) vont servir de filtres pour mettre en évidence un éventuel sens figuré du verbe (incohérence sémantique).

L'incohérence si elle porte sur l'un ou l'autre de ces cas ne met-elle pas en jeu des heuristiques différenciées ?

Les catégories sémantiques de ces éléments n'impliquent-elles pas des modes de résolution particuliers ? On pense notamment aux deux dimensions sémantiques : **abstrait/concret** et **animé/inanimé**.

La détection de sens figurés pose elle-même problème dans la mesure où la violation des préférences casuelles n'est pas systématique. L'exemple qui suit est significatif de ce problème :

Il a coupé les cartes.

Malgré le fait que les contraintes casuelles sont respectées, il subsiste une ambiguïté sur le sens du verbe *couper*. S'agit-il de couper effectivement les cartes ou de diviser le paquet, comme il se fait d'ordinaire au début d'une partie de cartes ?

Le modèle est limité au traitement des sens figurés induits par une incohérence sémantique entre les valeurs attendues par les cas du verbe et les valeurs en présence dans l'énoncé. Les ambiguïtés de sens vont être considérées comme relevant d'un niveau d'interprétation plus poussé fondé sur des règles sémantico-pragmatiques qui restent à préciser.

Il existe des expressions dont l'interprétation est subordonnée à l'application d'une suite de traitements. Nous les appellerons **sens figurés en chaîne**, généralisant ainsi la notion de chaîne proposée par Reddy (1979) pour la figure métonymique.

Ce type d'expressions nécessite parfois des structures de représentation pour les **substantifs** afin de mettre en œuvre des règles propres à traiter des sens figurés liés à ceux-ci.

décapiter une entreprise

Pour retrouver le sens de *décapiter*, il faut non seulement récupérer le cas but (détruire) de ce verbe, mais également conserver la valeur du cas objet (*tête*) utilisé dans un de ses sens métaphoriques et le traiter comme tel.

4. Un réseau utile pour d'autres applications du TAL

Aide à la structuration de textes scientifiques

Le réseau de verbes sera utilisé dans le cadre d'un système de structuration de textes scientifiques en vue de réaliser la génération automatique de résumés. La structuration consiste ici à déterminer et étiqueter les relations rhétoriques (de type argumentatives et logiques) entre des phrases ou des ensembles de phrases du texte. Ces relations sont en nombre fini et expriment la *causalité*, le *contraste*, le *renforcement d'idée*, l'*illustration*, etc. La structuration réalisée s'appuie sur deux modules :

- un module de repérage des bornes thématiques du texte. Celui-ci utilise des techniques statistiques qui portent sur l'analyse de la distribution des occurrences nominales dans le texte ;
- un module qui détermine le rôle des phrases les unes par rapport aux autres, c'est-à-dire établir les liens rhétoriques ou argumentatifs entre les phrases. Ce module utilise les résultats du précédent et met en œuvre une analyse linguistique de surface fondée sur un modèle linguistique d'analyse de textes scientifiques que nous avons développé (Masson, 1995). À ce niveau nous analysons différents phénomènes linguistiques tels que les *portées temporelles*, les *connecteurs*, certaines *expressions remarquables* (p. ex. *En conclusion...*) ainsi que les *verbes référents à l'argumentation*. C'est pour l'analyse de ces derniers que l'on utilise le réseau de verbes et le module de résolution de la polysémie précédemment explicités.

À chaque relation sont associés des verbes qui expriment la signification de la relation (p. ex. **renforcement d'idée** : *confirmer*). La détection, dans une phrase du texte, d'un verbe attaché à une relation (p. ex. *confirmer*, *étayer*, *renforcer*, *appuyer...*) donne le rôle argumentatif de la phrase. L'exemple suivant montre une relation de renforcement d'idée entre la première et la deuxième phrase.

(...) les peuples de ces régions l'utilisent [l'asphalte] dans le bâtiment, pour la réalisation de jouets, d'objets d'art, d'objets d'usage domestique. L'analyse détaillée des échantillons de bitume des diverses sources géologiques par des méthodes isotopiques et moléculaires a confirmé cette hypothèse. [« Les bitumes de Suse », Pour la Science, n° 204, octobre 1994].

Dans cet exemple, le verbe (*confirmer*) est un verbe d'argumentation strictement, c'est-à-dire qu'il est utilisé au sens littéral. Imaginons maintenant qu'à la place de *confirmer* on ait eu le verbe *étayer*, soit une phrase du type :

L'analyse détaillée des échantillons de bitume des diverses sources géologiques par des méthodes isotopiques et moléculaires étaye cette hypothèse.

Dans ce cas, le verbe *étayer* est utilisé au sens figuré, soit « *renforcer une hypothèse* », et non au sens littéral, soit « *supporter par des étais* ». Pour allouer une fonction argumentative ou rhétorique à certains verbes, il faut donc pouvoir résoudre les problèmes de polysémie.

Le réseau de verbes, accompagné du module de résolution de la polysémie verbale, permet de repérer les verbes utilisés au sens figuré susceptibles alors de référer à une fonction argumentative ou rhétorique.

Lors de la tâche de résumé, il sera également possible d'utiliser le réseau de verbes comme « dictionnaire des synonymes » pour des reformulations ou paraphrases (étape de « lissage ») dans le résumé engendré.

4.2. Autres applications

D'autres applications sont envisageables pour le réseau et les schémas descriptifs des verbes.

Les réseaux sémantiques des noms sont en général établis indépendamment des verbes. Puis, on tente de rattacher aux prédicats verbaux les catégories sémantiques ainsi définies comme valeurs de relations casuelles. Cette manière de procéder semble erronée au regard des restrictions imposées par les verbes.

Considérons par exemple le verbe *verser* auquel on associe généralement un *liquide* comme objet de l'action. Ce verbe admet, avec une signification similaire, des concepts de catégories sémantiques autres que liquide, *poudres* ou *granulés* par exemple. Il est donc essentiel d'envisager la construction des catégories des concepts nominaux relativement aux restrictions de sélection imposées par les verbes de manière à obtenir une cohérence entre les noms (leurs catégories et leurs propriétés) et les valeurs des relations casuelles attendues par les verbes. La difficulté de cette approche vient de ce qu'il n'existe pas toujours de termes dans la langue pour désigner des concepts aussi hétérogènes que *liquide* et *granulés*. Il faudra par conséquent créer des types conceptuels à l'aide de termes d'un métalangage ou d'un pseudo-langage qui regrouperaient les catégories en question.

Le réseau peut être utilisé comme outil de génération automatique de définitions. Il suffit pour cela de lier un descripteur (entre crochets dans l'exemple qui suit) à chaque trait sémantique du verbe à définir, afin de reconstruire une définition correcte du point de vue syntaxique.

Pour le verbe *couper* : diviser [**un**] objet solide [**en**] plusieurs parties [**au moyen d'un**] instrument tranchant...

La structure en arbre autorise aussi des renvois à tous les verbes ayant un apparentement de sens et offre donc la fonction de dictionnaire des synonymes. Associé à la description de chaque verbe, le réseau permet d'accéder à la paraphrase (*i.e.* exprimer de plusieurs manières un même contenu sémantique) : le verbe *décapiter* par exemple, peut être reformulé en *couper la tête* ou encore *trancher la tête*.

Conclusion

En nous fondant sur l'étude des modes d'interprétation de la métaphore verbale, nous avons proposé un formalisme de représentation des connaissances lexicales pertinent pour le traitement de la polysémie des verbes (descriptions maximalisées des verbes et construction d'un réseau hiérarchique). Nous avons montré qu'une hiérarchisation des verbes est d'emblée possible en fonction de la présence de valeurs pour les cas et de leur degré de spécificité. Enfin nous avons fait apparaître quelques heuristiques d'interprétation des sens figurés des verbes et de leurs sous-ordonnés.

Il reste à définir l'ensemble des règles pour le traitement de la polysémie et les conditions de choix entre les différents sens d'un verbe donné.

Nous souhaitons élargir l'étude aux autres catégories grammaticales (noms, adjectifs, adverbes) parce que ces représentations sont à notre sens indispensables à l'analyse sémantique de phrases et de textes : fréquence des phénomènes de paraphrases et de sens figurés dans les différents écrits (scientifiques, littéraires, journalistiques...). La robustesse des systèmes informatiques de traitement des langues (en analyse comme en génération) dépend grandement de leur capacité à intégrer la dimension polysémique du lexique.

Comme l'évoque G. Sabah (1988), la polysémie reste encore un des aspects du langage naturel qui pose le plus de difficultés en informatique linguistique.

Références

- ABEILLÉ, A. (1993) : *Les nouvelles syntaxes. Grammaires d'unification et analyse du français*, Paris, Armand Colin.
- ABEILLÉ, A. et Y. SCHABES (1989) : « Parsing Idioms in lexicalized TAGs », *Proceedings of EACL-89*, Manchester, pp. 1-9.
- Acad (1694) : *Dictionnaire de l'Académie française*, 1^{re} éd., Paris, Coignard, 1694 ; 2^e éd., Paris, Coignard, 1718 ; 3^e éd., Paris, Coignard, 1740 ; 4^e éd., Paris, Brunet, 1762 ; 5^e éd., Paris, Smits, 1798 ; 6^e éd., Paris, Firmin-Didot, 1835 ; 7^e éd., Paris, Firmin-Didot, 1878 ; 8^e éd., Paris, Hachette, 1932-1935.
- AGARWAL, R. (1994) : « (Almost) Automatic Semantic Feature Extraction from Technical Text », *Proceedings of the Human Language Technology Workshop*, Plainsboro, (N. J.), 8-11 mars, pp 378-382
- AGIRRE, E., ALEGRIA, I., ARREGI, X., ARTOLA, X., DIAZ DE ILARRAZA., SARASOLA, K. et M. URKIA (1989) : « Aplicación de la morfología de dos niveles al euskara », *SEPLN*, vol. 8, Barcelona, pp. 87-102.
- AHA, D. W., KIBLER, D. et M. K. ALBERT (1991) : « Instance-Based Learning Algorithms », *Machine Learning*, 6, pp. 37-66.
- AHLSEWEDE, T. & M. EVENS (1988) : « Generating a Relational Lexicon from a Machine-Readable Dictionary », *International Journal of Lexicography*, 1 (3), pp 214-237.
- AHO, A. V. & M. J. CORASICK (1975) : « Efficient String Matching : An Aid to Bibliographic Search », *Comm. of the ACM*, 18 (6), pp 330-340.
- AHO, A. V., KERNIGHAN, B. W. et P. J. WEINBERGER (1988) : *The AWK Programming Language*, AT&T Bell Laboratories, Murray Hill, New Jersey
- ALI, Nabil (1988) : *Al-Lughah Al-'Arabiyya wa Al-Hâsûb*, Dâr Taarîb.
- ALI, Nabil (1994) : *Al-'Arab wa 'Asr Al Ma'lûmât*, Al Koweït, 'Alam Al Ma'rifa, 48.
- ALLEN, M. (1978) : *Morphological Investigations*, unpublished Doctoral dissertation, University of Connecticut, Storrs
- ALLERTON, David J. (1982) : *Valency and the English Verb*, London, Academic Press.

- ALONSO RAMOS, M. (1993) : *Las funciones lexicales y el modelo lexicografico de I. Mel'čuk*, Universidad Nacional de Educacion a Distancias, Madrid, thèse de doctorat.
- ALONSO RAMOS, M. et S. MANTHA (1996) : « Description lexicographique des collocations dans un *Dictionnaire Explicatif et Combinatoire (DEC)* : articles de dictionnaire autonomes ? », *Actes de Lexicomatique et dictionnairiques*, Lyon, 28-30 septembre 1995.
- ALONSO RAMOS, M. et A. TUTIN (1996) : « A Classification and Description of the Lexical Functions for the Treatment of LF Combinations », L. Wanner (Ed.), *Lexical Functions in Lexicography and Natural Language Processing*, Amsterdam, Benjamins.
- ALONSO RAMOS, M., TUTIN, A. et G. LAPALME (1995) : « Lexical Functions of the *Explanatory Combinatorial Dictionary* for the Lexicalization in Text Generation », P. Saint-Dizier et E. Viegas (Eds), Cambridge University Press.
- ALSHAWI, H. (Ed.) (1992) : *The Core Language Engine*, Cambridge (MA), The MIT Press, 322 p.
- ALSHAWI, H., ARNOLD, D.-J., BACKOFEN, R., CARTER, D.-M., LINDOP, J., NETTER, K., PULMAN, S.-G., TSUJII, J. et H. USZKOREIT (1991) : *Eurotra ET6/1 : Rule Formalism and Virtual Machine Design Study (Final Report)*, Luxembourg, CEC.
- AMSLER, R. A. (1980) : *The Structure of the Merriam-Webster Pocket Dictionary*, Ph.D. Thesis, University of Texas at Austin, Austin.
- ANDERSON, Stephen (1977) : « On the Formal Description of Inflection », *Papers from the XIIIth Regional Meeting of the Chicago Linguistic Society*, Chicago, pp. 15-44.
- ANDERSON, Stephen (1985) : « Inflectional Morphology », T. Shopen (Ed.), *Language Typology and Syntactic Description*, Part III, Grammatical Categories and the Lexicon, Cambridge University Press, pp. 150-201
- ANDERSON, Stephen (1985) : « Typological Distinction in Word Formation », T. Shopen (Ed.), *Language Typology and Syntactic Description*, Part III, Grammatical Categories and the Lexicon, Cambridge University Press, pp. 3-56
- ANGLUIN, D. (1980) : « Inductive Inference of Formal Languages from Positive Data », *Information and Control*, 45, pp. 117-135.
- ANTONY-LAY, M.-H., FRANCOPOULO, G. & L. ZAYSSER (1994) : « A Generic Model for Reusable Lexicons : The GENELEX Project », *Linguistics and Literary Computing*, vol. 8.
- ANTWORTH, E. L. (1990) : *PC-KIMMO : A Two-Level Processor for Morphological Analysis*, Dallas (TX), Summer Institute of Linguistics, 273 p.
- APPELT, D. E., HOBBS, J. R., BEAR, J., ISRAEL, D. et M. TYSON (1993) : « FASTUS : A Finite-state Processor for Information Extraction from Real-world Text », *Proceedings of the 13th International Joint Conference on Artificial Intelligence*, Chambéry (France), 28 août-3 septembre, pp 1172-1178.
- ARMSTRONG-WARWICK, S. (1993) : « Acquisition and Exploitation of Textual Resources for NLP », *Proceedings of Knowledge Base and Knowledge Systems*, Tokyo.

- ARNTZ, R. (1993) : « Terminological Equivalence and Translation », Sonneveld, H. et K. Loening (Eds), *Terminology Applications in Interdisciplinary Communication*, Amsterdam et Philadelphia, John Benjamins Publishing Company, pp 5-19.
- ARONOFF, Mark (1974) : *Word-Structure*, unpublished Doctorate dissertation, MIT.
- ARONOFF, M. (1976) : *Word Formation in Generative Grammar*, Linguistic Inquiry, Monograph one, Cambridge (MA), The MIT Press
- ARONOFF, Mark (1978) : « Lexical Representations », *Papers from the Parasession on the Lexicon · XIVth Regional Meeting of the Chicago Linguistic Society*, Chicago, pp. 12-26.
- ASSAL, Allal et al. (1992) : « Sémantique et terminologie . sens et contextes », *Terminologie et Traduction*, n° 2/3, éd. Commission des Communautés européennes, pp. 411-421.
- ATKINS, B. T. & A. DUVAL (1978) : *Robert & Collins Dictionnaire Français-Anglais, Anglais-Français*. Paris, Le Robert/Glasgow, Collins.
- ATKINS, B. et A. ZAMPOLI (Eds) (1994) : *Automating the Lexicon*, Oxford University Press.
- ATKINS, B. T. & A. ZAMPOLLI (Eds) (1994) : *Computational Approaches to the Lexicon*, Oxford University Press.
- ATTALI, A., BOURQUIN, G. et al. (1992) : « Aide au transfert lexical dans une perspective de TAO : expérimentation sur un lexique non terminologique », *Meta*, 37 (4), pp. 770-790.
- AUROUX, S. (juin 1994) : « L'hypothèse de l'histoire et la sous-détermination grammaticale », *Langages*, 114, Paris, Larousse, pp. 25-40.
- AZOUGARH, M. (1992) : *Lexique berbère structures et signification*, thèse de DES, Faculté des Lettres et des Sciences Humaines, Oujda.
- BACHIMONT, B. (1995) : « Ontologie régionale et terminologie : quelques remarques méthodologiques et critiques », *La Banque des Mots*, Numéro spécial du Centre de Terminologie et de Néologie du CNRS, Actes de la Première journée "Terminologie et Intelligence Artificielle", Paris Villetaneuse.
- BAEZA-YATES, R. et G. GONNET (1992) : « A New Approach to Text Searching », *Communication of the ACM*, October 1992, 35 (10), pp. 74-82.
- BANKS, David (1994) : « Clause Organization in the Scientific Journal Article », *Unesco ALSED-LSP Newsletter*, 17 (2), pp. 4-16.
- BARBAUD, Philippe (1992) : « Recycling Words », Christiane Lauefer et Terrell A. Morgan (Eds), *Theoretical Analyses in Romance Linguistics*, coll. « CILT 74 », Philadelphie/Amsterdam, John Benjamins, pp. 197-217.
- BARBAUD, Philippe (1994) : « Conversion syntaxique », *Linguistic Investigations*, XVIII, pp. 1-26
- BARBAUD, Philippe (à paraître) : « La nominalisation d'un participe passé : la suppléance **mettre** mise en composition lexicale », *Revue canadienne de linguistique/Canadian Journal of Linguistics*.

- BARKER, K., COPECK, T., DELISLE, S. et S. SZPAKOWICZ (1993) : *A Case System for Interactive Knowledge Acquisition from Text*, Technical Report TR-93-08, Computer Science Department, University of Ottawa. février, 32 p.
- BARKER, K. et S. SZPAKOWICZ (1995) : « Interactive Semantic Analysis of Clause-Level Relationships », à paraître dans *Proceedings of the 1995 PACLING (Pacific Association for Computational Linguistics) Conference*, University of Queensland, Brisbane (Australie), 19-22 avril
- BASILI, R., PAZIENZA, M. T. et P. VELARDI (1992) : « Computational Lexicons : the Neat Examples and the Odd Exemplars », *Proceedings of the 3rd Conference on Applied Natural Language Processing*, Trento (Italy), 31 mars-3 avril, pp. 96-103
- BASSET, A. (1929) : *La langue berbère. Morphologie. Le verbe – Étude des thèmes*, Paris, Le-roux, LII + 269 p.
- BATEMAN, J. A. (1991) : « The Theoretical Status of Ontologie in Natural Language Processing », *KIT Report 97*, Technical University Berlin.
- BATEMAN, J. A. (1993) : « Ontology Construction and Natural Language », *International Workshop on Formal Ontology*, Padova (Italy), LADSEB-CNR National Research Council, pp. 83-93.
- BAUER, D., SEGOND, F. & A. ZAENEN (1995) : *Enriching an SGML-Tagged Bilingual Dictionary for Machine-Aided Comprehension*, Rank Xerox Research Centre Technical Report, Meylan
- BAUER, Daniel, SEGOND, Frédérique et Annie ZAENEN (1994) : « Enriching an SGML-Tagged Dictionary for Machine-Aided Comprehension », *Technical Report MLTT-011*, Rank Xerox Research Centre, Grenoble
- BAUER, Daniel, SEGOND, Frédérique et Annie ZAENEN (1995) : « LOCOLEX: the Translation Rolls off your Tongue », *Proceedings of the ACH-ALLC Conference*, Santa Barbara, pp 6-8
- BEAUJEAN, A. (1990) : *Le Petit Littré, dictionnaire de la langue française*, coll. « classiques modernes », Paris, La Pochothèque/Le Livre de Poche.
- BÉJOINT, H. (1993) : « La définition en terminologie », P. J. L. Arnaud & Ph. Thoiron (dir.), *Aspects du vocabulaire*, Lyon, Presses universitaires de Lyon, pp 18-25.
- BÉJOINT, H. et Ph. THOIRON (1987) : « Compte rendu de Benson *et al* (1986a) », *Les langues modernes*, 81 (3-4), pp. 152-159.
- BENMRAD, M. (1994) : *Agregats et composition de requêtes dans les hypertextes virtuels*, thèse de doctorat n 1284, Département d'informatique, École Polytechnique Fédérale de Lausanne, Octobre 1994.
- BENMRAD, M., CORAY, G. et C. VANOIRBEEK (1995) : « Designing Virtual Hyper-texts with Aggregates », *Proceedings of IWH'D'95*, Montpellier, France, June 1995
- BENSON, M. (1989) : « The Structure of the Collocational Dictionary », *International Journal of Lexicography*, 2 (1), pp 1-14.
- BENSON, M. (1990) : « Collocations and General-Purpose Dictionaries », *International Journal of Lexicography*, 3 (1), pp. 23-34.

- BENSON, M., BENSON E. et R. ILSON (1986a) : *The BBI Combinatory Dictionary of English. A Guide to Word Combinations*, Amsterdam/Philadelphia, John Benjamins, 286 p.
- BENSON, M., BENSON E. et R. ILSON (1986b) : *The Lexicographic Description of English*, Amsterdam, John Benjamins, 288 p.
- BENVENISTE, E. (1939) : « Nature du signe linguistique ». *Acta Linguistica*, 1, Copenhague, 1939. et *Problèmes de linguistique générale*, 1, Gallimard, 1974.
- BENVENISTE, Émile (1966) : « Formes nouvelles de la composition nominale », *Problèmes de linguistique générale*, Paris, Gallimard, pp. 163-173.
- BENVENISTE, Émile (1974) : *Problèmes de linguistique générale*, Paris, Gallimard.
- BENVENISTE, Émile (1974) : *Problèmes de linguistique générale* 2, coll. « TEL », Paris, Gallimard.
- BERNARD, P. et F. AL (1991) : « Dictionnaire bilingue et ordinateur ». *Wörterbücher, Ein internationales Handbuch zur Lexikographie*, Berlin/New York, Walter de Gruyter, pp 2804-2813.
- BERWICK, R. C. (1986) : « Learning from Positive-Only Examples: The Subset Principle and Three Case Studies », R. S. Michalski, J. G. Carbonell & T. M. Mitchell (Eds), *Machine Learning : An Artificial Intelligence Approach, Vol. II*, Los Altos (CA), Morgan Kaufmann, pp. 625-645.
- BIERWISCH, M. et K. E. HEIDOLPL (Eds) (1979) : *Progress in Linguistics*, The Hague, Mouton.
- BLACHE, P. (1995a) : « Contraintes et Héritage pour l'analyse syntaxique : pour une programmation multi-paradigmes », Actes de TALN'95.
- BLACHE, P. (1995b) : *Introduction à HPSG*, Rapport Technique #PB-9501, 2LC-CNRS.
- BLACK, A., RITCHIE, G., PULMAN, S. et G. RUSSELL (1987) : « Formalisms for Morphographic Description », *EACL-3*.
- BLAMPAIN, D. (1992) : « Traduction et écosystèmes terminologiques », *Terminologie & traduction*, 2 (3), pp. 457-466.
- BLAMPAIN, D. (1993) : « Notions et phraséologie. Une nouvelle alliance ? », Phraséologie. Actes du Séminaire international, *Terminologies nouvelles*, 10, pp. 43-49.
- BLAMPAIN, D., PETRUSSA, P. & M. VAN CAMPENHOUDT (1992) : « À la recherche d'écosystèmes terminologiques », *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.
- BLANCHON, H. (1994a) : « Perspectives of DBMT for Monolingual Authors on the Basis of LIDIA-1, an Implemented Mock-up », *15th International Conference on Computational Linguistics*, COLING-94, Kyoto, August 5-9
- BLANCHON, H. (1994b) : *LIDIA-1 : Une première maquette vers la TA interactive « pour tous »*, thèse de Doctorat, Université Joseph Fourier, Grenoble. (N. P.)

- BOGURAEV, B. (1988) : « A Natural Language Toolkit : Reconciling Theory with Practice », U. Reyle & C. Rohrer (Eds), *Natural Language Parsing and Linguistics*, Dordrecht, D. Reidel, pp 95-130.
- BOGURAEV, B. (1991) : « Building a Lexicon : The Contribution of Computers », *International Journal of Lexicography*, 4 (3), pp. 227-260.
- BOGURAEV, B. (1994) : « Machine-Readable Dictionaries and Computational Linguistics Research », Zampolli, Calzolari & Palmer (Eds), *Current Issues in Computational Linguistics : in Honour of Don Walker*, Series « Linguistica Computazionale », IX-X, pp. 119-154.
- BOGURAEV, B. & T. BRISCOE (1989) : *Computational Lexicography for Natural Language Processing*, London and New York, Longman.
- BOGURAEV, B., BRISCOE, T., CARROLL, J. & A. COPESTAKE (1992) : « Database Models for Computational Lexicography », *EURALEX'90 Proceedings*, Barcelona. Bibliograf. pp. 59-78
- BOGURAEV, B. et J. PUSTEJOVSKY (1990) : « Knowledge Representation and Acquisition from Dictionary », *Coling Tutorial*, August 16-18, 1990, Helsinki, Finland.
- BOISVERT, R. (1989) : *Programme de concordance*, rapport de projet de synthèse en informatique #89 1 06, Département de mathématiques et d'informatique, Université du Québec à Trois-Rivières.
- BOITET, C. & H. BLANCHON (1993) : « Dialogue-Based MT for Monolingual Authors and the LIDIA Project », H. Nomura (Ed.), *Proceedings NLPRS'93 (Natural Language Processing Rim Symposium)*, Fukuoka. 6-7/12/93, Kyushu Institute of Technology, pp. 208-222.
- BORESDON, Bernard et Irène TAMBA (1991) : « Verre à pied, moule à gaufres . préposition et noms composés de sous-classe », *Langue française*, 91, pp. 40-55.
- BOUCHARD, L. H. & L. EMIRKIANIAN (1990) : « Développement d'un analyseur morpho-syntaxique pour le français », *Actes du colloque « Les industries de la langue . Perspectives des années 1990 »*, Montréal, Office de la langue française et Société des traducteurs du Québec, pp. 115-130.
- BOUCHARD, L. & L. EMIRKIANIAN (1994) : « The Organization of the Lexicon in GSF : Structure and Implementation », Kiefer, Kiss & Pajzs (Eds), *Papers in Computational Lexicography – COMPLEX'94*, Research Institute for Linguistics, Hungarian Academy of Sciences, Budapest, pp. 13-22.
- BOUCHARD, L., EMIRKIANIAN, L. & F. GROS D'AILLON (1991) : « Extracting French Morphological and Syntactic Information from a Machine-Readable Dictionary », Kiefer (Ed.), *Computational Lexicography*. [Balatonfüred, Hungary 8-11 September 1990], Research Institute for Linguistics, Hungarian Academy of Sciences, pp. 9-24.
- BOUCHARD, L., EMIRKIANIAN, L. & J.-Y. MORIN (1992) : « Computational Grammar as Knowledge Representation », *Proceedings Sixth International Conference on Systems Research Informatics and Cybernetics (Volume II)*, Baden-Baden. International Institute for Advanced Studies in System Research and Cybernetics, pp. 121-132.
- BOUCHARD, L. H., EMIRKIANIAN, L. & S. RATTÉ (1989) : « L'analyse MLR une application au français », *ICO : Intelligence Artificielle et Sciences Cognitives au Québec*, 1(4), pp. 50-60

- BOUILLON, P. (1995) : « Le lexique génératif : une alternative au traitement de la polysémie. Le cas de *commencer* », communication présentée au cours du Colloque
- BOURIGAULT, D. (1992) : « Surface Grammatical Analysis for the Extraction of Terminological Noun Phrases », *Proceedings of the Fourteenth International Conference on Computational Linguistics*, COLING-92, Nantes, pp 977-981.
- BOURIGAULT, D. (1993a) : « An Endogenous Corpus-based Method for Structural Noun Phrase Disambiguation », *Proceedings of the 6th Conference of the European Chapter of the Association for Computational Linguistics (EACL'93)*, Utrecht.
- BOURIGAULT, D. (1993) : « Analyse syntaxique locale pour le repérage de termes complexes dans un texte », *T.A.L.*, 34 (2), pp 105-118
- BOURIGAULT, D. (1994) : *LEXTER, un Logiciel d'EXtraction de TERminologie. Application à l'acquisition des connaissances à partir de textes*. Thèse de doctorat de l'École des hautes études en sciences sociales, Paris, juin 1994
- BOURIGAULT, D. (1995) : « Lexter, a Terminology Extraction Tool for Knowledge Acquisition from Texts », *Proceedings of the 9th Knowledge Acquisition for KnowledgeBased Systems Workshop (KAW'95)*, Banff.
- BOURIGAULT, D. et P. LÉPINE (1995) : « Utilisation d'un logiciel d'extraction de terminologie (Lexter) pour l'acquisition des connaissances à partir de textes », *Acquisition et ingénierie des connaissances : tendances actuelles*, Eds. N. Aussenac-Gilles, P. Laublet, C. Reynaud, Toulouse, Cépaduès.
- BOUTIN-QUESNEL, R. et al. (1985) : « Vocabulaire systématique de la terminologie », *Cahiers de l'Office de la langue française*, Québec.
- BRESNAN, Joan (1982) : *The Mental Representation of Grammatical Relations*, Cambridge (MA), MIT Press.
- BRILL, E. (1992) : « A Simple Rule-Based Part of Speech Tagger », *Proceedings of the 3rd Conference on Applied Natural Language Processing*, Trento (Italy), 31 mars-3 avril, pp. 152-155.
- BRISCOE, T., DE PAIVA, V. et A. COPESTAKE (Eds) (1993) : *Inheritance, Defaults and the Lexicon*, Cambridge University Press.
- BRISCOE, T. et A. COPESTAKE (1993) : *Default Inheritance in the Lexicon*, Cambridge, CUP.
- BROWN, P., CHURCH, K., GODBY, J., LEWIS, D., REIGHART, R. et F. ZHOU (Eds.) (1993) : *Very Large Corpora : Academic and Industrial Perspectives*, Proceedings of the Workshop Sponsored by the Association for Computational Linguistics, American Chemical Society Chemical Abstracts, Mead Data Central, Inc., OCLC Online Computer Library Center, Inc. June 22, 1993, Ohio State University, Columbus, USA.
- BROWN, P. et al. (1991) : « Word-sense Disambiguation Using Statistical Methods », *Proceedings of ACL'91*.
- BROWN, P. et al. (1993) : « The Mathematics of Statistical Machine Translation : Parameter Estimation », *Computational Linguistics*, 19 (2), pp. 261-311.

- BRUANDET, M. F. (1980) : « A Conceptual Network for Automatic and Dynamic Thesaurus Updating in Information Retrieval System », *Proceedings of COLING-80*, 30 sept.-3 oct. Tokyo.
- BRUNDAGE, Jennifer, KRESSE, Maren, SCHWALL, Ulrike et Angelika STORRER (1992) : « Multiword Lexemes: A Monolingual and Contrastive Typology for NLP and MT IBM Deutschland GmbH », Institut für Wissensbasierte Systeme, Heidelberg, *IWBS Report 232*, September 1992, IBM TR-80.92-029.
- BUVET, P.-A. (1995) : « Lexicalisation et domaines d'emplois », communication présentée au cours du Colloque.
- BYRD, R. (1989) : « Discovering Relationships among Word Senses », *Dictionaries in the Electronic Age – Proceedings of the Fifth Annual Conference of the UW Centre for the New Oxford English Dictionary*, Oxford, pp. 67-79.
- BYRD, R., CALZOLARI, N., CHODOROW, M., KLAVANS, J., NEFF, M. & O. RIZK (1987) : « Tools and Methods for Computational Lexicology », *Computational Linguistics*, 13 (3/4), pp. 219-240.
- CADIOT, Pierre (1991) : « À la hache ou avec la hache ? Représentation mentale, expérience située et donation du référent », *Langue française*, 91, pp. 7-23.
- CALZOLARI, Nicoletta et Remo BINDI (1990) : « Acquisition of lexical information from a large textual Italian corpus », *Proceedings of the Thirteenth International Conference on Computational Linguistics*, Helsinki, Finlande.
- CAP-College of American Pathologists (1993) : *SNOMED International*, Introduction
- CARDIE, C. (1993) : « A Case-Based Approach to Knowledge Acquisition for Domain-Specific Sentence Analysis », *Proceedings of the 11th National Conference on Artificial Intelligence*, Washington (DC), 11-15 juillet, pp. 798-803.
- CARRÉ, R., DEGREMONT, J.-F., GROSS, M. et G. SABAH (1991) : *Langage humain et machine*, Paris, Presses du CNRS.
- CARROL, J. (1993) : *Lexical Database System, User Manual*, ESPRIT BRA 3030 Computer Laboratory, University of Cambridge, United Kingdom.
- CEUSTERS W., DEVILLE G., MOUSEL P., STREITER O. et G. THIENPONT (1994) : *Functional Specification of the ANTHEM Prototype*, ANTHEM Deliverable n D1-1.
- CHAKER, S. (1984) : *Textes en linguistique berbère*, Paris, éditions du CNRS, 291 p.
- CHAKRAVARTHY, A. S. (1995) : « Sense Disambiguation Using Semantic Relations and Adjacency Information », *Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics*, Cambridge (MA), 26-30 juin, pp. 293-295 (student session).
- CHANGEUX, Jean-Pierre (1983) : *L'homme neuronal*, Paris, Fayard.
- CHANOD, J.-P. & P. TAPANAINEN (1995) : « Statistical and Constraint Based Taggers for French », *Proceedings EACL'95*, Dublin, 17 p.
- CHAUCHÉ, Jacques (1984) : « Le Système Sygmart », *Actes de COLING'84*, Stanford.

- CHEN, S. F. (1993) : « Aligning Sentences in Bilingual Corpora Using Lexical Information », *Proceedings of the 13th Annual Meeting of ACL '93*, pp. 9-16.
- CHIBOUT, K. (1994) : *Traitement automatique des tropes*, document interne LIMSI, n° 94-04. Orsay, 66 p.
- CHOMSKY, Noam (1965) : *Aspects of the Theory of Syntax*, Cambridge (MA), The MIT Press.
- CHOMSKY, Noam (1970) : « Remarks on Nominalization », *Readings in English Transformational Grammar*, Rodeock Jacobs & Peter Rosenbaum (Eds), Waltman (MA), Ginn. pp. 184-221
- CHOMSKY, Noam (1981) : *Lectures on Government and Binding*, Dordrecht, Foris Publications.
- CHOUÉKA, Y. (1988) : « Looking for Needles in a Haystack », *Proceedings of the RIAO Conference on User-Oriented Context Based Text and Image Handling*, Cambridge, Ma.
- CHOUÉKA, Y. (1988) : « Looking for Needles in a Haystack or Locating Interesting Collocational Expressions in a Large Textual Database », *Actes de colloque du RIAO 88*. Cambridge, Cambridge University Press, pp. 609-623.
- CHOUÉKA, Y., KLEIN, T. et E. NEUWITZ (1983) : « Automatic Retrieval of Frequent Idiomatic and Collocational Expressions in a Large Corpus », *ALLC Journal*, Grande-Bretagne. 4 (1), pp 34-39.
- CHUKWU, U. (1993) : *Le repérage des termes dans un corpus bilingue anglais/français*. Thèse de doctorat, Université Lumière Lyon II.
- CHUKWU, U. et Ph. THOIRON (1989) : « Reformulation et repérage des termes », *La Banque des mots*, numéro spécial, pp. 23-50.
- CHURCH, K. & P. HANKS (1990) : « Word Association Norms, Mutual Information and Lexicography », *Computational Linguistics*, 16 (1), pp 22-29.
- CHURCH, K. & R. PATIL (1982) : « Coping with Syntactic Ambiguity or How to Put the Block in the Box on the Table », *Computational Linguistics*, 8 (3-4), pp. 139-149.
- CHURCH, K., GALE, W., HANKS, P. et D. HINDLE (1991) : « Using Statistics in Lexical Analysis », *Lexical acquisition : Using On-Line Resources to Build a Lexicon*, U. Zernik (Ed.), Lawrence Erlbaum.
- CHURCH, K., GALE, W., HANKS, P., HINDLE, D. & R. MOON (1994) : « Lexical Substitutability », Atkins & Zampolli (Eds), *Computational Approaches to the Lexicon*, Oxford University Press, pp. 153-177.
- CLAVIER, V. (1995) : *Modélisation de la suffixation pour le traitement automatique du français · Application à la recherche d'informations*, Thèse de doctorat, Grenoble.
- CLAVIER, V. et G. LALLICH-BOIDIN (1994) : « Modélisation linguistique de la suffixation en vue de l'analyse automatique », *T.A.L.*, 35 (2), pp. 129-143.
- CNET (1990) : Communication personnelle.

- COHEN, B. (1986) : *Lexique de cooccurrents. Bourse – Conjoncture économique*, Montréal, Linguattech.
- COHEN, D. (1993) : « Racines », *À la croisée des études libyco-berbères Mélanges offerts à Paulette Galang-Pernet et Lionel Galand. Comptes rendus de G.L.E.C.S.*, supplément 15, Paris, Librairie orientaliste Paul Geuthner, pp. 161-175.
- Collins Cobuild English Dictionary* (1995): London, Harper Collins Publishers Ltd.
- Collins Cobuild English Language Dictionary* (1987) : London and Glasgow, Collins Publishers, xxiv + 1703 p.
- CONDAMINES, A. (1995) : « Terminology . New Needs, New Perspectives », *Terminology*, Vol. 2 (2).
- COPECK, T., DELISLE, S. et S. SZPAKOWICZ (1992) : « Parsing and Case Analysis in TANKA », *Proceedings of the 14th International Conference on Computational Linguistics*, Nantes (France), 23-28 juillet, pp. 1008-1012
- COPESTAKE, A., SANFILIPPO, A. BRISCOE, T. et V. DE PAIVA (1993) : « The Aquilex LKB : An Introduction », T. Briscoe, V. De Paiva and A. Copestake (Eds), *Inheritance, Defaults and the Lexicon*, Cambridge University Press.
- CORBIN, D. (1987) : *Morphologie dérivationnelle et structuration du lexique*, 2 volumes, Tübingen, Niemeyer.
- CORBIN, Danielle (1987) : « Contre une transposition de la théorie X' à la morphologie dérivationnelle », *Acta Linguistica Academiae Scientiarum Hungaricae*, T. 37 (1-4), pp. 73-92.
- CORBIN, Danielle (1992) : « Hypothèse sur les frontières de la composition nominale », *Cahiers de grammaire 17*, Université de Toulouse-le-Mirail, pp. 25-55.
- CORREARD, M.-H. & V. GRUNDY (Eds) (1994) : *The Oxford-Hachette French Dictionary (French-English, English-French)*, Hachette et Oxford University Press.
- COWIE, A. P. (1986) : « Collocational Dictionaries – A Comparative View », Murphy (Ed.), *Fourth Joint Anglo-Soviet Seminar*, London, British Council, pp 61-69
- COWIE, A. P. (Ed.) (1989) : *Oxford Advanced Learner's Dictionary of Current English*, 4th edition, Oxford University Press.
- CROFT, W. (1984) : *The Representation of Adverbs, Adjectives and Events in Logical Form*, Technical Note 344, SRI International.
- CROFT, W. (1993) : « The Semantics of Mental Verbs », Pustejovsky (Ed.), *Semantics and the Lexicon*. Dordrecht, Kluwer.
- CRUSE, D. A. (1986) : *Lexical Semantics*, Cambridge Textbooks in Linguistics, Cambridge, London, New York, etc., Cambridge University Press.
- DACHELET, R. (1990) : « État de l'art de l'informatique documentaire », *Le Document électronique*, Cours INRIA dirigé par C Bornès, 11-15 juin 1990, Châtellailon, pp. 107-132.

- DAGAN, I. *et al.* (1991) : « Two Languages Are More Informative Than One », *Proceedings of ACL'91*.
- DAGAN, I. et I. ALON (1994) : « Word Sense Disambiguation Using a Second Language Monolingual Corpus », *Computational Linguistics*, vol. 20, pp. 563-596.
- DAHLBERG, I. (1981a) : « Conceptual Definitions for INTERCONCEPT », *International Classification*, 5 (3), pp 142-151.
- DAHLBERG, I. (1981b) : « A Referent-Oriented, Analytical Concept Theory for INTERCONCEPT », *International Classification*, Francfort, INDKS Verlag, 8 (1), pp. 16-22.
- DAHLBERG, I. (1982) : « Terminological Definitions Characteristics and Demands », *Problèmes de la définition et de la synonymie en terminologie*, Actes du colloque international de terminologie, Université Laval, Québec, mai 1982, pp 15-28.
- DAHLGREN, K. (1988) : *Naive Semantics for Natural Language Understanding*, Boston, Kluwer, 257 p.
- DAHLGREN, K. (1993) : A Linguistic Ontology, *International Workshop on Formal Ontology*, Padova (Italy), LADSEB-CNR National Research Council, pp 165-174.
- DAHLGREN, K. & J. McDOWELL (1986) : « Using Commonsense Knowledge to Disambiguate Preposition Phrase Modifiers », *AAAI'86*, Philadelphia, pp. 589-593.
- DAILLE, B. (1993) : « Extraction automatique de terminologie monolingue », *Actes du colloque Informatique et langue naturelle*, Nantes, 21 p.
- DAILLE, B. (1994) : « Extraction de noms composés terminologiques du domaine des télécommunications », *5^e Journées ERLA-GLAT (Études et Recherches Lexicales Appliquées)*, Brest, 13 p.
- DAILLE, Béatrice (1994) : « Approche mixte pour l'extraction automatique de terminologie : statistique lexicale et filtres linguistiques », thèse de doctorat, Université Paris 7.
- DAILLE, B. (1995) : « ACABIT : une maquette d'aide à la construction automatique des banques de données terminologiques », *Actes des IV^e Journées scientifiques du réseau LIT de l'AUPELF-UREF*, Lyon, sept. 1995.
- DAILLE, Béatrice, GAUSSIÉ, Éric et Jean-Marc LANGÉ (1994) : « Towards Automatic Extraction of Monolingual and Bilingual Terminology », *COLING-94*, Kyoto, Japon.
- DAILLE, Béatrice, GAUSSIÉ, Éric et Jean-Marc LANGÉ (1995) : « An Evaluation of Statistical Scores for Word Association », *The Tbilisi Symposium on Language, Logic and Computation*, octobre, Tbilisi, Georgia.
- DALLET, J.-M. (1982) : *Dictionnaire kabyle-français*, Paris, SELAF, XL + 1052 p.
- DANLOS, L. et P. SAMVELIAN (1992) : « Translation of the Predicative Element of a Sentence : Category Switching, Aspect and Diathesis », *Proceedings of the Fourth International Conference on Theoretical and Methodological Issues in Machine Translation*, Montréal, pp. 21-34.
- DAOUST, F. (1992) : *SATO . Système d'analyse de texte par ordinateur, Manuel de référence*, Centre ATO, Université du Québec à Montréal.

- DAVID, S. (1990) : « Le progiciel TERMINO : de la nécessité d'une analyse morphosyntaxique pour le dépouillement des textes », *Acte du colloque Les industries de la langue : perspective des années 1990*, 21 au 24 novembre 1990, Montréal, Office de la langue française et Société des traducteurs du Québec, pp. 71-89.
- DAVID, S. et P. PLANTE (1990) : « De la nécessité d'une approche morpho-syntaxique dans l'analyse de texte », *Revue ICO*, 2 (3).
- DE BESSÉ, B. (1991) : « Des fichiers terminologiques aux bases de connaissances ». Clas, A. et Safar, H. (dir.), *L'environnement traductionnel. La station de travail du traducteur de l'an 2001 – Journées scientifiques du Réseau thématique de recherche « Lexicologie, terminologie, traduction »*, Mons (Belgique), 25-27 avril 1991, Presses de l'Université du Québec, pp. 283-300.
- DELHEURE, J. (1984) : *Dictionnaire mozabite-français*, Paris, SELAF.
- DELISLE, S. (1994) : *Text Processing without A-Priori Domain Knowledge : Semi-Automatic Linguistic Analysis for Incremental Knowledge Acquisition*, Ph.D. thesis, Department of Computer Science Ottawa-Carleton Institute for Computer Science, TR-94-02, University of Ottawa, janvier.
- DELISLE, S. (1995) : *The Reattachment Module Design Document*, rapport technique, Université du Québec à Trois-Rivières, Département de mathématiques et informatique, 2 avril, 32 p.
- DELISLE, S., BARKER, K., COPECK, T. et S. SZPAKOWICZ (à paraître) : « Interactive Semantic Analysis of Technical Texts : Case Pattern Acquisition », à paraître dans le numéro de mai 1996 (vol.12, #2) de *Computational Intelligence*.
- DELISLE, S., COPECK, T., SZPAKOWICZ, S. et K. BARKER (1993) : « Pattern Matching for Case Analysis : A Computational Definition of Closeness », O. Abou-Rabia, C. K. Chang and W. W. Koczkodaj (Eds), *Proceedings of ICCI-93*, Sudbury (Canada), 27-29 mai, pp. 310-315.
- DELISLE, S. et S. SZPAKOWICZ (1991) : « A Broad-Coverage Parser for Knowledge Acquisition from Technical Texts », *Proceedings of the 5th International Conference on Symbolic and Logical Computing*, Madison (S.D.), avril, pp. 169-183.
- DELISLE, S. et S. SZPAKOWICZ (1995) : « Realistic Parsing : Practical Solutions of Difficult Problems », à paraître dans *Proceedings of the 1995 PACLING (Pacific Association for Computational Linguistics) Conference*, University of Queensland, Brisbane (Australie), 19-22 avril, 11 p.
- DELL, F. (1970) : *Les règles phonologiques tardives de la morphologie dérivationnelle du français*, Thèse de doctorat, Cambridge (MA), The MIT Press.
- DELPUI, M. (1993) : *Intégration d'une base de connaissances lexicales pour un analyseur syntaxique*, Rapport de stage, ESSI, Sophia-Antipolis.
- DESCLÉS, J.-P. (1987) : « Réseaux sémantiques : La nature logique et linguistique des relateurs », *Langages*, n° 87, pp. 55-78.
- DESCLÉS, J.-P. (1990) : *Langages applicatifs, langues naturelles et cognition*, Paris, Hermès.
- DESCLÉS, J. P. (1993) : *Représentation des connaissances : archétypes cognitifs, chaînes conceptuelles et schémas grammaticaux*, Centre d'analyse et de mathématiques sociales, CNRS EHESS-Sorbonne.

- DESCLÉS, J.-P., H. ABAAB, J. DICHY, D. E. KOULOUGHLI, M. S. ZIADAH (1983) : *Conception d'un synthétiseur et d'un analyseur morphologiques de l'arabe, en vue d'une utilisation en Enseignement assisté par Ordinateur*, Rapport rédigé sous la direction de J.-P. Desclés.
- DESCLÉS, J.-P. et C. JOUIS (1993) : « L'exploration contextuelle : une méthode linguistique et informatique pour l'analyse automatique de textes », *Actes du colloque Informatique et Langue Naturelle, ILN'93*, Nantes, 2 & 3 déc. 1993, pp. 339-350
- DEVILLE, G. (1989) : *Modelization of Task-Oriented Utterances in a Man-Machine Dialogue System*, Ph.D. Thesis, Universitaire Instelling Antwerpen.
- DEVILLE, G. et E. HERBIGNIAUX (1994) : *Methodological Principles for the Elaboration of Multilingual Corpora of Medical Diagnostic Expressions*, ANTHEM Deliverable n. D2-2 – Part I.
- DICHY, J. et M. O. HASSOUN (dir) (1989) : *Simulation de modèles linguistiques et Enseignement Assisté par Ordinateur de l'arabe - Travaux SAMIA I*, Paris, Conseil international de la langue française.
- Dictionnaire du français contemporain* (DFC) (1971) : Paris, Librairie Larousse.
- DIK, S. (1989) : *A Theory of Functional Grammar*, Dordrecht, Foris.
- DIXON, R. M. W. (1991) : *A New Approach to English Grammar, On Semantic Principles*, Oxford University Press.
- DOSTIE, G., MEL'ČUK I. A. et A. POLGUÈRE (1992) : « Méthodologie d'élaboration des entrées lexicales du Dictionnaire Explicatif et Combinatoire (REPROCHER, REPROCHE et IRRÉPROCHABLE) », *International Journal of Lexicography*, 5 (3), pp. 165-198.
- DUBOIS, Jean et al. (1994) : *Dictionnaire de linguistique et des sciences du langage*, Paris, Larousse, 514 p.
- DUNNING, Ted (1993) : « Accurate Methods for the Statistics of Surprise and Coincidence », *Computational Linguistics*, 19 (1).
- EDR (1993) : *EDR Electronic Dictionary Technical Guide*, Japan Electronic Dictionary Research Institute Ltd. Project report n° TR-042, August 16. 144 p.
- EJERHED, E. I. (1988) : « Finding Clauses in Unrestricted Text by Finitary and Stochastic Methods », *Second Conference on Applied Natural Language Processing*, pp. 219-227.
- EL-BÈZE, Marc (1993) : « Les modèles de langage probabilistes · quelques domaines d'applications », *Habilitation à diriger des recherches*, LIPN, Université Paris-Nord.
- EMIRKANIAN, L. & L. H. BOUCHARD (1989) : « La correction des erreurs d'orthographe d'usage dans un analyseur morphosyntaxique du français », *Langue française*, 83, pp. 106-122.
- EMIRKANIAN, L. & L. H. BOUCHARD (1992) : *Approche computationnelle aux phénomènes morphologique et syntaxique du français*, Rapport de recherche, Université du Québec à Montréal.

- ENGELKE, Sabine (1994) : *Eigenschaften von Phraseolexemen : Eine Untersuchung zur syntaktischen Variabilität und internen Modifizierbarkeit von somatischen verbalen Phraseolexemen*, Magisterarbeit, Universität Tübingen, April 1994
- ENGUEHARD, C. (1992) : *ANA, Apprentissage Naturel Automatique d'un Réseau Sémantique*, thèse de Doctorat en Contrôle des Systèmes, Université de Technologie de Compiègne.
- ENGUEHARD, C. (1993) : « Acquisition automatique de terminologie à partir de gros corpus », *Actes du colloque Informatique & Langue Naturelle, ILN'93*, Nantes, 2 & 3 déc 1993, pp. 373-384.
- ERNST, T. B. (1984) : *Towards an Integrated Theory of Adverb Position in English*, Indiana Linguistics Club, Indiana.
- FARWELL, D., GUTHRIE, L. & Y. WILKS (1993) : « Automatically Creating Lexical Entries for ULTRA, a Multilingual MT System », *Machine Translation*, 8, pp. 127-145.
- FASS, D. (1991) : « MET* : A Method for Discriminating Metonymy from Metaphor by Computer », *Computational Linguistics*, 17 (1), pp. 49-88.
- FELBER, H. (1987) : *Manuel de terminologie*, Paris, UNESCO.
- FELBER, H. (1994) : « Terminology Research : Its Relation to the Theory of Science », *ALFA*, vol. 7/8, pp. 163-172.
- FELLBAUM, C. (1990) : « English Verbs as a Semantic Net », *International Journal of Lexicography*, 3 (4), pp. 278-301.
- FILLMORE, C. (1968) : « The Case for Case », E. Bach and R.T. Harms (Eds), *Universals in Linguistic Theory*, New York, Holt, Rinehart and Winston, pp. 1-88.
- FLEISCHER, Wolfgang (1982) : *Phraseologie der deutschen Gegenwartssprache*, Leipzig, VEB Bibliographisches Institut.
- FONTENELLE, T. (1992) : « Collocation Acquisition from a Corpus or from a Dictionary : a Comparison », Tommola, Varantola, Salmi-Tolonen & Schopp (Eds), *EURALEX'92 Proceedings I-II, Fifth EURALEX International Congress*, Studia Translatologica, Ser. A, Vol. 1, University of Tampere, pp. 220-228.
- FONTENELLE, T. (1994) : « Towards the Construction of a Collocational Database for Translation Students », *Meta*, Presses de l'Université de Montréal, 39 (1), pp. 47-56.
- FONTENELLE, T. (1995) : *Turning a Bilingual Dictionary into a Lexical-semantic Database*, thèse de Doctorat, Université de Liège, ms
- FOSTER, Georges (1991) : « Statistical Lexical Desambiguation », Master's thesis, School of Computer Science, McGill University.
- FOUCAULD, Ch. de. (1951) : *Dictionnaire touareg-français (dialecte de l'Ahaggar)*, Paris, Imprimerie Nationale.
- FOUCAULT, Michel (1969) : *L'archéologie du savoir*, Gallimard, 275 p.
- FOUQUERÉ, C. (1994) : Communication personnelle.

- FRADIN, B. (1993a) : *Organisation de l'information lexicale et interface morphologie/syntaxe dans le domaine verbal*. Thèse de doctorat d'État, Université Paris 8.
- FRADIN, B. (1993b) : « La théorie morphologique face à ces choix », *Cahiers de Lexicologie*, 63, pp. 5-42.
- FRADIN, B. (1994) : « L'approche à deux niveaux en morphologie computationnelle et les développements récents de la morphologie », B. Fradin (Ed.), *La morphologie computationnelle, TAL (Traitement automatique des langues)*, 34 (2), Paris, Association pour le Traitement Automatique des Langues, pp. 9-48.
- FRAMIS, F. R. (1994) : « An Experiment on Learning Appropriate Selectional Restrictions from a Parsed Corpus », *Proceedings of the 15th International Conference on Computational Linguistics*, Kyoto (Japon), 5-9 août, pp. 769-774.
- FRAWLEY, W. (1988) : « New Forms of Specialized Dictionaries », *International Journal of Lexicography*, 1 (3), pp. 189-213.
- FUCHS, C. (1988) : « Représentation linguistique de la polysémie grammaticale », *T.A. Informations*, Revue internationale du traitement automatique du langage, bulletin semestriel de l'ATALA, 29 (1-2), pp. 7-20.
- FUNG, P. (1995) : « A Pattern Matching Method for Finding Noun and Proper Noun Translations from Noisy Parallel Corpora », *Proceedings of ACL'95*.
- GALAND, L. (1974) : « Signe arbitraire et signe motivé en berbère », *Actes du premier congrès international de linguistique sémitique et chamito-sémitiques*, Paris, The Hague, Mouton, pp. 90-101.
- GALAND, L. (1979a) : « Berbère et "traits sémitiques communs" », *Comptes rendus du G.L.E.C.S. tomes XVII-XXIII années 1973-1979*, Paris, Librairie orientaliste Paul Geuthner, pp. 463-478.
- GALAND, L. (1979b) : « La langue berbère existe-t-elle ? », *Mélanges linguistiques offerts à Maxime Rodinson. Comptes rendus du G.L.E.C.S., supplément 12*, Paris, Librairie orientaliste Paul Geuthner, pp. 175-184.
- GALAND, L. (1984) : « Le comportement des schèmes et des racines dans l'évolution de la langue : exemples touaregs », *Current Progress in Afro-Asiatic Linguistics. Third International Hamito-Semitic Congress*. Londres, Benjamins Publishing Company, pp. 304-315.
- GALAND, L. (1988) : « Le berbère », *Les langues dans le monde ancien et moderne, troisième partie : les langues chamito-sémitiques*, Paris, éditions du CNRS, pp. 207-242.
- GALE, William A. et Kenneth W. CHURCH (1991) : « Concordances for parallel texts », *Proceedings of the Seventh Annual Conference of the UW Centre for the New OED and Text Research, Using Corpora*. Oxford (U.K.), pp. 40-62.
- GAMBIER, Yves et François GAUDIN (dir.) (1993) : « numéro spécial Socioterminologie », *Le langage et l'homme*, XXVIII (4), DeBoeck Université.
- GARCIA, D. et A. JACKIEWICZ (1995) : « Aide à l'acquisition des connaissances causales par exploration de textes », *Actes des 6^{es} Journées Acquisition, Validation, (JAVA 95)*, Grenoble, pp. 147-158.

- GARDIN, Bernard *et al.* (Eds) (1994) : *Aspects terminologiques des pratiques langagières au travail*, cahier n° 7, réseau Langage et travail, 74 p.
- GAUDIN, François (1992) : « Terminologie et démocratisation du savoir : à propos de dictionnaires scientifiques », *Le langage et l'homme*, XXVII (2-3), Bruxelles, Institut Libre Marie Haps, pp. 123-129.
- GAUDIN, François (1993) : « Socioterminologie : propos et propositions épistémologiques », Y. Gambier et F. Gaudin (dir.), « numéro spécial Socioterminologie », *Le langage et l'homme*, XXVIII (4), DeBoeck Université, pp. 247-257.
- GAUDIN, François (dir.) (1995) : « numéro spécial : Usages sociaux des termes . théories et terrains », *Meta*, 40 (2), juin 1995, Montréal, Presses de l'Université de Montréal, pp. 193-329.
- GAUDIN, François et Allal ASSAL (dir) (1991) : « Terminologie et sociolinguistique », *Cahiers de linguistique sociale*, n° 18, éd. URA CNRS 1164/Université de Rouen, 213 p
- GAZDAR, G., KLEIN, E., PULLUM, G. & I. SAG (1985) : *Generalized Phrase Structure Grammar*. Cambridge (MA), Harvard University Press, 276 p.
- GIREIL (1994) : *La sous-catégorisation et la cliticisation*, Rapport de recherche, Université du Québec à Montréal
- GOETSCHALCKX, J. (1992) : « Terminologie et phraséologie », *Terminologie & traduction*, 2 (3), pp 477-484.
- GOMEZ, F., HULL, R. et C. SEGAMI (1994) : « Acquiring Knowledge from Encyclopedic Texts », *Proceedings of the 4th Conference on Applied Natural Language Processing*, Stuttgart (Germany), 13-15 octobre, pp. 84-90
- GOUADEC, D. (1994) : « Traduction et informatique · les implications pour la formation ». *Langages*, n° 116. Paris, Larousse. pp. 59-74.
- Grand Robert (1985) : *Le grand Robert de la langue française. Dictionnaire alphabétique et analogique de la langue française*, de P. Robert, 2^e édition entièrement revue et enrichie par A. Rey, 9 vol., Paris, Le Robert
- GREFENSTETTE, G. (1994a) : « Corpus-Derived First, Second and Third-Order Word Affinities », *EURALEX'94 Proceedings*, Vrije Universiteit Amsterdam, pp. 279-290. ◦
- GREFENSTETTE, G. (1994b) : *Explorations in Automatic Thesaurus Discovery*, Boston, Kluwer Academic Press.
- GRISHMAN, R. (1994) : « Whither Written Language Evaluation? », *Proceedings of the Human Language Technology Workshop*, Plainsboro (N. J.), 8-11 mars, pp. 120-125.
- GRISHMAN, R. et J. STERLING (1994) : « Generalizing Automatically Generated Selectional Patterns », *Proceedings of the 15th International Conference on Computational Linguistics*, Kyoto (Japon), 5-9 août, pp. 742-747.
- GRISHMAN, R., MACLEOD, C. et A. MEYERS (1994a) : « Complex Syntax : Building a Computational Lexicon », *Proceedings of the 15th International Conference on Computational Linguistics*, Kyoto (Japon), 5-9 août, pp. 268-272.

- GROSS, G. (1990) : « Définition des noms composés dans un lexique-grammaire », *Langue française*, n° 87, Paris, Larousse.
- GROSS, G. (1990) : « Les mots composés », *Modèles Linguistiques*, 12, pp. 47-63.
- GROSS, G. (1992) : « Forme d'un dictionnaire électronique », *L'environnement traductionnel. La station de travail du traducteur de l'an 2001*, Actes du colloque de Mons, Sillery (Québec), Presses de l'Université du Québec et AUPELF-UREF, pp. 255-271.
- GROSS, G. (1994) : « Classes d'objets et description des verbes », *Langages*, n° 115, septembre 1994, Paris, Larousse, pp. 15-30
- GROSS, G. (1995) : « Une sémantique nouvelle pour la traduction automatique Les classes d'objets », *La tribune des industries de la langue et de l'information électronique*, n° 17-18-19, Paris.
- GROSS, Gaston, CHAURAND, Jacques, VIVÈS, Robert, MATHIEU-COLAS, Michel et Pierre BILLY (1986) : « Typologie des noms composés », *Rapport technique A.T.P.-Nouvelles Recherches sur le Langage*, Université Paris 13.
- GROSS, Maurice (1975) : *Méthode en syntaxe*, Hermann
- GROSS, M. (1986) : « Lexicon grammar. The representation of compound words », *COLING-86*, pp 1-6.
- GROSS, M. (1988) : « Sur les phrases figées complexes du français », *Langue française*, n° 77, Paris, Larousse.
- GROSS, M. (1990) : « Le programme d'extension des lexiques électroniques », *Langue française*, n° 87, Paris, Larousse.
- GROSS, M. (1990) : « Sur la notion harissienne de transformation et son application au français », *Langages*, 99, Paris, Larousse, pp 39-56.
- GROSS, Maurice (1990) : *Grammaire transformationnelle du français Syntaxe de l'adverbe*, Paris, ASSTRIL.
- GUERARD, F. (Ed.) (1989) : *Le dictionnaire de notre temps*, Paris, Hachette, 1714 + 48 p.
- GUILBERT, L. (1965) : *La formation du vocabulaire de l'aviation*, Paris, Larousse.
- GUILBERT, Louis (1975) : *La créativité lexicale*, Paris, Larousse, 285 p
- GUILLET, A. (1990) : *Communication personnelle*.
- GUIRAUD, G. (1900) : *Vocabulaires des dialectes Sango, Bakongo et A'zande*, Paris, Challamel, 58 p.
- HABAILI, Hussein (1976) : *Contraintes de structure morphématique en arabe*, mémoire de Maîtrise ès art en linguistique, Montréal, Université de Montréal.
- HABAILI, Hussein (1990) : *Phonologie et morphologie flexionnelle et dérivationnelle de l'arabe* thèse de Doctorat d'État, Paris, Université de la Sorbonne-Nouvelle, Paris III.

- HABERT, Benoît et Christian JACQUEMIN (1993) : « Présentation », *Traitement automatique des langues*, 34 (2).
- HABERT, B. et C. JACQUEMIN (1993) : « Noms composés, termes, dénominations complexes : problématiques linguistiques et traitements automatiques », *TAL*, 34 (2), ATALA, pp. 5-42.
- Hachette (1993) : *Dictionnaire de notre temps*, Paris, Hachette.
- HALLE, Moris (1973) : « Prolegomena to a Theory of Word Formation », *Linguistic Inquiry*, 4 (1), pp. 3-16.
- HALLIDAY, Michael A. K. (1988) : « On the Language of Physical Science », *Registers of Written English : Situational Features and Linguistic Features*, M. Ghadassy (Ed.), London, Pinter.
- HANSE, J. et D. BLAMPAIN (1994) : *Nouveau dictionnaire des difficultés du français moderne*, Louvain-la-Neuve, De Boeck/Duculot. 983 p
- HarperCollins German-English Dictionary* (1991) : P. Terell, V. Schnorr, W. V. A. Morris, R. Breitsprecher (Eds), 2nd Edition, Glasgow, HarperCollins Publishers.
- HARRIS, Mary Dee (1985) : *Introduction to Natural Language Processing*, Reston, Reston Publishing Company, 312 p.
- HAUSMANN, F. J. (1979) : « Un dictionnaire des collocations est-il possible ? », *Travaux de linguistique et de littérature*, 17 (1), pp. 187-195.
- HAUSMANN, F. J. (1985) : « Kollokationen im Deutschen Wörterbuch. Ein Beitrag zur Theorie des Lexikographischen Beispiels », Bergenholtz & Mugdan (Eds), *Lexikographie und Grammatik*, Tübingen, Niemeyer, pp. 118-129.
- HAUSMANN, F. J. (1989) : « Le dictionnaire des collocations », Hausmann, Reichmann, Wiegand & Zgusta (Eds), *Wörterbücher, Dictionaries, Dictionnaires : An International Encyclopedia of Lexicography*, Berlin and New York, Walter de Gruyter, pp. 1010-1019.
- HEID, U. (1991) : *EUROTRA-7. Feasibility and Project Definition Study on the Reusability of Lexical and Terminological Resources in Computerized Applications. Intermediate report* (non publié).
- HEID, U. (1992) : « Décrire les collocations Deux approches lexicographiques et leur application dans un outil informatisé », *Terminologie et traduction*, 2/3, pp 523-548
- HEID, U. (1992a) : « Décrire les collocations : deux approches lexicographiques et leur application dans un outil informatisé », *Terminologie et Traduction*, Commission des Communautés européennes, pp. 523-548.
- HEID, U. (1992b) : « Notes on the Use of Lexical Functions for the Description of Collocations in an NLP Lexicon », K. Haenelt & L. Wanner (Eds), *International Workshop on the Meaning-Text Theory*, Darmstadt, Schloss Birlinghoven, G.M.D., pp. 217-229.
- HEID, U. (1994) : « On Ways Words Work Together - Topics in Lexical Combinatorics », *EU-RALEX'94 Proceedings*, Vrije Universiteit Amsterdam, pp. 226-257

- HEID, U. et G. FREIBOTT (1991) : « Collocations dans une base de données terminologique et lexicale », *Meta*, 36 (1), pp. 77-91
- HEYLEN, A., MAXWELL, A. K. et S. WARWICK-ARMSTRONG (1989) : « Collocations, Dictionaries and Machine Translation », *Proceedings of the AAAI Symposium on Machine Translation and the Lexicon*, Stanford (CA).
- HEYLEN, D. et K. MAXWELL (1994) : « Lexical Functions and the Translation of Collocations », *Proceedings of EURALEX '94*, Amsterdam, pp. 298-305.
- HOCKEY, S. (1988) : « Creating and Using Large Text Databases for Scholarly Research in the Humanities Some Practical Issues », Gignoni et C. Peters (Eds), *Computational Lexicology and Lexicography*, Pisa, L. Giardini Editori e Stampatori n° 1, coll. « Linguistica computazionale », vol. VI
- HUDSON, R. (1988) : « The Linguistic Foundation for Lexical Research and Dictionary Design », *International Journal of Lexicography*, 1 (4), pp. 287-312
- ICHIKAWA, S. (1990) : *New Japanese-English Dictionary*. Kenkyuusha
- IDE, N., LE MAITRE, J. & J. VERONIS (1994) : « Outline of a Model for Lexical Databases », Zampolli, Calzolari & Palmer (Eds), *Current Issues in Computational Linguistics : in Honour of Don Walker*, Series « Linguistica Computazionale », IX-X, pp. 283-320.
- IDE, N., VERONIS, J., WARWICK-ARMSTRONG, S. et N. CALZOLARI (1992) : « Principles for Encoding Machine-readable Dictionaries », *EURALEX'92 Proceedings*, Tampere (Finlande), pp. 239-246
- IORDANSKAJA, L., KIM, M. et A. POLGUÈRE (1996) : « Some Procedural Problems in the Implementation of Lexical Functions for Text Generation », L. Wanner (Ed.), *Lexical Functions in Lexicography and Natural Language Processing*. Amsterdam, John Benjamins.
- ISO 1087 (1990) : *Terminology - Vocabulary = Terminologie - Vocabulaire*. Genève, Organisation internationale de normalisation.
- ISO 1951 (1973) : *Symboles lexicographiques particulièrement pour l'emploi dans les vocabulaires systématiques à définitions*. Genève, Organisation internationale de normalisation.
- ISO 704 (1987) : *Principes et méthodes de la terminologie*. Genève, Organisation internationale de normalisation.
- ISO R 1087 (1969) : *Vocabulaire de la terminologie*. Genève, Organisation internationale de normalisation
- JACKENDOFF, Ray (1975) : « Morphological and Semantic Regularities in the Lexicon », *Language*, vol. 31, pp. 639-671.
- JACQUEMIN, Christian (1991) : *Transformations des noms composés*, thèse de doctorat en informatique fondamentale, Université Paris 7.
- JACQUEMIN, Christian (1994) : « FASTR: A unification-based front-end to automatic indexing », *Proceedings, Intelligent Multimedia Information Retrieval Systems and Management (RIAO'94)*. New York.

- JANSEN, J. & T. FONTENELLE (1994) : *Short Description of an Implementation of the Robert & Collins English-French Dictionary under Database Format*, Deliverable D-2a of the DECIDE Project, Liège.
- JOUIS, C. (1993) : *Contributions à la conceptualisation et à la modélisation des connaissances à partir d'une analyse linguistique de textes. Réalisation d'un prototype : le système SEEK*, Thèse de doctorat de l'École des hautes études en sciences sociales, Paris, mars 1993.
- JOUIS, C. (1994) : « Contextual Approach: SEEK, a Linguistic and Computational Tool for Use in Knowledge Acquisition », *Proceeding of the First European Conference "Cognitive Science in Industry"*, 28th-30th September 1994, Luxembourg, pp. 259-274.
- JOUIS, C. (1995) : « SEEK, un logiciel d'acquisition des connaissances utilisant un savoir linguistique sans employer de connaissances sur le monde externe », *Actes des 6^{es} Journées Acquisition, Validation, (JAVA 95)*, Grenoble, Avril 1995, pp. 159-172.
- JOUIS, C., COMPAGNON, F., GAUDINAT, B., ROUSSEAU, J.-M. et C. TORA (1991) : « Metodac : Une méthodologie pour l'acquisition et la modélisation des connaissances », *Actes du 8^e congrès RFA*, AFCET, Villeurbanne, nov. 1991, vol. 1, pp. 35-44.
- JOUIS, C. et W. MUSTAFA-ELHADI (1995) : « Conceptual Modeling of Database Schema Using Linguistic Knowledge. Application to Terminological Databases », *First Workshop on Application of Natural Language to Databases (NLDB'95)*, Versailles, Juin 1995, pp. 103-118.
- JUDGE, Anne et Solange LAMOTHE (1995) : *Stylistic Developments in Literary and Non-literary French Prose*, coll. « Studies in French Literature », Vol 19., Lampeter, The Edwin Mellen Press Ltd.
- KARP, D., SCHABES, Y., ZAIDEL, M. et D. EGEDI (1992) : « A Freely Available Wide Coverage Morphological Analyzer for English », *Proceedings of the 14th International Conference on Computational Linguistics*, Nantes (France), 23-28 juillet, pp. 950-955.
- KARTTUNEN, L. (1993) : *Finite-State Lexicon Compiler*, Research Report, XEROX Palo Alto Research Center.
- KARTTUNEN, Lauri et Kenneth BEESLEY (1992) : « Two Level Rules Compiler », *Technical Report ISTL-92-2*, Xerox Palo Alto Research Center.
- KARTTUNEN, Lauri et Todd YAMPOL (1993) : « Interactive Finite-State Calculus », *Technical Report ISTL-NLIT-1993-04-01*, Xerox Palo Alto Research Center
- KAVI, M., (en préparation) : *Ontology Development : Ideology and Methodology*, Computing Research Laboratory, New Mexico State University.
- KAY, M. et M. ROSCHEISENJ (1993) : « Text Translation Alignment », *Computational Linguistics*, 19 (1), pp. 121-142.
- KAYE, Jonathan Derek (1974) : « Morpheme Structure Conditions Live », *Recherches Linguistiques Montréal*, n° 3, pp. 55-62.
- KAYSER, D. (1987) : « Une sémantique qui n'a pas de sens », *Langages*, n° 87, Paris, Larousse, pp. 33-45

- KAYSER, Daniel et Pierre LERAT (1990) : « La notion de définition dans les systèmes de traitement du langage naturel », *La définition*, coll. « Langue et Langage », Paris, Larousse, pp. 113-124.
- KEMBLE, I. R. (1991) : « Lexicography », *Computers as a Tool in Language Teaching*, Brierly, W and I. R. Kemble (Eds), Chichester, Ellis Horwood.
- KEY SUN CHOI YOUNG, S. HAN (1992) : « Syntactic Analysis Based Automatic Indexing for Korean Texts », *International Conference on Terminology, Standardization and Technology Transfer*, Pekin, Science Press.
- KIEFER, F. (1973) : « Morphology in Generative Grammar », M. Gross *et al.* (Eds), *The Formal Analysis of Natural Languages*, The Hague.
- KIPARSKY, P. et C. KIPARSKY (1979) : « Fact », Bierwisch et Heidolph (Eds), *Progress in Linguistics*, The Hague, Mouton.
- KITTREDGE, R. (1983) : *Sublanguage – Specific Computer Aids to Translation – a survey of the most promising areas*, Contract n° 2-5273, Université de Montréal et Bureau des traductions, mars 1983.
- KLEIBER, G. et I. TAMBA (1990) : « L'hyponymie revisitée : inclusion et hiérarchie », *Languages*, n° 98, pp 7-32
- KNOWLES, Frank E. (1993) : « Review of *English Adverbial Collocations* », (voir Kozłowska), *International Journal of Lexicography*, 6 (4), pp 300-302.
- KOCOUREK, R. (1991) : *La langue française de la technique et de la science : vers une linguistique de la langue savante*, 2^e édition, Wiesbaden, Oscar Brandsetter Verlag, 259 p.
- KOINE, Y. (1990) : *New English-Japanese Dictionary*, Kenkyuusha.
- KOSKENNIEMI, K. (1983) : *Two-Level Morphology : A General Computational Model for Word-Form Recognition and Production*, Research Report, University of Helsinki, University of Helsinki, Department of General Linguistics. Publications 11.
- KOZŁOWSKA, Christian D. (1991) : *English Adverbial Collocations*, PWN, Varsovie.
- KRIEGER, H.-U. (1994) : « Derivation without lexical rules », C. Rupp *et al.* (Eds), *Constraints, Language and Computation*, Academic Press, pp. 277-313
- KRIEGER, H.-U. et J. NERBONNE (1993) : « Feature-Based Inheritance Networks for Computational Lexicons », *Research Report 31*, Saarbrücken, Deutsches Forschungszentrum für Künstliche Intelligenz.
- KUPIEC, J. (1992) : « Robust Part-of-Speech Tagging Using a Hidden Markov Model », *Computer Speech and Language*, 6, pp 225-242.
- LADL (1993) : *Outils de traitement linguistique, applications à l'analyse documentaire*, Rapport Technique LADL #43.
- LAFON, Pierre (1984) : *Dépouillements et statistiques en lexicométrie*, Genève, Slatkine et Champion.

- LAKOFF, G. (1987) : *Women, Fire and Dangerous Things*, Chicago.
- LALLICH-BOIDIN, G., HENNERON, G. et R. PALERMITI (1990) : *Analyse du français. Achèvement et implantation de l'analyseur morpho-syntaxique*, Grenoble, Cahiers du CRISS, n° 16.
- LAUER, M. (1994) : « Conceptual Association for Compound Noun Analysis », *Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics, Student Session*, juin, Las Cruces, 6 p.
- LAURISTON, A. (1994) : « Automatic Recognition of Complex Terms : Problems and the TERMINO Solution », *Terminology : International Journal of Theoretical and Applied Issues in Specialized Communication*, 1 (1). John Benjamins Publishing Company, pp 147-170.
- LE NY, J.-F. (1989) : *Science cognitive et compréhension du langage*, Paris, PUF.
- LEBART, L. et A. SALEM (1988) : *Analyse statistique des données textuelles . questions ouvertes et lexicométrie*, Paris, Dunod, 210 p.
- LECLÈRE, C. (1990) : « Organisation du lexique-grammaire des verbes du français », *Langage*, 87, pp. 112-122.
- LEHRER, A. (1990) : « Polysemy. Conventionality and the Structure of the Lexicon », *Cognitive Linguistics*, 1-2, pp. 207-246.
- LELUBRE, X. (1992) : *La terminologie arabe contemporaine de l'optique : faits - théories - évaluation*, thèse de nouveau doctorat, Université Lyon-2, 546 p.
- LERAT, P. (1988) : « Terminologie et sémantique descriptive », *La banque des mots*, numéro spécial, pp. 11-30.
- LERAT, P. (1990) : « L'hyperonymie dans la structuration des terminologies », *Langages*, n° 98, pp. 79-86.
- LERAT, P. (1990) : « Sélection et analyse de termes nouveaux dans une base de données prédictionnaires », *Cahiers de lexicologie*, n° 56-57, Paris. Didier Érudition, pp 255-260.
- LERAT, P. (1995) : *Les langues spécialisées*, coll. « Linguistique nouvelle », Paris, PUF, 206 p.
- LEROUX, D., MINEL, J.-L. et J. BERRI (1994) : « Seraphin Project (Expert System for Automatic Marking of Important Sentences in a Text) the Industrial Approach », *Proceeding of the First European Conference "Cognitive Science in Industry"*, 28-30 Sept. 1994, Luxembourg.
- LEROY-TURCAN, I. (1994a) : « L'informatisation du *Dictionnaire étymologique ou Origines de la langue française* de Gilles Ménage (1694) », I. Lancashire et T. R. Wooldridge (Éd.), *Early Dictionary Databases*, University of Toronto, Centre for Computing in the Humanities, pp. 131-142. = T. R. Wooldridge (Éd.), *Informatique et dictionnaires anciens*, Paris, Didier Érudition, 1995, pp. 131-142.
- LEROY-TURCAN, I. (1994b) : « Intérêt d'une base informatisée pour le *Dictionnaire étymologique ou Origines de la langue française* de Ménage (DEOLF 1694) ; les modalités de mise en œuvre », *Actes du Séminaire sur Dictionnaires historiques et dictionnaires anciens : problèmes de méthode et d'édition*, CNRS-HESO, Ivry-sur-Seine, déc. 1994. À paraître

- LEROY-TURCAN, I. et T. R. WOOLDRIDGE (1995) : « L'informatisation des premiers dictionnaires de langue française : les difficultés propres à la première édition du *Dictionnaire de l'Académie française* », J. Pruvost (dir.), *Actes de la Journée des dictionnaires*, Université de Cergy-Pontoise (1995).
- LEVIN, B. (1993) : *English Verb Classes and Alternations (A Preliminary Investigation)*, The University of Chicago Press.
- LEVRAT, B. (1993) : *Le problème du sens dans les systèmes de traitement automatique du langage naturel : une approche alternative au travers de la paraphrase*, thèse de doctorat d'État, Université de Paris-Nord Villetaneuse, 214 p
- LEVRAT, B. et G. SABAH (1990) : « 'Sorte de' : une façon de rendre compte de la relation d'hyponymie/hyperonymie dans les réseaux sémantiques », *Langages*, n° 98, pp. 87-102.
- LEXIS (1987) : *Dictionnaire de la langue française Lexis*, Paris, Larousse.
- LIEBERT, W. A. (1994) : « Lascaux – a Hypermedia lexicon of Metaphor Models for Scientific Imagination », W. Martin et al. (Eds), *Euralex 1994*, Amsterdam, pp. 494-500
- LUN, S. (1983) : « A Two-Level Morphological Analysis of French », *Texas Linguistic Forum*, 22, pp. 271-278.
- LYONS, J. (1970) : *Linguistique générale. Introduction à la linguistique théorique*, Paris, Larousse, coll. « Langue et langage ».
- LYONS, J. (1977) : *Semantics*, vol. I, Cambridge, London, New York, Melbourne, Cambridge University Press.
- MACLEOD, C. et R. GRISHMAN (1994) : *COMLEX Syntax Reference Manual*.
- MANNING, C. D. (1993) : « Automatic Acquisition of a Large Subcategorization Dictionary from Corpora », *Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics*, Columbus (Ohio), 22-26 juin, pp. 235-242.
- MARCUS, M., KIM, G., MARCINKIEWICZ, M. A., MACINTYRE, R., BIES, A., FERGUSON, M., KATZ, K. et B. SCHASBERGER (1994) : « The Penn Treebank : Annotating Predicate Argument Structure », *Proceedings of the Human Language Technology Workshop*, Plainsboro (N. J.), 8-11 mars, pp. 114-119.
- MARTIN, John R. (1993) : *Life as a Noun . Arresting the Universe in Science and Humanities*, London, Falmer Press Ltd.
- MARTIN, R. (1994) : « Dictionnaire informatisé et traitement automatique de la polysémie », E. Martin (Éd.), *Textes et informatique*, coll. « Études de sémantique lexicale », Paris, CNRS-INaLF, Didier Érudition, pp. 77-113.
- MASSON, N. (1995) : « An Automatic Method for Text Structuring », *Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Edward A. Fox, Peter Ingwersen and Raya Fidel (Eds), Seattle (WA), July 9-13, pp. 372-373.
- MATHIEU, J.-P., KASTLER, A. et P. FLEURY (1985) : *Dictionnaire de physique*, 2^e édition, Paris, Masson/Eyrolles, 568 p. (MKF)

- MATHIEU-COLAS, Michel (1988) : « Typologie des noms composés », *Rapport technique n° 7, Programme de Recherches Coordonnées « Informatique et Linguistique »*, Université Paris 13.
- MATHIEU-COLAS, M. (1994) : *Les mots à trait d'union. Problèmes de lexicographie informatique*, coll. « Études de sémantique lexicale », Paris, CNRS-INaLF, Didier Érudition, 351 p.
- MATTHEWS, PH. (1974) : *Morphology : An Introduction to the Theory of Word-structure*, Cambridge, Cambridge University Press.
- MAUREL, Denis (1993) : « Passage d'un automate avec tables d'acceptabilité à un automate lexical », *Actes du colloque Informatique et langue naturelle*, Nantes, pp. 269-279.
- Mc CARTHY, J. J. (1979) : *Formal Problems in Semantic Phonology and Morphology*, Ph.D., MIT, Inédit
- McCAWLEY, N. A. (1976) : « On Experiencer Causatives », Shibatani (Ed.), *The Grammar of Causative Constructions*, coll. « Syntax and Semantics », vol. 6, New York.
- McNAUGHT, J., NKWENTI-AZEH, B., MARTIN, W. et E. TEN PAS (1991) : *EU-ROTRA-7 Study DOC-11 Feasibility of Standards for Terminological Description of Lexical Items*, Final Version (non publié).
- MEL'ČUK, I. (1981) : « Meaning-Text Models : a Recent Trend in Soviet Linguistics », *Annual Review of Anthropology*, 10, pp. 27-62.
- MEL'ČUK, I. A. (1982) : « Lexical Functions in Lexicographic Description », *Proceedings of the Eighth Annual Meeting of the Berkeley Linguistics Society*, Berkeley, UCB, pp. 427-444.
- MEL'ČUK, I. A. (1982a) : *Towards a Language of Linguistics, A System of Formal Notions for Theoretical Morphology*, München, Wilhem Fink Verlag.
- MEL'ČUK, I. A. (1982b) : « Élaboration d'un langage formel pour la morphologie », C. Bertaux, J-P. Desclés, D. Dubarle et al., *Linguistique et mathématiques Peut-on construire un discours cohérent en linguistique ?*, Berne, Peter Lang Verlag, pp. 99-119
- MEL'ČUK, I. A. (1988) : « Principes et critères de description sémantique dans le DEC », *Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexicosémantiques II*, Montréal, Les Presses de l'Université de Montréal, pp. 27-39.
- MEL'ČUK, I. A. (1994) : « Collocations and Lexical Functions », *Proceedings of the Leeds Colloquium on Collocations*.
- MEL'ČUK, I. A. (1994) : « Typologie des phrasèmes et leur présentation dans un dictionnaire de langue », Conférence plénière au Colloque International *La Locution*, Paris, ENS St-Cloud.
- MEL'ČUK, I. A. (1995a) : « Phrasemes in Language and Phraseology in Linguistics », M. Everaert et R. Schreuder (Eds), *Idioms : Structural and Psychological Perspectives*, Hillsdale/Hove, Lawrence Erlbaum Associates, pp. 167-232.
- MEL'ČUK, I. A. (1995b) : « The Future of the Lexicon in Linguistic Description . The Explanatory Combinatorial Dictionary », *Linguistics in the Morning Calm 3. Selected Papers of Seoul International Conference on Linguistics 1992*, Seoul, Hanshin Publishing Company, pp. 181-270.

- MEL'ČUK, I. A. (1996) : « Lexical Functions : A Powerful Tool for the Description of Lexical Relations in a Lexicon ». L. Wanner (Ed.), *Lexical Functions in Lexicography and Natural Language Processing*, Amsterdam, John Benjamins, pp. 37-102.
- MEL'ČUK, I. A., CLAS A. et A. POLGUÈRE (1995) : *Introduction à la lexicologie explicative et combinatoire*, coll. « Universités francophones » et « Champs Linguistiques », Louvain-la-Neuve, AUPELF-UREF et Duculot, 256 p.
- MEL'ČUK, I. et A. POLGUÈRE (1987) : « A Formal Lexicon in the Meaning-Text Theory (or how to do lexica with words) », *Computational Linguistics*, 13, pp. 261-275.
- MEL'ČUK, I. et al. (1984) : *DEC : Dictionnaire explicatif et combinatoire du français contemporain, recherches lexico-sémantiques I*, Montréal (Québec), Canada, Presses de l'Université de Montréal
- MEL'ČUK, I. et al. (1988) : *Dictionnaire explicatif et combinatoire du français contemporain Recherches lexico-sémantiques II*, Montréal, Presses de l'Université de Montréal
- MEL'ČUK, I. et al. (1992) : *DEC : Dictionnaire explicatif et combinatoire du français contemporain, recherches lexico-sémantiques III*, Montréal (Québec), Canada, Presses de l'Université de Montréal.
- MEUNIER, J.-G. (1970) : Communication personnelle.
- MEYER, C. (1993) : *Correction orthographique et grammaticale : vers une machine à dicter*, Rapport de diplôme, Université de Neuchâtel.
- MEYER, I. et B. MCHAFFIE (1993) : « De la focalisation à l'amplification . nouvelles perspectives de représentation des données terminologiques », *TA-TAO : recherches de pointe et applications immédiates Actes du colloque de Montréal*. Beyrouth, AUPELF-UREF et FMA, pp. 425-440
- MEYER, I., ONYSHKEVYCH, B. et L. CARLSON (1990) : *Lexicographic Principles and Design for Knowledge-based Machine Translation*, Technical Report CMU-CMT-90-118, Carnegie Mellon University.
- MICHIELS, A. (1995) : « Feeding LDOCE Entries into HORATIO », Alberto & Bennett (Eds), *Lexical Issues in Machine Translation*, Studies in Machine Translation and Natural Language Processing, Vol 8. Luxembourg, European Commission, pp. 93-115.
- MICHIELS, A. & J. NOËL (1984) : « The Pro's and Con's of a Controlled Defining Vocabulary in a Learner's Dictionary », *LEXeter'83 Proceedings*. Tübingen, Max Niemeyer, pp. 386-394.
- MICLET, L. (1980) : « Regular Inference with a Tail-Clustering Method », *I.E.E.E. Trans. on Systems, Man and Cybernetics*, 10, pp. 737-743.
- MILLER, A. G., FELLBAUM, C. et D. GROSS (1989) : « WORDNET a Lexical Database Organised on Psycholinguistic Principles », Zernik (Ed.), *Proceedings of the First International Lexical Acquisition Workshop, IJCAL*, Détroit.
- MILLER, G. A. (1990) : « Nouns in Wordnet . a Lexical Inheritance System », *International Journal of Lexicography*, 3 (4), pp. 245-264.
- MILLER, G. A. (dir.) (1990) : « WordNet : An On-Line Lexical Database », *International Journal of Lexicography*, special issue, 3 (4), pp. 235-312.

- MILLER, G. A. et C. FELLBAUM (1991) : « Semantic Networks of English », *Cognition*, n° 41, pp. 197-229.
- MILNER, Jean-Claude (1987) : *Introduction à une science du langage*, coll. « Des Travaux », Paris, Le Seuil.
- MONTEIL, Vincent (1960) : *L'arabe moderne*, Paris, Klincksiek (thèse de doctorat)
- MOODY, M. (1978) : « Some Preliminaries to a Theory of Morphology », *Glossa*, 12, pp. 16-38
- MOUSEL, P. et G. THIENPONT (1994) : *Technical Specification of the ANTHEM Prototype*, ANTHEM Deliverable n. D1-2.
- MULTEXT (1994) : *Common Specifications and Notation for Lexicon Encoding*, MULTEXT Report WP1.6.
- MUSTAFA-ELHADI, W. (1989) : *Terminologie arabe des télécommunications : théorie et faits de variation*. Thèse de doctorat en linguistique, Université Lyon II.
- MUSTAFA-ELHADI, W. (1990) : « The Contribution of Terminology to the Theoretical Conception of Classificatory Languages and Thesaurus Indexing », *Tools for Knowledge Organization and the Human Interface, Advances in Knowledge Organization*, vol. 1, Frankfurt, Index VERLAG, pp. 98-106.
- MUSTAFA-ELHADI, W. (1993) : *État de l'art de l'extraction de concepts à partir du langage naturel*, Rapport SGDN/STS/VST/5 & CREDO (CNRS-URA 1423), Université Charles De Gaulle-Lille III.
- MYAENG, S. H., KHOO, C. et M. LI (1994) : « Linguistic Processing of Text for a Large-Scale Conceptual Information Retrieval System », *Proceedings of the 2nd International Conference on Conceptual Structures*, Maryland (USA), août, dans *Lecture Notes in Artificial Intelligence*, n° 835, Springer-Verlag, Teufelhart, W. M., Dick, J. P. et Sowa, J. F. (dir.), pp. 69-83.
- NIDA, E. A. (1975) : *Componential Analysis of Meaning and Introduction to Semantic Structure*, The Hague, Mouton, 272 p.
- NIRENBURG, S., RASKIN, V. et B. ONYSHKEVYCH (1994) : *Apologiae ontologiae*, Memoranda in Computer and Cognitive Science MCS-95-281, New Mexico State University, Computing Research Laboratory
- NIWA, Yoshiki et Yoshihiko NITTA (1995) : « Co-occurrence vectors from corpora vs. distance vectors from dictionaries », *The Computation and Language E-Print Archive*, sur Internet.
- NOALLY, Michèle (1990) : *Le substantif épithète*, Paris, PUF.
- Nouveau Petit Robert* (1993) : *Le nouveau petit Robert Dictionnaire alphabétique et analogique de la langue française. Nouvelle édition remaniée et amplifiée*, Paris, Le Robert
- NUNBERG, Geoff, WASOW, Thomas et Ivan SAG (1994) : « Idioms », *Language. Oxford-Hachette French Dictionary* (1994) : Oxford University Press.

- Observatoire wallon des industries de la langue (1993) : *Des industries de la langue pour quoi faire ? Technologies, usages et marchés*, A. Moulin (dir.), Liège, Université de Liège.
- OGONOWSKI, A., HERVIOU, M. L. et E. DAUPHIN (1994) : « Tools for Extracting and Structuring Knowledge from Texts », *Proceedings of the 15th International Conference on Computational Linguistics*, Kyoto (Japon), 5-9 août, pp. 1049-1053.
- OH, H.-G., JOUIS, C., MAIRE-REPPERT, D. et J. BERRI (1992) : « Analyse de textes par exploration contextuelle : application aux problèmes des temps et à l'extraction des connaissances », *ECCO-92*, Orsay, 29 juin-1er juillet 1992, pp. 193-210
- ONYSHKEVYCH, B. et S. NIRENBURG (1994) : *The Lexicon in the Scheme of KBMT Things*, Technical Report MCCS-94-277, Computing Research Laboratory, New Mexico State University.
- ORACLE Corporation (1993) : *ORACLE, SQL*Forms, SQL*Plus, PL/SQL, Text*Retrieval and so on*, Redwood City.
- OTMAN, G. (1989) : « Terminologie et intelligence artificielle », *La Banque des mots*, pp. 63-95.
- OTMAN, G. (1995) : *Les représentations sémantiques en terminologie*, thèse de doctorat, Paris IV Sorbonne, 357 p
- PAASCH, H. (1901) : *From Keel to Truck = De la quille à la pomme de mâit = Vom Kiel zum Flaggenknopf. Dictionnaire de marine en anglais, français et allemand illustré de nombreux dessins explicatifs [...]*, 3^e édition, Anvers, H. Paasch, Hamburg, Eckardt & Messtorff.
- PAJZS, J. (1990) : « Számítógép és lexikográfia » (Ordinateur et lexicographie), *Linguistica*, Series A., 4. Budapest, MTA Nyelvtudományi Intézet (Institut de Linguistique de l'Académie des Sciences de Hongrie).
- PERENNOU, G. (1988) : « Le projet BDLEX de base de connaissances lexicales et phonologiques », *Actes des Premières Journées du GDR-PRC Communication Homme-Machine*
- PETITPIERRE, D., ROBERT, D. & S. WARWICK-ARMSTRONG (1994) : « DICO : A Network-Based Dictionary Consultation Tool », poster presented at the EURALEX'94 International Congress, Vrije Universiteit Amsterdam.
- PETITPIERRE, P. et G. ROBERT (1995) : *DICO, Technologiestandort Schweiz*, CeBIT 1995, Hanover.
- PETITPIERRE, P., ROBERT, G. et S. ARMSTRONG (1994) : *Design of an On-line Dictionary Consultation Tool*.
- PICABIA, L. (1978) : *Les constructions adjectivales en français. Systématique transformationnelle*, Genève et Paris, Droz.
- PICCHI, E., PETERS, C. & E. MARINAI (1992) : « The Pisa Lexicographic Workstation : The Bilingual Components », Tommola, Varantola, Salmi-Tolonen & Schopp (Eds), *EURALEX'92 Proceedings I-II, Fifth EURALEX International Congress*, Studia Translatologica, Ser. A, Vol. 1, University of Tampere, pp. 277-285.
- PICHT, H. (1987) : « Terms and their LSP Environment – LSP Phraseology », *Meta*, 23 (2), pp. 149-155.

- POLLARD, C. & I. A. SAG (1987) : *Information-Based Syntax and Semantics*, Stanford, CSLI Lecture Notes No. 13, 233 p.
- POLLARD, C. et I. SAG (1994) : *Head-Driven Phrase Structure Grammar*, CSLI/University of Chicago Press.
- PROCTER, P. (Ed.) (1978) : *Longman Dictionary of Contemporary English*, 2nd edition edited by D. Summers, Harlow, Longman Group Ltd.
- PROCTER, P. (Ed.) (1995) : *Cambridge International Dictionary of English*, Cambridge University Press.
- PUGEAULT, F., SAINT-DIZIER, P. et M.-G. MONTEIL (1994) : « Knowledge Extraction from Texts : A Method for Extracting Predicate-argument Structures from Texts », *Proceedings of the 15th International Conference on Computational Linguistics*, Kyoto (Japon), 5-9 août, pp 1039-1043.
- PUSTEJOVSKY, J. (1991) : « The Generative Lexicon », *Computational Linguistics*, 17(1)
- PUSTEJOVSKY, J. (Ed.) (1993) : *Semantics and the Lexicon*, Dordrecht, Kluwer.
- PUSTEJOVSKY, J. (1995 à paraître) : *The Generative Lexicon*. Cambridge, MIT.
- PUSTEJOVSKY, J. et P. BOUILLON (1995) : « Aspectual Coercion and Logical Polysemy », *Journal of Semantics*, vol. 2
- PUSTEJOVSKY, J. et S. BERGLER (1991) : « Lexical Semantics and Knowledge Representation », *Proceedings of a Workshop Sponsored by the Special Interest Group on the Lexicon of the Association for Computational Linguistics*, 17 June 1991, University of California, Berkeley, California, USA
- QUILLIAN, M. R. (1967) : « Word Concepts : a Theory and Simulation of some Basic Semantic Capabilities », *Behavioral Science*, 12 (5), pp. 410-443.
- QUILLIAN, R. (1968) : « Semantic Memory », *Semantic Information Processing*, M. Minsky (Ed.), Cambridge (Mass.), MIT Press, pp. 227-270.
- QUIRK, R., GREENBAUM, S., LEECH, G. et J. SVARTVIK (1994) : *A Comprehensive Grammar of the English Language*. London et New York, Longman, 12^e édition
- RAHMSTORF, G. (1993) : « Role and Representation of Terminological Definitions », *Actes du colloque Terminology & Knowledge Engineering*, Cologne, 25-27 août 1993, pp. 39-48.
- RAO, R., PEDERSEN, J. O., HEARST, M. A., MACKINLAY, J. D., CARD, S. K., MASINTER, L., HALVORSEN, P.-K. et G. G. ROBERTSON (1995) : « Rich Interaction in the Digital Library Communication », *ACM*, April 1995, 38 (4), pp. 29-39.
- RAPP, R. (1995) : « Identifying Word Translations in Non-Parallel Texts », *Proceedings of ACL '95*.
- RASTIER, François (1987) : *Langages : sémantique et intelligence artificielle*, Bernard Willerval Jouve, 14192 édition.

- RASTIER, F. (1987) : *Sémantique interprétative*, Paris, Presses Universitaires de France.
- RASTIER, F. (1991) : *Sémantique et recherches cognitives*, Paris, Presses Universitaires de France, 262 p.
- RASTIER, F. (1995) : « Le terme entre ontologie et linguistique ». *La Banque des Mots*, Numéro spécial du Centre de Terminologie et de Néologie du CNRS, actes de la Première Journée « Terminologie et Intelligence Artificielle », Paris, Villetaneuse.
- RASTIER, F., CAVAZZA, M. et A. ABEILLÉ (1994) : *Sémantique pour l'analyse – De la linguistique à l'informatique*, Paris, Masson
- REDDY, M. J. (1979) : « The Conduit Metaphor – a Case of Frame Conflict in our Language about Language », Andrew Ortony (Ed.), *Metaphor and Thought*, Cambridge University Press, pp. 284-324.
- REN, X. et F. PERRAULT (1992) : « The Typology of Unknown Words : An Experimental Study of Two Corpora », *Proceedings of COLING-92*, Nantes.
- RESNIK, Philip (1995) : « Using Information Content to Evaluate Semantic Similarity in a Taxonomy », *Actes de IJCAI'95*.
- REY-DEBOVE, J. (1971) : *Étude linguistique et sémiotique des dictionnaires français contemporains*, La Haye/Paris, Mouton.
- RILOFF, E. (1993) : « Automatically Constructing a Dictionary for Information Extraction Tasks », *Proceedings of the 11th National Conference on Artificial Intelligence*, Washington (DC), 11-15 juillet, pp. 811-816.
- RITCHIE, G. D., RUSSELL, G. J., BLACK, A. W. & S. G. PULMAN (1992) : *Computational Morphology : Practical Mechanisms for the English Lexicon*, Cambridge (MA), The MIT Press, 291 p.
- Robert (1994) : *Le Nouveau Petit Robert, dictionnaire alphabétique et analytique de la langue française*, Paris, Dictionnaires Le Robert.
- ROCHETTE, A. (1988) : *Semantic and Syntactic Aspects of Romance Sentential Complementation*, PhD thesis, Massachusetts Institute of Technology
- ROMAN, A. (1993) : « La voie des hypertextes ? », P. J. L. Arnaud & Ph. Thoiron (dir), *Aspects du vocabulaire*, Lyon, Presses universitaires de Lyon, pp. 103-132.
- RONDEAU, G. (1981) : *Introduction à la terminologie*, Centre éducatif et culturel, Québec
- ROULEAU, M. (1994) : *La traduction médicale (une approche méthodique)*, Brossard (Canada), Linguattech.
- RUESSINK, H.-A. (1990) : « Two-level Formalisms », Coopmans, P., Schouten, B , Zonneveld, W (Eds), *OTS Yearbook*, University of Utrecht.
- RUSSELL, G., BALLIM, A., CARROLL, J. et S. WARWICK-ARMSTRONG (1992) : « A Practical Approach to Multiple Default Inheritance for Unification-based Lexicons », *Computational Linguistics*, 18 (3).

- SABAH, G. (1988/89) : *L'intelligence artificielle et le langage*, vol 1 : *Représentation des connaissances* (1988), vol 2 : *Processus de compréhension* (1989), Paris, Hermès.
- SAGER, J. C. (1982) : « Definitions in Terminology », *Problèmes de la définition et de la synonymie en terminologie*, Actes du colloque international de terminologie, Université Laval, Québec, mai 1982, pp 113-140
- SAGER, J. C. (1990) : *A Practical Course in Terminology Processing*, Amsterdam, John Benjamins Publishing Company, 254 p.
- SALTON, Gerard (1988) : *Term-weighting Approaches in Automatic Text Retrieval*.
- SALTON, Gerard et J. M. MCGILL (1983) : *Introduction to Modern Information Retrieval*, McGraw-Hill Computer Science Series, New York. McGraw-Hill.
- SANFILIPPO, A. (Ed.) (1992) : *The (Other) Cambridge Acquilex papers*, Technical Report n° 253, University of Cambridge Computer Laboratory, New Museums Site
- SANFILIPPO, A. (1994) : « Word Knowledge Acquisition, Lexicon Construction and Dictionary Compilation », *Proceedings of the 15th International Conference on Computational Linguistics*, Kyoto (Japan), 5-9 août, pp. 273-277.
- SANFILIPPO, A. et V. POZNANSKI (1992) : « The Acquisition of Lexical Knowledge from Combined Machine-Readable Dictionary Sources », *Proceedings of the 3rd Conference on Applied Natural Language Processing*, Trento (Italy), 31 mars-3 avril, pp. 80-87.
- SCHMITT, L., OLIVAN, E., LANDI, B., ROYAUTÉ, J. et J. DUCLOY (1992) : « STDI · Une Station de travail pour une indexation assistée », *Natural Language Processing and its Applications*, Avignon
- SEFFAH, A. et J.-G. MEUNIER (1995) : « Un atelier génie logiciel orienté objets pour l'analyse cognitive de texte », *Actes du congrès JADT*, Rome.
- SEGOND, F. & A. ZAENEN (1994) : « Multi-word Expressions in Bilingual Dictionaries and in Compass », paper read at the workshop on « *The Future of the Dictionary* » co-sponsored by Rank Xerox Research Centre and Acquilex-II, Grenoble.
- SEGOND, Frédérique et Pasi TAPANAINEN (1995) : « Using a Finite-state Based Formalism to Identify and Generate Multiword Expressions », *Technical Report MLTT-019*, Rank Xerox Research Centre, Grenoble, July 1995
- SEKINE, S., CARROLL, J. J., ANANIADOU, S. et J. TSUJII (1992) : « Automatic Learning for Semantic Collocation », *Proceedings of the 3rd Conference on Applied Natural Language Processing*, Trento (Italy), 31 mars-3 avril, pp. 104-110
- SÉRASSET, G. (1994) : *SUBLIM : un système universel de bases lexicales multilingues et NADIA : sa spécialisation aux bases lexicales interlingues par acceptions*, Thèse nouveau doctorat, Université Joseph Fourier-Grenoble 1, 194 p.
- SÉRASSET, G. (1994a) : « Internal Lexical Organization for Multilingual Lexical Databases », *15th International Conference on Computational Linguistics, COLING-94*, Kyoto, August 5-9.
- SHIBATANI, M. (Ed.) (1976) : *The Grammar of Causative Constructions*, coll. « Syntax and Semantics », vol. 6, New York.

- SHIEBER, S. M. (1986) : *An Introduction to Unification-based Approaches to Grammar*, Stanford (CA), Stanford University CSLI, 105 p
- SIEGEL, D. (1974) : *Topics in English Morphology*, unpublished Doctoral dissertation, MIT.
- SILBERZTEIN, M. (1993) : *Dictionnaires électroniques et analyse automatique de textes . le système INTEX*, Paris, Masson, XIV + 233 p.
- SINCLAIR, John (Ed.) (1987) : *Collins Cobuild English Language Dictionary*. London et Glasgow, Collins.
- SINCLAIR, J. (1987) : *Looking UP. An account of the COBUILD project in lexical computing*, London, Collins Cobuild.
- SINCLAIR, J. (1991) : *Corpus, concordance, collocations*, Oxford, Oxford University Press.
- SLODZIAN, M. (1994) : « La doctrine terminologique, nouvelle théorie du signe au carrefour de l'universalisme et de du logicisme », *Actes de langue française et linguistique*, volume 7/8
- SLODZIAN, M. (1995) : « Comment revisiter la doctrine terminologique aujourd'hui ? », *La Banque des Mots*, Numéro spécial du Centre de Terminologie et de Néologie du CNRS, actes de la Première Journée "Terminologie et Intelligence Artificielle", Paris. Villetaneuse.
- SMADJA, F. (1991) : « Macrocoding the Lexicon with Co-occurrence Knowledge », Zernik (Ed.), *Lexical Acquisition : Using On-Line Resources to Build a Lexicon*, Hillsdale, Lawrence Erlbaum Associates, pp 165-189.
- SMADJA, F. (1993) : « Retrieving Collocations from Text . Xtract », *Computational Linguistics*, 19 (1), pp. 143-177.
- SMADJA, Frank A. et Kathleen R. MCKEOWN (1990) : « Automatically extracting and representing collocations for language generation », *Proceedings of the 28th Annual Meeting of the Association for Computational Linguistics*, pp. 252-259
- SMADJA, F. et K. MCKEOWN (1991) : « Using Collocations for Language Generation », *Computational Intelligence*, 7 (4), pp. 229-239.
- SNELL-HORNBY, M. (1988) : *Translation Studies . An Integrated Approach*, Amsterdam/Philadelphia, John Benjamins.
- SODERLAND, S., FISHER, D., ASELTINE, J. et W. LEHNERT (1995) : « CRYSTAL : Inducing a Conceptual Dictionary », *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, Montréal (Canada), 20-25 août, pp. 1314-1319.
- SOMERS, H. L. (1987) : *Valency and Case in Computational Linguistics*, Edinburgh University Press.
- SOWA, J. (1984) : *Conceptual Structures . Processing in Mind and Machine*, Reading (MA), Addison-Wesley.
- SPERBERG-MCQUEEN, C. M. et L. BURNARD (Eds) (1994) : *Guidelines for Electronic Text Encoding and Interchange*.

- STRAUSS, Steven L. (1979) : *Some Principles of Word Structure in English and German*, unpublished Doctoral dissertation, City University of New York, Graduate Center.
- STRAUSS, Steven L. (1982) : *Lexicalist Phonology of English and German*, Dordrecht-Holland/Cinnaminson-USA, Foris publications.
- STREITER, O., HALLER, J., SHARP, R., SCHMITT-WIGGER, A. et C. PEASE (1994) : « Aspects of a Unification Based Multilingual System for Computer Aided Translation », *Proceedings of the 14th International Conference « Avignon '94 »*, May 30th-June 3rd 1994.
- SURRIDGE, Marie E. (1985) : « Le genre grammatical des composés en français », *Revue canadienne de linguistique/Canadian Journal of Linguistics*, 30 (3), pp. 246-271.
- SWALES, John M. (1990) : *Genre Analysis*, Cambridge University Press.
- TAIFI, M. (1988) : « Problèmes méthodologiques relatifs à la confection d'un dictionnaire du tamazight », *Awal, Cahiers d'Études Berbères*, n° 4, Paris, Awal, pp. 15-26
- TAIFI, M. (1989) : *Le lexique berbère (parlers du Maroc central) : formes, sens et évolution*, thèse de doctorat d'État, Université de Paris III, Sorbonne nouvelle, XLIX + 940 p
- TAIFI, M. (1990a) : « Pour un théorie des schèmes en berbère », *Études et Documents Berbères*, n° 7, Paris, La Boîte à Documents, pp 92-110.
- TAIFI, M. (1990b) : « L'altération des racines berbères : la diachronie dans la synchronie », *Awal, Cahiers d'Études Berbères, numéro spécial en hommage à Mouloud Mammeri*, Paris, Awal, pp. 219-232.
- TAIFI, M. (1992) : *Dictionnaire tamazight-français (parlers du Maroc central)*, Paris, L'Harmattan/Awal, XXII + 879 p.
- TAIFI, M. (1995) : « Unité et diversité du berbère : détermination des lieux linguistiques d'intercompréhension », *Études et Documents Berbères*, n° 12, Paris, La Boîte à Documents/Edisud, pp 119-138.
- TAKEBE, Y. et al. (1976) : *The Japanese Thesaurus*, Sanseidou.
- TALMY, L. (1985) : « Lexicalization Patterns : Semantic Structure in Lexical Forms », Shopen, T. (Ed.), *Language Typology and Syntactic Description*, vol 3, Cambridge University Press.
- TANAKA, K. et K. UMEMURA (1994) : « Constuction of a Bilingual Dictionary Intermediated by a Third Language », *Proceedings of the International Conference for Computational Linguistics '94*, pp. 293-393.
- TAPANAINEN, Pasi (1994) : « RXRC Finite-State Rule Compiler », *Technical Report MLTT-020*, Rank Xerox Research Centre, Grenoble.
- TARSKI, A. (1936) : « Der Wahrheitsbegriff in den formalisierten Sprachen », *Studia philosophica*, I.
- TESNIÈRE, Louis (1959) : *Éléments de syntaxe structurale*, Paris, Librairie C. Klincksieck.
- THIELE, Johannes (1987) : *La formation des mots en français moderne*, Montréal, Presses de l'Université de Montréal.

- THOIRON, Ph. (1994) : « La terminologie multilingue . une aide à la maîtrise des concepts », *Meta*, 39 (4), déc. 1994, pp. 765-773
- THOIRON, Ph. et H. BÉJOINT (1991) : « La place des reformulations dans les textes scientifiques », *Meta*, 36 (1), pp 101-110.
- THOMAS, Patricia (1995) : *Orientation in Multiple Lexical Terms and Verb Phrases: A Model for Special Language Combinants*. Ph.D. Thesis, University of Surrey, Guildford, U.K.
- TOKUNAGA, T. et H. TANAKA (1990) : « The Automatic Extraction of Conceptual Items from Bilingual Dictionaries », *PRICAI*.
- TOMITA, M. (1984) : « Disambiguating Grammatically Ambiguous Sentences by Asking », *Proceedings 22nd Annual Meeting of the ACL*, Stanford, pp 476-480.
- TOURETZKY, D. (1994) : « Continuity, Polysemy and Representation : Understanding the Verb *Cut* », Fuchs, C. and Victorri, B. (Eds), *Continuity in Linguistic Semantics*, Amsterdam, John Benjamins, pp. 231-240.
- TROST, H. (1990) : « The application of two-level morphology to non-concatenative German morphology », *Proceedings of COLING-90*. vol. 2, Helsinki, pp. 371-376.
- UTSURO, T. *et al.* (1994) : « Bilingual Text Matching Using Bilingual Dictionary and Statistics », *Proceedings of the International Conference for Computational Linguistics '94*, pp. 1076-1082.
- VACHON-L'HEUREUX, Pierrette (1995) : « Table ronde sur les marques lexicographiques. Compte rendu », *Terminogramme*, n° 75. Office de la langue française, pp 1-6.
- VAN CAMPENHOUDT, M. (1991) : « *TI*, le logiciel d'expérimentation notionnelle de Termisti », *Terminologies nouvelles*, n° 5, pp. 11-14
- VAN CAMPENHOUDT, M. (1994) : *Un apport du monde maritime à la terminologie notionnelle multilingue : étude du dictionnaire du capitaine Heinrich Paasch « De la quille à la pomme du mât » (1885-1901)*, thèse de doctorat en sciences du langage, Université de Paris XIII.
- VERLINDE, S., BINON, J. & J. VAN DYCK (1992) : *Dictionnaire contextuel du français économique – Tome A : L'entreprise*, Louvain, Garant.
- VERONIS, J. & N. IDE (1994) : « From Dictionaries to Knowledge Bases... and Back », paper read at the workshop on « *The Future of the Dictionary* » co-sponsored by Rank Xerox Research Centre and Acquilex-II, Grenoble (abstract published under the title « Machine-Readable Dictionaries : Have we wasted our time? », *Cambridge Language Reference News*, Cambridge University Press, Number 4, p. 1).
- VIEGAS, E. (1995) : *Lexicon Development : Pro a Semi-automated Approach*. Technical Report (in preparation), Computing Research Laboratory, New Mexico State University
- VIEGAS, E. et M. GONZALES (1995) : *Derivational Morphology Rules to Enhance Lexicons Acquisition*, Technical Report (in preparation), Computing Research Laboratory, New Mexico State University.

- VIEGAS, E. et S. NIRENBURG (1995) : « The Semantic Recovery of Event Ellipsis: its Computational Treatment », *Proceedings of the Workshop on Context and Natural Language*, IJCAI 95, Montréal.
- VIEGAS, E. et S. NIRENBURG (1996) : « The Ecology of Lexical Acquisition : Computational Lexicon Making Process », Soumis à Euralex '96.
- WANG, William S. Y. (1977) : *The Lexicon in Phonological Change*, Paris. The Hague, Mouton.
- WANNER, L. (Ed.) (1996) : *Lexical Functions in Lexicography and Natural Language Processing*, Amsterdam, Benjamins.
- WARWICK, S. (1994) : « Automated Lexical Resources in Europe : A Survey », D Walker & A Zampolli (Eds.), *Automating the Lexicon*, Clarendon Press, forthcoming.
- WARWICK, S., J. HAJIC et G. RUSSELL (1990) : « Searching on Tagged Corpora. Linguistically Motivated Concordance Analysis », *Electronic Text Research. Proceedings of the Conference*, Waterloo, UW Centre for the New OED and Text Research.
- WASOW, T., SAG, I. et G. NUNBERG (1994) : « Idioms », *Language*, volume 70, pp. 491-538.
- WASTON, Thomas (1977) : « Transformations and the Lexicon », *Formal Syntax*, P. W. Culicover, Thomas and Adrian Akinajian (Eds), New York, Academic Press.
- WEBELHUTH, Gert (Ed.) (1995) : *Government and Binding Theory and the Minimalist Program*, Oxford (UK) & Cambridge (USA), Blackwell.
- WEHRLI, E. (1985) : « Design and Implementation of a Lexical Data Base », *Proceedings of the 2nd European ACL Conference*, pp. 146-153
- WILKS, Y. (1978) : « Making Preference more Active », *Artificial Intelligence*, 11, pp. 197-223.
- WILKS, Y., HUANG, X. & D. FASS (1985) : « Syntax, Preference and Right Attachment », *Proceedings IJCAI*, Los Angeles (CA), pp. 779-784.
- WITTGENSTEIN, L. von (1969) : *Philosophische Untersuchungen*, Frankfurt.
- WOOLDRIDGE, T. R. (1977) : *Les Débuts de la lexicographie française · Estienne, Nicot et le Thresor de la langue françoise (1606)*, Toronto/Buffalo, University of Toronto Press.
- WOOLDRIDGE, T. R. (1988) : « Les vocabulaire et fréquence métalinguistiques du discours lexicographique des principaux dictionnaires généraux monolingues français des XVI^e-XX^e siècles », *Travaux de linguistique et de philologie*, 26, Paris, Klincksieck, pp. 305-313.
- WOOLDRIDGE, T. R. (1993) : « Le flou en informatique textuelle », *Texte*, 13/14, Toronto, Édts Paratexte, pp. 275-289.
- WOOLDRIDGE, T. R. (1994) : « Projet d'informatisation du Dictionnaire de l'Académie (1694-1935) », B. Quemada (dir.), *Actes du Colloque sur le Dictionnaire de l'Académie française et la lexicographie institutionnelle européenne*, Institut de France (1995). À paraître.
- WOOLDRIDGE, T. R. (à paraître) : « Le mot métalinguistique du discours dictionnaire », *Cahiers de lexicologie*, Paris, Didier.

- WU, S. et U. MANBER (1992) : « Fast Text Searching Allowing Errors », *Communications of the ACM*, 35 (10), pp. 83-91.
- WURBEL, N. (1995) : *Dictionnaires et bases de connaissances : traitement automatique des données dictionnaires de langue française*, thèse de Doctorat en informatique, Université d'Aix-Marseille III, 262 p.
- WÜSTER, E. (1968) : *Dictionnaire multilingue de la machine-outil. Notions fondamentales, définies et illustrées, présentées dans l'ordre systématique et l'ordre alphabétique. Volume de base anglais-français = The Machine Tool An Interlingual Dictionary of Basic Concepts comprising an Alphabetical Dictionary and a Classified Vocabulary with Definitions and Illustrations. English-French Master Volume*, London, Technical Press.
- WÜSTER, E. (1971) : « Les classifications de notions et de thèmes. Différences essentielles et applications » = « Begriffs- und Themaklassifikationen. Unterschiede in ihrem Wesen und ihrer Anwendung », *Nachrichten für Dokumentation*, vol. 22 (3), pp. 98-104 et n° 4, pp. 143-150, traduit par INFOTERM, Bibliothèque d'INFOTERM.
- WÜSTER, E. (1981) : « L'étude scientifique générale de la terminologie, zone frontalière entre la linguistique, la logique, l'ontologie, l'informatique et la science des choses », Rondeau, G. et H. Felber (dir.), *Textes choisis de terminologie. Vol. 1. Fondements théoriques de la terminologie*, Québec, Université Laval - GIRSTERM, pp. 55-113.
- XEROX (1993) : *Lexical Tools : French Lexicon*, Palo Alto (CA), XEROX Corporation.
- XEROX (1995) : *Part of Speech Disambiguator and Text Tokenizer Reference : French Version 1.0*, Palo Alto (CA), XEROX Corporation.
- YAOLIANG, J. & D. ZHENDONG (1991) : « As a Participant in CICC MMT (ODA) Project », *Proceedings International Symposium on Multilingual Machine Translation*, (MMT-91), Beijing, 19-21 August 1991, vol. 1, pp. 13-15
- YAROWSKY, David (1992) : « Word-Sense Disambiguation Using Statistical Models of Roget's Categories Trained on Large Corpora », *Actes de COLLING-92*, Nantes, 23-28 août 1992.
- YAROWSKY, D. (1995) : « Unsupervised Word Sense Disambiguation Rivaling Supervised Methods », *Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics*, Cambridge (MA), 26-30 juin, pp. 189-196
- ZAMPOLLI, A., CALZOLARI, N. & M. PALMER (Eds) (1994) : *Current Issues in Computational Linguistics : in Honour of Don Walker*. Series « Linguistica Computazionale », IX-X. Pisa and Dordrecht, Giardini Editori and Kluwer Academic Publishers.
- ZERNIK, U. (1991) : *Lexical Acquisition : Exploiting On-Line Resources to Build a Lexicon*, Lawrence Erlbaum Associates.
- ZWICKY, A. M. (1992) : « Some Choices in the Theory of Morphology », R. Levine (Ed.), *Formal Grammar Theory and Implementation*, Oxford, Oxford University Press
- ZYZOMYS (1989) : Paris, ACT Informatique.

*Achévé d'imprimer sur
les presses de la SIEL (Beyrouth)
en décembre 1996*

La collection **Universités francophones** créée en 1988 à l'initiative de l'UREF, propose des ouvrages de référence, des manuels spécialisés et des actes de colloques scientifiques aux étudiants des 2^e et 3^e cycles universitaires ainsi qu'aux chercheurs francophones et se compose de titres originaux paraissant régulièrement.

Leurs auteurs appartiennent conjointement aux pays du Sud et du Nord et rendent compte des résultats des recherches et des études récentes entreprises en français à travers le monde. Ils permettent à cette collection pluridisciplinaire de couvrir progressivement l'ensemble des enseignements universitaires en français.

Enfin, la vente des ouvrages à un prix préférentiel destinés aux pays du Sud tient compte des exigences économiques nationales et assure une diffusion adaptée aux pays francophones.

Ainsi la collection **Universités francophones** constitue une bibliothèque de référence comprenant des ouvrages universitaires répondant aux besoins des étudiants de langue française.

Prix : 140 FF • Prix préférentiel UREF (Afrique, Asie, Amérique du Sud, Moyen-Orient) : 60 FF



9 782920 021709

ISSN 0993 - 3948