

**Lexicologie
Terminologie Traduction:
nouvelles recherches au cœur
d'un système**

Actes de la première journée
« Jeunes Chercheurs » du réseau
Lexicologie Terminologie Traduction

organisée à Bruxelles le 1^{er} février 2013
à l'Institut supérieur de traducteurs
et d'interprètes (ISTI)

Comité scientifique

Coordonnateur : Mathieu-Mangeot-Nagata (GETALP-LIG, Grenoble)

Ibrahim Ben Mrad (VALD, Université de La Manouba, Tunis)

Xavier Blanco Escoda (Laboratoire fLexSem, Université Autonome de Barcelone)

Aïcha Bouhjar (Institut royal de la culture amazighe, Rabat)

Mame Thierno Cissé (Laboratoire SOLDILAF, Université Cheikh Anta Diop de Dakar)

Anne Condamines (ERSS, CNRS-Université de Toulouse-le-Mirail)

Marie-Claude L'Homme (OLST, Université de Montréal)

Teresa Lino (Centro de Linguística, Universidade Nova de Lisboa)

François Maniez (CRTT, Université Lumière Lyon 2)

Salah Mejri (Laboratoire LDI, Université de Paris 13)

Franck Neveu (Université de Paris IV et ILF)

Patrice Pognan (Lalic, Institut national des langues et civilisations orientales)

Philippe Thoiron (CRTT, Université Lumière Lyon 2)

Amalia Todirascu (UR LILPA, Université de Strasbourg)

Marc Van Campenhoudt (Termisti, Institut supérieur de traducteurs et interprètes, Bruxelles)

Soutien

La *Première journée « Jeunes chercheurs »* du réseau *Lexicologie, Terminologie, Traduction* et la publication de ses actes ont bénéficié du soutien financier des organismes suivants :

Agence universitaire de la francophonie (Bureau Europe de l'Ouest)

Service de la langue française de la Fédération Wallonie-Bruxelles

Institut supérieur de traducteurs et interprètes de la Haute École de Bruxelles

Sommaire

AVANT-PROPOS p.8

Mathieu MANGEOT-NAGATA,
Groupe d'étude en traduction automatique –
traitement automatisé des langues et de la parole
Laboratoire d'informatique de Grenoble

Marc VAN CAMPENHOUDT,
Centre de recherche en linguistique appliquée Termisti
Institut supérieur de traducteurs et interprètes
Haute École de Bruxelles

AU CŒUR DE LA SÉMANTIQUE

Gaëlle DOUALAN, Université de Paris-Sorbonne
La synonymie, relation d'équivalence, un artefact de la pensée? ... p.15

Lassaâd OUESLATI, Université de Jendouba
Unités monolexicales et unités polylexicales : entre équivalence
sémantique et synonymie, le cas des adverbes p.43

Abdellatif CHEKIR, Université de Gabès
La traduction linguistique vers l'arabe : la métalangue polylexicale
comme exemple p.73

AUX FRONTIÈRES DU LEXIQUE

Ying ZHANG, Université de Grenoble
et Mathieu MANGEOT-NAGATA, Université de Savoie
Bases lexicales multilingues : traitement des acronymes p.93

Fethi JMIL, Université de la Manouba
Le traitement des interjections dans le dictionnaire
bilingue français-arabe p.117

DIRE LA PENSÉE COMPLEXE

Claire DUCARME, Université de Liège
Analyse de la microstructure comme prolégomènes à l'informatisation
d'un dictionnaire ancien : le cas particulier des Curiositez françaises
(1640) d'Antoine Oudin p.135

Ward VAN de VELDE,
Institut supérieur de traducteurs et interprètes
Le corpus BABELIRIS : d'une méthodologie de constitution
aux principales caractéristiques linguistiques p.163

Christina DECHAMPS, Université nouvelle de Lisbonne,
L'enseignement/apprentissage du français juridique :
une proposition de démarche terminologique et didactique p.191

Téléphone NGARAMBE, Université nationale du Rwanda
La traduction des aspects juridiques de la loi Gacaca du Rwanda.
Regards croisés p.225

AU CŒUR DE LA THÉORIE SENS-TEXTE

Yauheniya YAKUBOVICH,
Universitat Autònoma de Barcelona
Défigement dans les textes poétiques. Typologies et exemples
en français, espagnol, catalan, polonais, russe et biélorusse p.251

Mariana BOTTA,
Universidade de São Paulo et Université de Montréal
La terminologie de l'environnement en vulgarisation scientifique :
la famille lexicale de la régénération des forêts en portugais p.277

Avant-propos

Mathieu MANGEOT-NAGATA
*Groupe d'étude en traduction automatique –
traitement automatisé des langues et de la parole
Laboratoire d'informatique de Grenoble*
Marc VAN CAMPENHOUDT
*Centre de recherche en linguistique appliquée Termisti
Institut supérieur de traducteurs
et interprètes Haute École de Bruxelles*

Le réseau *Lexicologie, Terminologie, Traduction* (LTT) fédère seize centres de recherche et rassemble plus de huit cents chercheurs dans le domaine de la linguistique. Il a été lancé il y a plus de vingt ans au sein de l'AUELF-UREF¹, devenue plus tard l'Agence universitaire de la Francophonie. En 2010, il a acquis une personnalité juridique d'association internationale distincte de l'AUF, avec laquelle il continue de collaborer étroitement, tout en poursuivant les objectifs et les idéaux de ses fondateurs. Son siège social a été établi à l'Institut supérieur de traducteurs et interprètes, qui en exerce la présidence. Ce nouveau volume d'*Équivalences*, réalisé avec le soutien financier des Éditions du Hazard, de l'AUF et du Service de la langue française de la Fédération Wallonie-Bruxelles, tisse les liens d'une collaboration étroite – et déjà ancienne – entre le réseau LTT et l'ISTI.

À l'occasion de sa première assemblée générale, le comité scientifique du réseau a décidé d'organiser une journée d'étude des jeunes chercheurs dans les locaux de l'ISTI le 1^{er} février 2013. Destinée à réunir en priorité des doctorants et des post-doctorants travaillant au sein d'équipes affiliées au réseau ou intéressées à le rejoindre, cette rencontre avait pour ambition de maintenir la dynamique d'échange

1. Association des universités partiellement ou entièrement de langue française – Université des réseaux d'expression française.

et de transmission qui a animé le réseau depuis sa fondation. Dans la foulée des 9^{es} *Journées scientifiques de Villetaneuse* (2011) et des *Journées d'animation scientifique régionale de Tunis* (2012), elle a aussi été l'occasion de rendre un vibrant hommage à Marc Cheymol, ancien directeur du Programme *Langue française et diversité culturelle et linguistique* de l'AUF, auquel ce numéro d'*Équivalences* est dédié. Pendant de longues années, Marc Cheymol a coordonné l'action des réseaux de langue de l'Agence, percevant et ménageant avec beaucoup de finesse les spécificités du réseau LTT.

L'une de celles-ci – qu'il a toujours voulu préserver – est de constituer une communauté de chercheurs particulièrement attentive à la transmission des savoirs et des compétences entre les générations et par-delà les frontières. Les membres actuels du comité scientifique sont particulièrement soucieux de maintenir cette dynamique qui leur a jadis permis de valoriser leurs premières recherches universitaires. Lors des prestigieuses journées scientifiques du réseau, le « jeune » diplômé de troisième cycle, ou en passe de l'être, a autant de chance de présenter une communication que le chercheur blanchi sous le harnais, pourvu qu'il passe – comme le second, d'ailleurs – la barre d'une évaluation scientifique rigoureuse. En aucun cas, une « journée des jeunes chercheurs LTT » ne doit être perçue comme une forme d'antichambre à la pleine participation à un colloque de chercheurs patentés. En l'occurrence, en organisant cette rencontre, le réseau souhaitait avant tout valoriser des recherches doctorales ou post-doctorales menées dans les laboratoires qui en forment l'épine dorsale. Elles ont été présentées devant un public nombreux, dans la meilleure tradition de convivialité, d'écoute et de respect qui fait la réputation des rencontres LTT.

Volontairement large, l'appel à communications a permis aux conférenciers de présenter des recherches consacrées au

système linguistique et situées au cœur des préoccupations du réseau, notamment :

- l'équivalence et la synonymie ;
- l'évolution de la langue spécialisée ;
- la modélisation des dictionnaires et des corpus ;
- les outils d'aide à la traduction ;
- la description linguistique au service de l'intercompréhension.

Vingt résumés de communication ont été proposés. Ils étaient issus de douze pays et quatre continents, ce qui démontre, une fois de plus, la vitalité de la francophonie universitaire. Chacun a été évalué par au moins deux membres du comité scientifique, qui n'en a finalement conservé que treize. Les onze textes retenus pour publication dans ce numéro d'*Équivalences* ont, à leur tour, fait l'objet d'une évaluation scientifique. Au-delà de cette garantie de qualité, la diversité linguistique est également présente puisque quatorze langues figurent dans les articles acceptés : anglais, arabe, bélarus, chinois, espagnol, italien, français, japonais, kinyarwanda, néerlandais, polonais, portugais, russe et wolof. Les articles sont signés par leurs seuls auteurs véritables, une caractéristique qui tend à se perdre dans un monde de la course à la publication scientifique. Les éventuelles cosignatures signalent de véritables duos.

Les textes acceptés relèvent de quatre thèmes principaux correspondant aux séances du colloque. *Au cœur de la sémantique* regroupe des communications qui posent la question de la synonymie et de l'équivalence, au-delà des frontières du mot et d'une approche « entendue » de ces notions. Il s'agit des contributions de Gaëlle Doualan (Université de Paris-Sorbonne), de Lassaad Oueslati (Université de Jendouba) et d'Abdellatif Chekir (Université de Gabès). *Aux frontières du lexique* aborde le traitement lexicographique des unités lexicales le plus souvent exclues du dictionnaire. Cette séance comprenait les interventions de Ying Zhang (Université de Grenoble) et de Fethi Jmil (Université de la Manouba).

La séance intitulée *Dire la pensée complexe* était consacrée aux interventions de Claire Ducarme (Université de Liège), de Ward Van de Velde, (Institut supérieur de traducteurs et interprètes), de Christina Dechamps (Université nouvelle de Lisbonne) et de Téléphore Ngarambe (Université nationale du Rwanda). Ces chercheurs tentent de mettre de la structure, de l'ordre, de la clarification là où la communication spécialisée n'envisage pas l'intervention du linguiste et son souci du métalangage – qu'il s'agisse d'écrire un dictionnaire, de dire le droit ou d'informer un malade.

Les travaux d'Igor Mel'čuk, dont le centre de recherche fut longtemps le « camp de base » du réseau, ont nourri de nombreuses recherches autour de la théorie Sens-Texte au sein de LTT. Les articles de Yauheniya Yakubovich (Universitat Autònoma de Barcelona) et de Mariana Botta (Universidade de São Paulo et Université de Montréal) relèvent de la dernière séance de la journée, tout naturellement dénommée *Au cœur de la théorie Sens-Texte*.

S'il était un point commun à trouver à cet ensemble, au-delà des qualités communes, c'est assurément une approche fondée sur l'observation des faits de langue. Ceux-ci sont le plus souvent décrits sur la base de corpus authentiques, qu'ils soient ou non informatisés.

Cet avant-propos serait incomplet sans l'expression des plus vifs remerciements du réseau LTT aux professeurs Martine Bracops et Jean-Marie Van der Meerschen, qui ont proposé de publier les actes du colloque dans la prestigieuse revue de l'Institut supérieur de traducteurs et interprètes. Ils sont aussi adressés à Nathalie Lemaire, chercheuse au centre de recherche Termisti, qui, après la préparation des actes des Journées de Lisbonne et de Bukavu, a veillé à la révision et au toilettage des textes ici rassemblés.

Au cœur de la sémantique

La synonymie, relation d'équivalence, un artefact de la pensée ?

Gaëlle DOUALAN
L'Équipe d'accueil Sens, Texte, Informatique, Histoire
Université de Paris-Sorbonne

Résumé

Cette communication interroge l'histoire de la notion de synonymie pour tenter de comprendre si cette relation d'équivalence ne serait pas un simple artefact de la pensée. En effet, l'histoire de cette notion nous apprend, depuis l'Antiquité grecque, que l'équivalence sémantique est souvent rejetée entre les synonymes au profit de leurs différences de sens. Cette recherche des différences de sens ne connaît pas les mêmes motivations au cours du temps : nous rencontrons aussi bien un souci rhétorique de l'usage de la langue que la recherche d'une norme linguistique ou un usage didactique de la synonymie. C'est vers Aristote et sa conception toute personnelle de la synonymie qu'il faut se tourner pour retrouver une prédominance des ressemblances de sens sur les distinctions.

Mots-clés :

synonymie, équivalence, histoire de la synonymie, métalangage.

1. Introduction

La synonymie est la relation d'équivalence sémantique entre deux ou plusieurs unités lexicales dont la forme diffère (Lehmann & Martin-Berthet 1998 : 54). Ainsi, la synonymie est communément définie comme étant une relation d'équivalence. Seulement, elle fait également l'objet d'une définition en termes de substitution : *deux termes ou expressions sont synonymes lorsque la substitution de l'un à l'autre dans un même énoncé permet d'obtenir deux énoncés synonymes* (Arrivé et al. 1986 : 663). On comprend que substitution et équivalence sont étroitement liées ; il ne serait pas possible de substituer deux termes l'un à l'autre s'ils n'étaient pas équivalents.

Toutefois, l'équivalence peut être comprise comme une identité ou comme une approximation. La synonymie comme identité de sens a déjà été longuement étudiée et critiquée, notamment par Marie-Luce Honeste (2007). Nous prendrons le parti d'étudier la synonymie, relation d'équivalence, comme une approximation de sens, en tenant compte des définitions axées sur l'équivalence et celles axées sur la substitution.

Cependant, la synonymie comme équivalence de sens est souvent rejetée, car elle est trop floue ou car elle ne laisse que peu de possibilités de substitution. N'ayant pas de réalité dans la langue, on en vient à penser qu'elle serait seulement un artefact de la pensée, une commodité construite en vue de développer des études fines de la langue et du sens des mots.

Le but de notre démarche est de discuter le bien-fondé de la synonymie-artefact en nous appuyant sur des définitions passées et présentes de la synonymie. Nous nous plaçons donc résolument dans une perspective historique et terminographique. Ne pouvant prétendre à l'exhaustivité, nous avons constitué une petite anthologie de définitions de la synonymie, de l'Antiquité à nos jours. L'intérêt est d'envisager la plus large période historique en s'arrêtant sur les auteurs-clés de chacune de ces périodes. Ainsi, nous pouvons confronter les définitions de la synonymie au cours de l'histoire et examiner la place laissée à la synonymie comme relation d'équivalence.

Les définitions de la synonymie sont regroupées en trois sections correspondant chacune à un type de définition. La première section renvoie à la conception différentielle de la synonymie, conception dominante que l'on décèle dès l'Antiquité et qui perdure de nos jours. La seconde section expose la synonymie en pédagogie et en didactique de la Renaissance à nos jours en passant par le XIX^e siècle. Loin d'être simplifiée et réduite à une simple équivalence, la synonymie dans ce domaine tend à appliquer la conception distinctive de la synonymie. Enfin, la dernière section opère un retour aux sources puisqu'elle est consacrée à la définition aristotélicienne de la synonymie. Cette définition en marge permet de renouveler le débat sur la synonymie.

2. Une conception différentielle de la synonymie

Nombreuses ont été les conceptions différentielles de la synonymie. Le premier à avoir initié une telle conception a été Prodicos de Céos, sophiste grec. Malgré le peu d'éléments que nous avons sur lui, nous pouvons avancer qu'il s'intéressait au langage et au sens des mots. Il s'était notamment attaché à différencier des synonymes, du fait de leurs sens proches (Gomperz 2008). Mais son but premier était de renouer le lien entre le réel et la langue : un mot doit renvoyer de manière univoque à une chose, deux mots différents ne sauraient donc renvoyer à la même chose. François Rastier (2004 : 30) pose cela comme étant *l'axiome de Prodicos de Céos selon lequel deux synonymes ne sont jamais équivalents*. Il est donc impératif de différencier les mots synonymes. Cette conception de la synonymie a perduré et a donné lieu à une tradition séculaire (nommée par Sylvain Auroux (1986) « axiome Girard-Prodicos ») : *on pourrait même dire que la distinction des synonymes fut à la base de la dialectique, jusqu'au differentiae verborum de la seconde sophistique ; elle est au fond de la réflexion sémantique des Lumières, comme à l'origine de la sémantique différentielle saussurienne* (Rastier, 2004 : 30).

Nous n'examinerons que quelques-uns des auteurs qui ont participé à cette tradition, de l'Antiquité romaine à nos jours.

2.1. Les rhéteurs de l'Antiquité romaine : Quintilien

Pour la rhétorique antique, la synonymie est une figure de mots relevant de l'élocution. Cette conception de la synonymie – en rupture

par rapport à la synonymie des sophistes grecs ou celle d'Aristote – a été introduite par le rhéteur grec Caecilius de Calé-Acté (Desbordes 2007 : 336). *En tant que figure de mots, la synonymia est un phénomène du discours et non du lexique* (op. cit. : 338). Seulement, un rhéteur tel que Quintilien a approfondi la réflexion sur la synonymie car selon lui, elle *est loin d'être une équivalence absolue* (ibid.).

Mais comme plusieurs mots peuvent très souvent signifier la même chose (c'est ce que l'on appelle la synonymie), il y en a qui sont plus distingués, plus sublimes, plus brillants, plus agréables, plus éclatants que d'autres (Quintilien, Inst. Or. 8, 3, 16, trad. Jean Cousin, C.U.F).

Dans ce passage [...], Quintilien fait ainsi apparaître que dans les appréciations portées sur l'emploi des synonymes, il faut faire intervenir des notions telles que celles de la distinction, du ton, du genre littéraire, de l'âge du mot (Moussy 1994 : 9).

De par les propos qu'il tient sur la synonymie, nous pouvons replacer Quintilien dans la tradition différentielle même s'il est avant tout rhéteur, ce qui l'inscrit dans une conception de la synonymie comme procédé d'amplification. Il dépasse les équivalences entre les mots pour s'intéresser à ce qui les distingue. Ces distinctions ne sont pas toujours inhérentes au sens dénotatif des mots, elles peuvent aussi provenir de leurs connotations¹.

1. [...] la seule distinction des sens s'avère insuffisante pour justifier les choix faits par les écrivains et qu'il faut recourir à d'autres principes d'explications qui varient avec les mots considérés, mais dont souvent plusieurs sont valables en même temps : « âge, origine, particularités de construction, valeur intensive ou expressive, intellectuelle ou affective, degré de distinction ou de vulgarité (Marouzeau 1921 : 22) » (Moussy 1994 : 9).

Françoise Desbordes (2007 : 338) développe la conception de la synonymie chez Quintilien :

on peut bien, dans un énoncé, renvoyer à la même chose en juxtaposant plusieurs mots, mais, si des mots synonymes peuvent être additionnés, ils ne sont pas automatiquement interchangeables ; chacun a sa valeur propre, sinon son sens, et tel qui conviendra à tel endroit ne conviendra pas ailleurs.

Elle met en avant la place du contexte dans la synonymie. Les synonymes peuvent être accumulés dans un procédé d'amplification, autrement dit sur l'axe syntagmatique mais ils ne sauraient se substituer les uns aux autres dans un contexte donné, ce qui renvoie à l'axe paradigmatique. Ce qui importe donc dans la conception de Quintilien est son attachement à montrer que les mots possèdent des différences qui les empêchent d'être substitués les uns aux autres en contexte. En tant que rhéteur, il met l'accent sur l'effet produit sur le discours par tel ou tel mot. Ainsi, il *en tire un long développement sur le genre du mot qui convient à tel ou tel contexte* (*op. cit.*: 339) avec pour but la recherche d'une conformité entre le discours et la pensée, entre le mot et la chose.

François Desbordes résume ainsi la conception de la synonymie sous la Rome antique (*ibid.*) :

si les Latins ont bien conscience qu'il y a des mots qui « signifient la même chose », des synonymes, ils ont peut-être plus encore conscience d'une différence entre mots « de sens proches », entre synonymes.

Cela donne lieu à « la pratique assidue de la *differentia* » (*ibid.*) qui se manifeste sous la forme de recueils. Cette pratique marque un souci de précision, « un idéal de conformité du discours aux choses » (*op. cit.*: 338).

Les mots divergeant par leurs connotations, par les effets qu'ils produisent dans le discours, ils ne peuvent être substitués en contexte. Une telle conception de la synonymie place Quintilien dans la perspective différentielle de la synonymie. Notons que, même si les mots sont de sens proches, c'est-à-dire équivalents, cette équivalence est toujours invalidée par les différences de sens. Elle est donc comme un artefact de la pensée, sur lequel les différences de sens prennent pied.

2.2. Les remarqueurs du XVII^e siècle : le Père Bouhours

La synonymie distinctive a connu un essor considérable au XVII^e siècle avec les travaux des remarqueurs². Pour notre part, nous ne mentionnerons que l'un d'entre eux : le Père Bouhours. Il se place dans la filiation des *Remarques* de Vaugelas (1647). La définition donnée de la synonymie est assez pauvre puisqu'il s'agit de mots de même sens³. Cependant, cette définition s'enrichit de commentaires qui dénotent la pensée de la langue qui s'initie au XVII^e siècle⁴.

Selon Gilles Siouffi (2008), c'est avec le Père Bouhours que la synonymie évolue du syntagme au paradigme. Considérée comme un phénomène discursif, la synonymie se réalise sous forme de binômes synonymiques, c'est-à-dire des juxtapositions ou des coordinations de mots synonymes. En cela, elle est un procédé rhétorique

2. [Pour une synthèse complète sur la question, voir Vlassov (2012).

3. Bouhours, *Nouvelles remarques*, p. 196 : *les mots synonymes, c'est-à-dire qui signifient une même chose.*

4. [...] *elle [la langue française] ne peut supporter les périodes qui sont trop longues, les épithètes qui ne sont point nécessaires, les purs synonymes qui n'ajoutent rien au sens, & qui ne servent qu'à remplir leur nombre* (Bouhours 1671: 61).

d'amplification⁵. À cette synonymie *in praesentia*, est désormais opposée une synonymie *in absentia* par les remarqueurs. Cette seconde synonymie, de l'ordre du paradigmatique, permet d'examiner la concurrence des mots du lexique. Cette nouvelle conception de la synonymie oblige à une description fine du sens des mots, pris en couple⁶. Elle incite à prendre conscience de leur proximité de sens qui est telle qu'ils ne peuvent plus être associés en contexte, dans un binôme. *Une certaine forme de synonymie in absentia rend alors impossible la synonymie in praesentia* (Siouffi 2008 : 32).

Ce changement inhérent à la synonymie est à mettre en lien avec les changements que connaît la conception de la langue : *l'exigence de justesse va se substituer à l'ancienne poétique de l'amplification* (Siouffi 2012 : 370). Cette exigence de justesse se concrétise en une esthétique du bon usage, qui se veut normative. *Ne pourrait-on compter les Synonymes inutiles entre les fautes qui se commettent contre la justesse ?* (Bouhours 1674 : 241). Ainsi, la question de la synonymie – surtout celle de sa superfluité – prend pleinement part dans cette nouvelle esthétique de la langue.

Les remarqueurs de l'« après-Vaugelas » se trouvaient ainsi placés devant un véritable dilemme. Soit ils suivaient la pente normative qui semblait indiquée dans les Remarques, et ils étaient conduits à écarter un nombre de plus en plus important de mots de l'usage ; soit ils suivaient celle, tout aussi présente, de libéralité, et ils se trouvaient engagés dans la direction d'une recherche de nuances. (Siouffi 2012 :370)

5. Cette synonymie provient de la rhétorique antique. Elle se poursuit au Moyen-âge. Mais Claude Buridant (1980) montre que ces synonymies n'étaient pas de pures accumulations, elles avaient de nombreuses justifications : précision du sens, néologie, traduction...

6. Les traités des remarqueurs sont constitués d'articles confrontant des couples de mots.

Le père Bouhours se fait donc synonymiste et procède à l'étude des nuances existant entre les mots considérés comme synonymes. Il fonde cette synonymie distinctive en majeure partie sur l'observation de l'usage. Dès lors, bon nombre de mots synonymes sont sauvés par les nuances qui leur sont propres. Ces différenciations empêchent toute confusion entre les mots, ils ne sont donc plus aisément interchangeable. Cela participe de l'esthétique du bon usage qui se traduit par la volonté de parler avec clarté en utilisant de manière appropriée la langue et donc les mots.

Cette attention portée au sens des mots ne fait plus de la synonymie une simple relation d'équivalence. Au contraire, ce sont les différences de sens qui priment ; les ressemblances sont rejetées afin de sauver la richesse de la langue.

2.3. À l'aube des Lumières : l'abbé Girard

Au XVII^e siècle, les remarqueurs procédaient à des études ponctuelles sur des couples de mots dont ils mettaient au jour les distinctions de sens. Au XVIII^e siècle naissent les premiers « dictionnaires de synonymes » qui sont en réalité une systématisation des remarques. Du point de vue quantitatif, le nombre de mots traités augmente considérablement.

Le premier de ces dictionnaires est la *Justesse de la langue française* (1718) de l'abbé Girard. Il se présente sous la forme d'un recueil à entrées multiples. L'article correspondant à ces entrées faites de paires (ou plus) de synonymes a pour objectif de faire apparaître leurs nuances propres.

Dans la préface de son ouvrage, Girard distingue deux types de synonymie. Une synonymie que nous appellerions « absolue », où les synonymes sont des mots dont le sens est exactement le même. Et une seconde synonymie où les mots synonymes connaissent des idées communes mais se distinguent par des idées accessoires qui leur sont propres (Girard 1718 : XXVII-XXIX).

Il faut premièrement convenir de ce que c'est que synonyme. Ce terme peut être pris en deux sens ; l'un plus étendu, l'autre plus resserré. On peut le prendre pour une ressemblance de signification qui souffre cependant quelque variété ; c'est-à-dire que les termes synonymes profèrent tous une même idée principale ; mais que chacun d'eux y ajoute néanmoins quelques idées accessoires, qui diversifient la principale ; ensorte qu'elle paroisse dans ces différents mots, comme une même couleur paroît sous diverses nuances. On peut aussi entendre par synonyme une ressemblance de signification si entière & si parfaite que le sens, pris dans toute la force & dans toutes les circonstances, soit toujours & absolument le même ; ensorte qu'un des synonymes ne signifie li plus ni moins que l'autre ; qu'on puisse les employer indifféremment dans toutes les occasions ; & qu'il n'y ait pas plus de choix à faire entre eux, pour le sens, qu'entre les gouttes d'eau d'une même source, pour le goût.

Girard reconnaît qu'une synonymie absolue n'est d'aucune utilité pour la langue. Il retient donc la synonymie faite d'idées communes et d'idées accessoires. Partant des idées communes, il recherche les idées accessoires qui distinguent des synonymes, la dissection de ces idées accessoires étant l'objet des articles de son ouvrage. De ce fait, la synonymie qui serait retenue par Girard fait la part belle aux nuances de sens.

L'abbé Girard se place donc dans le prolongement des remarqueurs puisqu'il adopte la synonymie distinctive qu'ils ont élaborée et s'attache à décrire les distinctions de sens existant entre les synonymes. Cependant, il conceptualise davantage ces différences de sens

puisqu'il décompose les unités de la langue en deux types d'idées. Il n'en demeure pas moins que son travail différentiel est, lui aussi, guidé par l'esthétique du bon usage. En effet, comme le rappelle Pierre Swiggers (2008), le public de l'abbé Girard est un public instruit qui recherche le maniement de la langue dans l'art de la conversation. La justesse de l'expression se traduit donc par le choix correct des mots.

À la fin de sa préface, l'abbé Girard s'interroge : *le rapport qu'il y a des mots aux idées ne peut-il pas être encore une preuve que la différence des uns marque toujours la différence des autres* (Girard 1718 : XLV). Ainsi, la substitution d'un mot par un autre engendrerait une modification dans les idées exprimées et la justesse de la pensée ne serait plus de mise. Une fois de plus, les mots ne sauraient être considérés comme des équivalents, ils ne peuvent se laisser prendre au jeu de la substitution.

2.4. Le XIX^e siècle

2.4.1. Les lexicographes

La lexicographie de la synonymie s'est poursuivie au XVIII^e mais aussi au XIX^e siècle, avec des synonymistes tels que François Guizot ou Pierre-Benjamin Lafaye. Ils héritent de la vision distinctive de la synonymie et configurent leurs dictionnaires de la même manière que précédemment : des entrées multiples suivies d'articles cherchant à faire ressortir leurs différences de sens. François Guizot reprend à l'abbé Girard certains de ces termes : « idée commune » et « idée accessoire » (Aruta Stampacchia 2006). Mais surtout, il définit

les synonymes comme *les termes dont le sens a de grands rapports et des différences légères, mais réelles* (Guizot 1809/1859 : XVIII). Ici encore les synonymes ont des différences de sens qui sont l'objet de toutes les attentions dans les dictionnaires.

Chez Pierre-Benjamin Lafaye, les synonymes sont également distingués en idées principales et accessoires⁷ car un dictionnaire des synonymes doit *faire ressortir les nuances de signification qui séparent un terme d'un autre* (Aruta Stampacchia 2006 : 32), ce que ne font pas les dictionnaires de langue, auxquels il fait ce reproche. En revanche, sa singularité réside dans la tentative de proposer des représentations géométriques des synonymes : il s'agit de diagrammes en formes de cercles se chevauchant. Les chevauchements renvoient aux idées communes à plusieurs synonymes, les autres sections des cercles symbolisant les idées accessoires, propres à chaque synonyme. Ces schématisations montrent donc une impossible superposition des synonymes, qui possèdent nécessairement des différences.

De par leur travail lexicographique, François Guizot et Pierre-Benjamin Lafaye, s'inscrivant dans la lignée de la synonymie distinctive, ne sauraient accepter une conception de la synonymie se réduisant à une simple équivalence. Au contraire, les différences de sens sont au cœur de leurs recherches. Lafaye semble même rejeter la substitution au vu des représentations géométriques qu'il élabore.

7. En cela, il suit la tradition de la synonymie distinctive concrétisée sous forme de dictionnaires, le sien s'articulant comme ceux de ses prédécesseurs.

2.4.2. Naissance de la théorisation linguistique : Saussure

À la fin du XIX^e siècle, Saussure s'appuie sur les recherches distinctives depuis l'abbé Girard pour proposer sa théorie de la valeur. C'est ce que postule Sylvain Auroux (1984 : 105) :

l'hypothèse historique la plus vraisemblable consiste à admettre que la théorie classique de la synonymie, non seulement est intégrée à la théorie linguistique saussurienne, mais encore qu'elle en est probablement un élément générateur.

En effet, Saussure prend pour exemple des mots synonymes pour expliciter ce qu'il entend par valeur : *dans l'intérieur d'une même langue, tous les mots qui expriment des idées voisines se limitent respectivement : des synonymes comme redouter, craindre, avoir peur n'ont de valeur propre que par leur opposition ; si redouter n'existait pas, tout son contenu irait à ses concurrents* (Saussure 1916/1986 : 160). Comme le formule Valentina Bisconti (2012 : 715), *Saussure se penche sur le mécanisme de la synonymie pour illustrer la solidarité sémiologique entre les signes et, par conséquent, la notion de valeur*. Les unités du lexique font système, elles s'attribuent leur propre valeur, en se répartissant les significations. Ainsi, elles se différencient les unes des autres et ne peuvent donc être équivalentes. Elles ont toujours une partie de leur signification qui leur est propre si bien qu'elles ne peuvent se substituer les unes aux autres.

2.5. La linguistique contemporaine

Avec le développement de la linguistique au XX^e siècle, les définitions de la synonymie sont légion. Cependant, nous ne pouvons nous

attacher qu'à un auteur afin de pouvoir développer sa pensée. Nous choisissons tout particulièrement Marie-Luce Honeste pour son inscription dans la mouvance de la synonymie distinctive.

Rejetant une synonymie totale – relation d'identité, dans tous les contextes – (Honeste, 2007 : 161-162), elle reconnaît en revanche la prégnance de la synonymie partielle dans les recherches en linguistique. Cette synonymie serait une équivalence au sens où nous l'entendons, c'est-à-dire une approximation. Elle aurait donc la particularité d'offrir des possibilités de substitution dans certains contextes (*Ibid.* : 163). Voici une définition consensuelle de cette synonymie partielle par Sabine Ploux et Bernard Victorri (1998) :

[...] deux unités lexicales sont en relation de synonymie si toute occurrence de l'une peut être remplacée par une occurrence de l'autre dans un certain nombre d'environnements sans modifier notablement le sens de l'énoncé dans lequel elle se trouve.

Cependant, Marie-Luce Honeste rejette la synonymie partielle car *ce qui fonde la synonymie, c'est alors le fait de rendre prégnante en discours les ressemblances et non les différences* (Honeste, 2007 : 163). Selon elle, une telle synonymie implique un « irréductible implicite référentiel » qu'elle combat puisqu'elle se place dans une approche conceptuelle du sens. Dans cette approche, la dénotation qui fondait la synonymie est évacuée, il ne saurait donc y avoir de synonymie en langue. Il n'y a pas davantage de synonymie en discours. En effet, suite à des enquêtes, elle peut affirmer que *les locuteurs manifestent une grande sensibilité aux plus infimes nuances des signifiés ; c'est un facteur défavorable à l'usage de la synonymie* (*op. cit.* : 168).

Elle résume ainsi son propos : d'un point de vue linguistique, *la synonymie relève d'une approche référentielle* (*Ibid.* : 171) en déphasage avec le fonctionnement réel de la langue, et, d'un point de vue discursif, la synonymie brouille les différences sémantiques données par les auteurs ou sert à manipuler ou travestir le réel. En définitive, elle considère qu'il est préférable de s'attacher aux différences de sens plutôt qu'aux ressemblances, car les ressemblances sont plus enrichissantes intellectuellement et plus respectueuses de la diversité du monde et de sa perception (*op. cit.* : 171).

À ce sujet, elle propose dans le développement de son article une étude lexico-sémantique de deux verbes considérés comme synonymes : *perdre et égarer*. Après confrontation de ces deux items dans des contextes similaires, elle en conclut une dissemblance nette, qui est passée sous silence lorsqu'on les emploie en discours. Elle met donc bien en avant les différences de sens, qui se doivent de primer sur les ressemblances, ce qui aboutit à une impossible substituabilité.

Marie-Luce Honeste n'est toutefois pas la seule à proposer de telles études lexico-sémantiques. D'autres linguistes analysent les mots considérés comme synonymes par paires afin de dégager leurs différences de sens. C'est le cas notamment de Peter Blumenthal, qui dans « De la logique des mots à l'analyse de la synonymie » (2006), propose la confrontation entre *surprise* et *étonnement* et entre *chagrin* et *tristesse*. Tout comme Marie-Luce Honeste, il recherche les logiques sous-jacentes qui permettent de différencier les synonymes. Mais pour sa part, il s'attache plus précisément à l'étude des collocations de ces synonymes.

Ces études centrées sur un couple de synonymes perpétuent la technique utilisée par les synonymistes des siècles passés, pour aboutir à des résultats semblables : dégager les différences de sens qui distinguent deux synonymes pour montrer qu'ils ne sont pas équivalents et qu'ils ne doivent donc pas être confondus en discours.

Il ressort de cette étude de la conception différentielle de la synonymie, aussi appelée « synonymie distinctive » que la synonymie est pensée comme un rapprochement de mots de sens proches. Cependant, il est souvent fait peu de cas de ces ressemblances de sens car l'attention de l'étude est aussitôt portée sur les différences sémantiques qui séparent ces synonymes. D'ailleurs, ces différences semblent toujours prendre le pas sur les ressemblances puisqu'elles sont l'objet d'articles lexicographiques ou de commentaires métalinguistiques.

Ainsi, la synonymie se place dans l'optique d'une sémantique différentielle où il est question de distinguer les mots les uns des autres. Cela peut être motivé par différentes raisons, notamment la norme du bon usage ou les études de conditions d'emploi des items. En somme, les ressemblances entre les mots ne sont qu'un point de départ et un prétexte à une étude plus approfondie des unités de la langue afin de proposer une description fine du lexique.

L'équivalence entre les synonymes étant rejetée au nom des distinctions de sens, la synonymie en tant que relation d'équivalence est, selon cette perspective, un artefact de la pensée linguistique. Elle est une construction de la pensée qui apporte un fondement à l'étude différentielle du lexique.

3. Les synonymies cumulatives et didactiques

3.1. Amplification et manuels de pédagogie au Moyen-âge et à la Renaissance

Au Moyen-âge, homonymie et synonymie⁸ ont été étudiées de pair dans des manuels qui *apportent du matériel aux discussions des dialecticiens, en particulier quand il s'agit de définir les relations de l'aequivocatio et de l'univocatio*⁹ (Chevalier, 1997 : 9). Au XIII^e siècle, les manuels les plus répandus sont attribués à Jean de Garlande (*ibid.*). Outre leur usage dialectique, ces manuels *servent aussi aux travaux d'écriture, la synonymie étant une des figures les plus courantes de la prose médiévale* (*ibid.*). De plus, *ils figurent parmi les livres de base de l'étudiant et sont souvent insérés dans des recueils d'enseignement* (*ibid.*).

À la Renaissance, Guy de Fontenay se fait le continuateur de Jean de Garlande comme l'indique le titre de l'un de ses ouvrages *Magna synonyma noviter per eum repurgata ... cui Olla patella nomen inditum est auctore Joanne de Garlandia* (1517). Cet ouvrage fait une large place à la synonymie qu'il définit ainsi : *des mots* (dictiones) *qui se rapportent à la même signification* (in eandem cadentes significationem) (Chevalier 1997 : 18). Ce manuel, dédié aux jeunes gens aussi bien qu'aux *amateurs de composition poétique et oratoire* (*ibid.*), est constitué d'articles dont l'entrée est un mot français suivi d'une

8. Ce couple de notions est issu d'Aristote qui les définit ensemble dans *Catégorie*.

9. Ces deux termes sont les équivalents latins de l'*homonyma* et de la *synonyma* grecques.

liste de synonymes latins, souvent conséquente¹⁰. Le lecteur n'est aidé par aucune indication de registres, de genre, etc. dans ces listes. Cela laisse une grande liberté au compositeur si bien que *les termes seront utilisés aussi bien en substitution qu'en juxtaposition* (op. cit. : 19). Cette juxtaposition renvoie à la synonymie amplification qui se concrétise sous forme de binômes synonymiques. Quant à la substitution, elle laisse entendre que la synonymie paradigmatique est en germe, déjà à cette époque¹¹.

La liste pléthorique de synonymes ainsi que la définition donnée par Guy de Fontenay laisse à penser que cette synonymie, cumulative et didactique, est perçue comme une approximation qui, de par son caractère flou, laisse une grande marge de manœuvre aux compositeurs et aux étudiants. Les manuels compilent autant de synonymes que possible afin d'offrir le plus grand choix aux scripteurs. Cependant, ce serait avoir une vision réductrice de cette synonymie que de penser cela. Claude Buridant (1980) montre par de nombreux exemples que les binômes synonymiques du Moyen-âge et de la Renaissance n'étaient pas de pures redondances. Bien au contraire, leurs emplois se justifient systématiquement : pour préciser le sens d'un mot vague (hyperonymique), pour faciliter la compréhension d'un néologisme à l'aide d'un mot connu, comme technique de traduction quand aucun équivalent n'existe dans la langue cible... De ce fait, cette synonymie cumulative n'est pas une équivalence : les synonymes juxtaposés se dissimilent pour mieux se compléter sémantiquement.

10. *Pauvreté* ne compte pas moins de dix-huit synonymes et *lacif (sic)* quarante-trois (Chevalier 1997 : 19).

11. Même si elle semble apparaître pour de bon avec les remarqueurs du XVII^e siècle.

En outre, la synonymie commence à se penser en termes différentiels. En effet, à la fin de son manuel, Guy de Fontenay y joint un petit « Traité des différences » (Chevalier, 1997 : 21). Les mots sont examinés dans des contextes et en termes de « bon voisinage » (*ibid.*). Mais cette émergence d'une nouvelle synonymie est plus frappante chez Gérard de Vivre (1569) dans ses *Synonymes*. Son manuel était destiné aux jeunes Allemands apprenant le français. Gérard de Vivre leur proposait des synonymes afin d'apprendre le français et d'enrichir leur expression. Il s'attachait donc tout particulièrement à faire fonctionner les synonymes en contexte : *les phrases [d'exemple] sont forgées avec le souci de mettre en valeur, pour l'apprenant étranger, les contextes propres à chaque synonyme* (Leclercq 2008 : 19). Le contexte prend donc une place plus importante dans l'étude de la synonymie. Il aide les apprenants à comprendre les différences de sens qui existent entre synonymes. *La substitution possible d'un mot par un autre [...] est toujours subordonnée à un contexte d'emploi particulier* (*ibid.*). Ces études en contexte aident aussi les apprenants à employer le bon mot dans le bon contexte, ce qui améliore les vertus pédagogiques de la synonymie. Ainsi, comme l'avance Odile Leclercq, Gérard de Vivre serait un précurseur de la synonymie distinctive, avant les remarqueurs et l'abbé Girard.

En résumé, même si la synonymie dans les manuels de pédagogie est souvent présentée par accumulation, c'est-à-dire sous forme de listes, elle n'est pas une simple approximation de sens. Au contraire, elle tend vers la synonymie distinctive dès lors que le contexte est introduit dans l'étude des synonymes. Là encore, les ressemblances entre synonymes sont nécessaires pour les regrouper en listes mais

ces rapprochements donnent lieu à un examen plus attentif qui aboutit toujours à la mise en avant des différences de sens. Les exemples étudiés demeurent affiliés à la synonymie cumulative et didactique mais la conception distinctive de la synonymie commence à se faire jour.

3.2. La synonymie comme outil didactique

La synonymie présente dans les manuels scolaires, apprise aux écoliers dès le cours élémentaire devient un outil de l'apprentissage de la langue, plus précisément du vocabulaire. Valentina Bisconti (2008 : 711) le considère comme un « dispositif pédagogique ». Elle étudie à ce sujet les manuels de Pierre Larousse¹² selon qui *la synonymie est conçue comme stratégie d'un éveil lexical et sémantique (ibid.)*. D'après Valentina Bisconti, Larousse estime *qu'il n'y a jamais deux mots qui sont synonymes parce que la langue ne s'embarrasse pas de la redondance [...] la pauvreté de la langue serait également un défaut nuisible à l'expression de la pensée (op. cit. : 712)*. Dans cette définition héritée des Lumières, la synonymie est donc nécessairement distinctive. Larousse va donc proposer aux élèves des exercices où il sera question de différencier des synonymes à partir de contextes d'usage ou par des rapprochements car *c'est précisément l'écart qui fait ressortir la signification (op. cit. : 713)*.

La synonymie pédagogique chez Larousse s'affilie également à l'esthétique du bon usage, non pour un maniement de la langue par

des personnes cultivées mais dans l'optique de son projet d'instruction. Les jeunes écoliers doivent apprendre leur langue dans sa richesse et apprendre à la parler correctement, si bien que ce projet d'instruction prend forcément un tour normatif.

L'exemple de Pierre Larousse montre que la synonymie pédagogique n'est nullement une approximation. Ce n'est pas parce qu'elle sert à apprendre le lexique aux écoliers qu'elle est simplifiée, réduite à une simple approximation de sens. Au contraire, elle peut se montrer distinctive afin de faire prendre conscience aux apprenants des nuances de la langue pour qu'ils les intègrent.

La synonymie enseignée aux écoliers ne se limite pas à reconnaître et différencier des synonymes. Elle est également employée dans la production écrite afin de favoriser la mémorisation et la mobilisation des synonymes de manière pertinente (Garcia-Debanc *et al.* 2009 : 215). Cependant, en production écrite, ce n'est plus seulement la synonymie qui est convoquée mais aussi la cohyponymie (*op. cit.* : 213) :

Le travail de formulation passe parfois par une recherche de synonyme mais il implique aussi très souvent la recherche d'un terme plus spécifique, dans une relation d'hyponymie/hyponymie.

La synonymie peut être utilisée pour éviter les répétitions lors de la production mais la cohyponymie intervient quand les élèves ont tendance à employer de manière récurrentes des termes très courants et donc généraux et vagues, ce qui ne favorise pas l'enrichissement du vocabulaire des apprenants. *Le travail porte moins sur la recherche de synonymes que de cohyponymes [de verbes] qualifiés*

12. *Lexicologies des écoles* (1853-1858). Voir (Bisconti 2012 : 720, note 6).

de « passe-partout », tels que *faire, dire, mettre, aller*¹³ (*op. cit.* : 209). En conséquence, la synonymie en didactique laisse souvent la place à la cohyponymie. Elle permet davantage l'enrichissement lexical des élèves mais aussi, elle contribue à leur faire prendre conscience de la structuration du lexique, en champs thématiques et sémantiques.

La synonymie en didactique ne se réduit pas à de simples accumulations. Au contraire, elle cherche à distinguer les nuances de sens entre synonymes, via notamment leurs contextes d'emploi. Elle se fait donc volontiers distinctive. Cela incite les apprenants à un usage judicieux de la langue afin d'éviter toute confusion. Par là, elle enrichit leur vocabulaire et leur fait prendre conscience de la structuration du lexique. Cependant, la synonymie se partage le terrain avec la cohyponymie. Ces deux notions sont justement liées dans la conception aristotélicienne de la synonymie, dernière conception de la synonymie que nous étudions.

4. La synonymie aristotélicienne

Une troisième conception de la synonymie se dégage au cours de l'histoire de cette notion : il s'agit de la synonymie telle que définie par Aristote. Cette synonymie est le plus souvent rapprochée de la notion de cohyponymie, c'est-à-dire une relation entre deux unités

qui renvoient au même hyperonyme. Voici la définition donnée par Aristote (2002 : 59) :

[1a] On dit synonymes les items qui ont le nom en commun, et dont l'énoncé de l'essence, correspondant au nom, est le même par exemple [on dit] zôion [être animé] pour l'homme et pour le bœuf ; [on les dit synonymes] car on les appelle tous les deux du commun nom de zôion et l'énoncé de l'essence, correspondant au nom, est le même. Si en effet, on doit donner la définition de chacun d'eux, de ce que c'est, pour chacun d'eux, d'être zôion, on donnera la même définition.

Ainsi, pour Aristote, *homme* et *bœuf* sont synonymes dans la mesure où la caractéristique « être animé » est valable aussi bien pour l'un que pour l'autre, en sachant que cette caractéristique renvoie à la même chose chez l'un comme chez l'autre. Si on lit cette définition au travers du prisme de la cohyponymie, *homme* et *bœuf* sont les hyponymes de l'hyperonyme *être animé*. Cette synonymie aristotélicienne étant plutôt de nature hyponymique, il ne saurait être question d'équivalence ou de substitution entre les items concernés par cette relation. D'ailleurs, Aristote n'en fait nullement mention, il est donc en dehors de la tradition de la synonymie distinctive.

Franco Lo Piparo (2000) propose une analyse de la synonymie aristotélicienne en termes de commensurabilité (concept mathématique).

'Homme' et 'bœuf' sont synonymes non pas parce qu'ils ont des significations partiellement ou totalement interchangeables (acception post-aristotélicienne de la synonymie), mais parce qu'ils sont commensurables. Le nom et le lógos 'animal' sont le mètre commun à 'homme' et à 'bœuf', il permet donc d'évaluer ou de mesurer leur différence de signification (*ibid.* : 61)

13. L'article de Garcia-Debanc *et alii* (2009) porte plus précisément sur les verbes de déplacement.

De ce fait, avec la synonymie d'Aristote, deux mots ou deux choses sont rendus comparables par un même mètre étalon, qui leur est commun. Les rapprochements entre deux mots s'effectuent à l'aide d'un troisième terme : c'est exactement le cas d'*homme* et *bœuf* avec être *animé*. Le troisième terme servant d'intermédiaire, il n'est nullement question de considérer les deux autres mots comme équivalents ou substituables.

Cette définition, vue sous l'angle des mathématiques, permet de repenser la synonymie et invalide toute définition en termes d'équivalence. Les synonymes – ou cohyponymes – ne se substituent pas car ils ne sont pas équivalents. En revanche, ils se rattachent à un même hyperonyme. Cela renvoie à ce que nous évoquions en didactique avec les contraintes inhérentes à la production écrite et à la recherche de « synonymes ».

5. Conclusion

Cet historique de la synonymie donne un aperçu des différentes définitions qui ont pu être données à cette notion. La définition de la synonymie en tant que relation d'équivalence n'est pas la plus opératoire. Bien souvent, cette définition est une version appauvrie des définitions de la synonymie que nous avons rencontrées. La synonymie comme relation d'équivalence n'est que le premier stade d'une réflexion plus poussée sur la signification des unités. Elle n'est donc qu'une construction de la pensée, un artefact, servant à mener une réflexion sur la langue.

Laissons à Georges Kleiber (2009 : 9) le soin de résumer la situation :

Il est ainsi piquant de noter que le but principal des dictionnaires de synonymes n'est pas tellement de montrer l'identité sémantique des lexèmes retenus comme synonymes, mais consiste précisément à mettre au jour leurs différences, si minimes soient-elles. Et de nombreux articles de linguistique, même et surtout dans les recueils consacrés à la synonymie, suivent ce mouvement : l'auteur prend deux ou trois expressions considérées habituellement comme synonymes et déploie ensuite tout son talent de fin sémanticien pour prouver qu'il n'en est finalement rien, que chaque expression a son sens et son fonctionnement propres, qui conduisent à rejeter in fine l'assimilation sémantique « grossière » qui a servi de point de départ

Notre analyse terminographique a permis de changer le regard porté sur la synonymie : elle n'est pas nécessairement pensée comme une relation d'équivalence car les ressemblances de sens sont toujours déconsidérées au profit des différences de sens. La synonymie distinctive détruisant la synonymie-équivalence, elle en fait une impasse. Toutefois, notre analyse est l'occasion de repenser la notion d'équivalence afin de réhabiliter la relation de synonymie. L'équivalence ne doit plus forcément être pensée comme donnant lieu à la substitution. Elle doit être le théâtre des ressemblances de sens, ressemblances notamment regroupées en champs sémantiques. Cadrée par un champ sémantique, la synonymie peut renouer avec la conception aristotélicienne et la cohyponymie. Dès lors, elle prend un tour résolument cognitif puisqu'il s'agit de hiérarchiser le lexique, de constituer des champs sémantiques ou des ontologies dans lesquels pourront évoluer des synonymes. D'ailleurs, Benjamín García-Hernández (1997 : 2) définit les synonymes comme des noms associés par leur voisinage sémantique :

el prefijo griego συν- de la palabra sinónimo dista mucha de implicar esa noción de « igualdad »; al contrario representa, lo mismo que el latín com-, un valor sociativo, que es el que conviene a los sinónimos como « nombres asociados » por su vecindad significativa.

Une approche de la synonymie basée sur les ressemblances de sens, faisant intervenir la cohyponymie et les champs sémantiques, peut réorienter et renouveler le débat sur la synonymie en tant qu'équivalence, en se détachant de la tradition différentielle.

Bibliographie

- ARISTOTE 2002, *Catégories, I*, Présentation, traduction et commentaires de F. Ildefonse et J. Lallot, Paris, Seuil.
- ARUTA STAMPACCHIA (A.) 2006, *Introduction à la théorie des synonymes de Pierre- Benjamin Lafaye*, Paris, Editions Lanore.
- AUROUX (S.) 1984, « D'Alembert et les synonymistes », in *Dix-Huitième siècle*, Paris, PUF, 16, pp.93-108.
- AUROUX (S.) 1986, « La synonymie et la contrainte de la science : Roubaud, 1785 », in *Autour de Féraud. La lexicographie en France de 1762 à 1835*, GEHLF éd., Paris, Collection de l'École Normale Supérieure de Jeunes Filles, 29, pp. 73-81.
- BISCONTI (V.) 2012, « La synonymie dans la seconde moitié du XIX^e siècle : description, pédagogie et théorisation », *Congrès Mondial de Linguistique Française 2012*, pp. 707-721 (www.linguistiquefrancaise.org).
- BLUMENTHAL (P.) 2006, « De la logique des mots à l'analyse de la synonymie », in *Langue française*, Paris, 150, pp. 14-31.
- BOUHOURS (D.) 1671/2003, *Les Entretiens d'Ariste et d'Eugène*, Paris, Honoré Champion.
- BOUHOURS (D.) 1674, *Doutes sur la langue française. Proposez à Messieurs de l'Académie par un gentilhomme de province*, Paris, Sébastien Mabre-Cramoisy.
- BURIDANT (C.) 1980, « Les binômes synonymiques. Esquisse d'une histoire des couples de synonymes du Moyen Âge au XVII^e siècle », in *Bulletin du centre d'analyse du discours*, Lille, 4, pp. 5-79.
- CHEVALIER (J.-C.) 1997, « La synonymie dans les manuels pédagogiques du début de la Renaissance », in *Langages*, Paris, 128, pp. 8-24.
- DESBORDES (F.) 2007, « Homonymie et synonymie d'après les textes théoriques latins », in *Idées grecques et romaines sur le langage. Travaux d'histoire et d'épistémologie*, Paris, ENS Éditions, pp. 307-357.
- GARCIA-DEBANC (C.), DUVIGNAU (K.), DUTRAIT (C.) & GANGNEUX (M.) 2009, « Enseignement du lexique et production écrite. Un travail sur les verbes de déplacement à la fin de l'école primaire », in *Pratiques*, Metz, 141/142, pp. 208-232.
- GARCIA-HERNANDEZ (B.) 1997, « Sinonimia y diferencia de significado », in *Revista Española de Lingüística*, Madrid, 27/1, pp. 1-31.

- GIRARD (G.) 1718, *La Justesse de la langue françoise ou Les différentes significations des mots qui passent pour synonymes*, Paris, Veuve d'Houry.
- GOMPERZ (T.) 2008, *Les Sophistes. Les Penseurs de la Grèce. Histoire de la philosophie antique (Tome I, Livre III, Chapitre V, VI, VII)*, Houilles, Editions Manucius.
- GUIZOT (F.) 1809/1859, *Nouveau Dictionnaire universel des synonymes de la langue française*, Paris, Maradan (1809), Paris, Didier (1859).
- HONESTE (M.-L.) 2007, « Entre ressemblance et différence : synonymie et cognition », in *Le Français moderne*, Paris, 75/1, pp. 160-174.
- KLEIBER (G.) 2009, « La synonymie-“identité de sens” n'est pas un mythe », in *Pratiques*, Metz, 141/142, pp. 9-25.
- LECLERCQ (O.) 2008, « Pratiques synonymiques et régulation du sens lexical aux XVIe et XVIIe siècles », in *Cahiers de lexicologie*, Paris, 92/1, pp. 13-29.
- LO PIPARO (F.) 2000, « Le théorème de Pythagore dans la linguistique grecque », in *Histoire Épistémologie Langage*, Paris, 22/1, pp. 51-67.
- MAROUZEAU (J.-M.) 1921, « Synonymes latins », in *Cinquantenaire de l'EPHE*, Paris, pp. 13-22.
- MOUSSY (C.) 1994, « Avant-propos » in *Les problèmes de la synonymie en latin*, C. Moussy éd., Paris, PUPS, pp. 7-9.
- RASTIER (F.) 2004, « Ontologie(s) », in *Revue des sciences et technologies de l'information*, Paris, 18/1, pp. 15-40.
- SAUSSURE (F.) 1916/1986, *Cours de linguistique générale*, Paris, Payot.
- SIOUFFI (G.) 2008, « Du syntagme au paradigme : autour de la synonymie chez D. Bouhours », in *Cahiers de lexicologie*, Paris, 92/1, pp. 31-50.
- SIOUFFI (G.) 2012, « Les binômes synonymiques et la question de la figure au XVIIe siècle : quelques investigations dans l'usage et dans les remarques », in *La synonymie*, F. Berlan et G. Berthomieu eds, Paris, PUPS, pp. 367-378.
- SWIGGERS (P.) 2008, « L'économie (interne) de la synonymie : valeur et prix des mots chez Girard », in *Cahiers de lexicologie*, Paris, 92/1, pp. 51-68.
- VLASSOV (S.) 2012, « La synonymie distinctive dans les remarques sur la langue française du XVIIe siècle », in *La synonymie*, F. Berlan et G. Berthomieu eds, Paris, PUPS, pp. 379-413.

Unités monolexicales et unités polylexicales : entre équivalence sémantique et synonymie, le cas des adverbes

Lassâad OUESLATI
Institut Supérieur des Sciences Humaines de Jendouba, Tunisie
Lexiques, Dictionnaires, Informatique (Paris 13)

Résumé

Les unités polylexicales relèvent le plus souvent du figement dans la mesure où elles se caractérisent par une certaine globalité sémantique. Plus précisément, elles se caractérisent par leur non-compositionnalité. Cette caractéristique implique que les unités polylexicales se comportent exactement comme les unités monolexicales sur les plans syntaxique et sémantique. Tout comme les lexèmes simples, les séquences polylexicales établissent un réseau de relations sémantiques impliquant l'antonymie, la synonymie, l'hyponymie, etc. Certains linguistes parlent d'équivalence sémantique en esquivant la synonymie. Par exemple, certains lexicologues donnent *faire peur* pour synonyme de *effrayer* ; *avoir l'air* pour synonyme de *sembler* ; *en dépit* pour synonyme de *malgré* ; *panier percé* pour *dépensier*, etc. La synonymie existe donc entre les unités monolexicales et les unités polylexicales, toutes les parties du discours confondues.

Les dictionnaires définissent les unités polylexicales en donnant et les antonymes et les synonymes. Par exemple, dans le *Petit Robert électronique*, on donne *faususement*, *indûment*, *injustement* pour synonymes de *à tort* ; et on donne aussi *inconsidérément* pour synonyme de *à tort et à travers*. Nous nous intéresserons, dans ce travail à

la synonymie entre les adverbes simples et les séquences adverbiales figées. Nous mettrons l'accent sur le degré d'équivalence sémantique entre des unités du type *en vain* et *vainement*.

Cette investigation est de nature à nous permettre de vérifier le degré d'équivalence entre les unités polylexicales et les lexèmes simples. Nous pourrions aussi vérifier si les unités polylexicales servant de paraphrases aux adverbes en *ment* leur servent vraiment de synonymes. La notion d'emploi développée dans les travaux du LDI nous servira d'outil d'analyse pour mesurer le degré de synonymie entre les différentes unités lexicales adverbiales.

Mots-clés :

unités polylexicales adverbiales, synonymie, congruence, globalisation.

1. Introduction

Après avoir été pendant longtemps considérée comme le parent pauvre de l'analyse sémantique, la synonymie semble actuellement regagner du terrain dans les études sémantiques. Les deux publications récentes portant sur cette question, en l'occurrence Berlan et Bouverot (2007) et Berlan et Pruvost (2008), témoignent de ce regain d'intérêt. Dans ces travaux, un constat est fait : la notion de synonymie est problématique à plus d'un titre dans la mesure où, comme le souligne à juste titre G. Petit (2008), *le fait de parler de « synonymie*

partielle » ou de « parasynonymie » et non plus de « synonymie » vient confirmer, dans la dénomination même de la relation, le déficit dont elle est porteuse.

Outre le problème qu'elle pose au niveau de sa dénomination, cette notion pose problème également au niveau du lexique. En effet, si on la définit essentiellement par la substituabilité entre deux mots ou séquences, la question qui se pose est de savoir si l'on peut parler de synonymie entre un lexème simple et une unité polylexicale et *vice versa*. Cette problématique devient d'autant plus cruciale que les séquences polylexicales sont, elles-mêmes, problématiques aux deux niveaux syntaxique et sémantique.

Nous nous proposons d'étudier, dans ce cadre, la synonymie entre les unités polylexicales et les unités monolexicales, et plus particulièrement les séquences adverbiales polylexicales et les adverbes simples. Notre étude s'articulera autour des trois axes suivants. En premier lieu, nous nous interrogerons sur le statut des séquences adverbiales polylexicales et leur fonctionnement syntactico-sémantique. Nous axerons la deuxième partie sur une étude critique de la synonymie entre unités adverbiales polylexicales et unités adverbiales simples, telles qu'elles se présentent dans le Dictionnaire électronique des synonymes de l'Université de Caen. La dernière partie portera sur l'étude de la synonymie de ces unités dans la théorie des classes d'objets.

2. Les unités polylexicales : statut et fonctionnement

2.1. Le caractère problématique de l'unité lexicale et la synonymie

L'une des caractéristiques de la synonymie sur laquelle tous les travaux sont unanimes est celle de la substituabilité. Deux mots ou deux séquences de mots ne sont substituables que lorsqu'ils sont pourvus du même contenu sémantique. Il s'agit à la fois d'un critère syntaxique et sémantique. Du point de vue syntaxique, deux unités ou deux séquences sont substituables quand elles appartiennent toutes les deux à la même catégorie syntaxique. C'est pour cela que, dans certains cas, on parle d'un fonctionnement adjectival, adverbial, nominal ou autre d'une séquence de mots ou même proposition subordonnée. Par exemple, les grammairiens parlent souvent de proposition subordonnée relative adjectivale dans le sens où une telle proposition assure une des fonctions grammaticales assignées à l'adjectif. Sur le plan sémantique, deux mots ou deux séquences de mots ne peuvent être substituables que lorsqu'ils véhiculent le même contenu sémantique. G. Petit (2008) précise à ce propos que : *Deux mots ou séquences n'étant substituables que dans un ou des contextes restreints (et contraints) et non pas dans la totalité de leurs emplois en langue, la portée de la synonymie s'en voit considérablement restreinte et son rendement entaché d'un fort coefficient de déceptivité.*

Il apparaît que la substitution n'est possible qu'entre deux mots ou deux séquences équivalents comme dans les exemples suivants :

Jean-Jacques **fait peur** aux enfants/Jean-Jacques **effraie** les enfants

Outre le test de substitution, le contexte semble être déterminant pour mesurer le degré de synonymie. La question qui se pose à ce propos est de savoir si une séquence polylexicale est substituable par une unité monolexicale et *vice versa*. Pour répondre à cette question, nous sommes amenés à réfléchir sur le statut de l'unité lexicale, c'est-à-dire sur la notion problématique de mot. En d'autres termes, qu'est-ce qui fait qu'une unité polylexicale peut fonctionner exactement comme une unité monolexicale de façon qu'elles deviennent réciproquement substituables, dans une relation lexicale de synonymie ?

Pour répondre à une telle question, on se heurte inéluctablement au problème de la définition du mot. Le statut problématique de l'unité lexicale, en l'occurrence le mot, a fait, en effet, l'objet d'un numéro du *Français moderne* intitulé « La problématique du mot » (2009-1). Il a fait aussi l'objet de nombreux articles dont nous citons particulièrement celui de S. Mejri (2006) intitulé « Polylexicalité, monolexicalité et double articulation du mot ». La définition du *mot* se heurte à deux problèmes principaux. Le premier problème est de nature morphologique. En effet, la définition du mot est d'autant plus difficile que celui-ci peut prendre des formes variables et différentes les unes des autres. Ces formes peuvent aller d'un phonème du type *de*, *y*, etc., jusqu'à prendre la forme d'une unité polylexicale phrastique comme un *suivez-moi-jeune-homme*, en passant par des unités lexicales polymorphématiques du type indissociable ou des unités polylexicales syntagmatiques du type *à tout bout de champ*, *de plein fouet*, *à petites gorgées*, *à pleurer*, etc. La délimitation de ce qu'on appelle *mot* demeure difficile dans ce sens et, comme le souligne à juste titre S. Mejri (2005 : 20), *le mot peut avoir la configuration formelle de n'importe quel segment de la combinatoire linguistique.*

S'ajoute à ce problème de nature morphologique un autre de nature syntaxique. En effet, un mot ne peut être défini comme tel qu'en fonction de sa distribution dans la phrase. Or, le mot peut, selon les cas, avoir différentes distributions lui permettant de se rattacher à une catégorie syntaxique. En d'autres termes, c'est cette distribution qui assure l'unité syntaxique à un mot quelconque, quel que soit son gabarit. S. Mejri (2009 : 71) trouve que la langue est dotée d'un mécanisme profond qui lui permet, en dehors de toute analogie, de verser n'importe quel segment linguistique dans une catégorie d'unité capable de s'insérer tout naturellement dans la combinatoire syntagmatique et syntaxique. Citons à ce propos des syntagmes du type à pleurer, de justesse, à bâtons rompus, à grandes enjambées, chemin faisant, à la mode, etc. qui, ainsi que les constituants du syntagme le confirment, ne sont pas prédestinés à former une unité lexicale.

Ces deux critères de reconnaissance du statut d'une unité lexicale – aussi plurielle soit-elle –, à savoir la morphologie et la syntaxe distributionnelle et combinatoire, font partie d'un principe général qu'est la congruence. Mais à ces deux éléments de la congruence s'ajoute un troisième : la sémantique. Cet élément est le plus important dans ce sens qu'il permet d'actualiser une unité lexicale quelconque. S. Mejri (2009 : 74) précise justement que *le mot n'est plus employé pour lui-même : il est au service d'un contrat sémiotique qui le rattache à une actualité énonciative qui justifie son emploi*. Il en résulte que la notion de congruence est englobante. Elle dispose de plusieurs strates, dont la forme, la combinatoire, le sens et même la pragmatique, puisqu'un mot ne peut être actualisé que dans le discours. L'intérêt de cette notion de congruence réside

dans la congruence d'actualisation qui est, selon S. Mejri (2009 : 74), la synthèse de toutes les autres congruences. Pour lui, c'est cette congruence qui :

est soit de nature syntaxique soit de nature pragmatique. La syntaxe intervient au niveau des contraintes relatives à l'ordre des mots, à l'expression des différentes catégories grammaticales et à l'ancrage de l'énoncé dans un contexte textuel ou extra-textuel. La pragmatique prend en charge la relation entre l'énoncé et tous les paramètres référentiels extralinguistiques : l'impératif, l'interjection, la performativité, la signalétique, etc. sont autant d'expressions de la congruence pragmatique.

Il ressort de cet extrait l'importance du discours dans la définition même de toute unité lexicale. En effet, un mot ne peut former une unité – ayant un fonctionnement unitaire, aussi polylexicale qu'elle soit – que lorsqu'il est employé dans le discours. Sa définition se résume donc en une position et un fonctionnement de la phrase. En d'autres termes, *est considéré comme mot toute unité lexicale pouvant assurer une fonction dans le cadre de la phrase en occupant une position dans la structure et en portant les marques morphosyntaxiques qui lui sont assignées par son appartenance catégorielle* (Mejri 2009 : 78). Dans les exemples suivants :

- (1) La sauce a pris.
- (2) Luc a pris son sac.
- (3) Luc a pris une décision.
- (4) Luc a pris la tangente.

Le verbe *prendre*, comme toute unité lexicale, n'a d'existence qu'après son emploi dans un cadre phrastique, lequel emploi lui assigne sa marque catégorielle verbale et lui confère un sens qui le distingue de tous les autres emplois. Pour s'en convaincre, il suffit

d'opérer un certain nombre de modifications sur sa distribution ou sur ses combinaisons :

- (5) *Luc a pris
- (6) *La sauce a pris son sac
- (7) *La sauce a pris une décision
- (8) *La sauce a pris la tangente

Il faut préciser à la suite de S. Mejri (2009) que si les énoncés (de 6 à 8) ne sont pas acceptables parce qu'on a porté atteinte à la combinatoire de ce verbe – ce qui, par conséquent, influence directement le contenu sémantique et compromet leur acceptabilité. Cette relation étroite entre la combinatoire peut aussi être illustrée par les schémas d'arguments suivants :

- (1') N0 : substance <prendre>, synonyme : durcir, épaissir
- (2') N0 : humain <prendre> N1 : objet concret, synonyme : saisir
- (3') N0 : humain, verbe support : prendre <décision>
- (4') N0 : humain <prendre la tangente>, synonyme : partir

En y regardant de plus près, la différence entre les synonymes proposés est de nature à montrer l'importance de la congruence morphologique, syntaxique, sémantique et pragmatique.

2.2. Le caractère problématique des séquences adverbiales figées

La classe des adverbes est une classe qui suscite beaucoup d'intérêt chez les linguistes. Pour s'en convaincre, il suffit de voir la richesse de la bibliographie dédiée à cette classe grammaticale. Cet

intérêt s'explique non seulement par la polymorphie, mais aussi par des propriétés syntactico-sémantiques qui distinguent les adverbes de toutes les autres parties du discours. L'une des propriétés morphologiques de cette classe grammaticale réside dans le fait que l'adverbe, contrairement aux autres parties du discours, telles que l'adjectif, le nom, le verbe, etc., ne dispose que d'un seul suffixe pour former des adverbes. Il suffit d'ajouter le suffixe en *ment* à une base adjectivale ou autre pour obtenir un adverbe. De plus, l'adjonction d'un tel suffixe n'est pas possible pour tous les adjectifs. Par exemple, les adjectifs relationnels du type *tunisien* ou les adjectifs de couleur comme *jaune*, *rouge*, etc., échappent à la règle de la formation des adverbes en *ment*. D'ailleurs, il est aisé de constater que les adjectifs qui résistent à cette règle morpho-syntaxique qu'est l'adverbialisation sont de loin plus nombreux que les adjectifs réguliers. Pour compenser un tel déficit morphologique, la langue a recours au recyclage que présente le figement lexical. En effet, la formation des unités polylexicales adverbiales permet de remédier à ces lacunes morphologiques.

Les locutions adverbiales présentent, outre la morphologie – puisqu'elles n'ont pas de schème unique —, des problèmes de nature syntaxique. De nombreux adverbiaux sont syntaxiquement contraints ou changent de sens en changeant de position. Certaines séquences subissent un transfert catégoriel leur permettant de passer d'une catégorie syntaxique à une autre. Par exemple, la séquence polylexicale à *califourchon* peut fonctionner comme un adjectival : *Il est à califourchon* comme elle peut avoir un fonctionnement adverbial *monter à califourchon*.

Sur le plan sémantique, les adverbiaux posent un autre problème. Nous savons que la polysémie est de règle pour les unités lexicales simples, toutes les catégories syntaxiques confondues. Les adverbes ne dérogent pas à la règle : les adverbes simples, notamment les adverbes en ment, se caractérisent par leur polysémie, alors que les séquences adverbiales sont le plus souvent polysémiques. Prenons, par exemple, l'adverbe *honnêtement* et l'adverbial *de bonne foi*. Considérons les exemples suivants :

- (9) Il gagne honnêtement sa vie./syn. correctement
- (10) Il se conduit honnêtement avec les autres./syn. civilement, poliment
- (11) La pensée est ici honnêtement exprimée./syn. convenablement
- (12) Il se comporta honnêtement/syn. civilement, poliment
- (13) Il est honnêtement logé/syn. correctement
- (14) Il est honnêtement payé/ syn. suffisamment, correctement

L'adverbe *honnêtement* change de sens en fonction du prédicat sur lequel il porte. Ce changement sémantique se manifeste à travers le choix du synonyme adéquat que l'on trouve à la suite de chaque énoncé. Si on choisit *correctement* comme synonyme de cet adverbe dans la phrase (9), on choisit *civilement* comme synonyme du même adverbe dans la phrase (12). Pour vérifier si la synonymie est partielle ou totale, il suffit de substituer un adverbe à son synonyme. Par exemple, il suffit de remplacer *honnêtement* dans la phrase (9) par *civilement* :

- (15) *Il gagne sa vie civilement
- (16) * Il est civilement payé

La pluralité des synonymes témoigne de la polysémie de cet adverbe. Or, une unité polylexicale adverbiale synonyme de *honnêtement* n'admet qu'un seul emploi, sachant qu'un emploi correspond entre autres à un seul sens. Ainsi, cet adverbial ne peut se substituer à *honnêtement* que dans quelques cas :

- (17) *Il gagne de bonne foi sa vie.
- (18) Il se conduit de bonne foi avec les autres.
- (19) ?? La pensée est ici exprimée de bonne foi.
- (20) Il se comporta de bonne foi.
- (21) *Il est logé de bonne foi.
- (22) *Il est de bonne foi payé.

Là nous constatons une dissymétrie entre l'adverbe monolexical et l'adverbe polylexical, une dissymétrie qui semble intimement liée au fonctionnement de la polysémie. S. Mejri (2005) souligne à ce propos que *le fonctionnement de la polysémie se fonde sur une dissymétrie qui fait que l'économie s'exprime par le biais d'une surcharge du côté des significations apportées par le même signifiant*. Parallèlement, la polylexicalité réduit la polysémie. L'opposition entre unité polylexicale et unité monolexicale n'est pas réductible à la morphologie (c'est-à-dire la pluralité ou non des signifiants). Elle la dépasse pour toucher le contenu sémantique. Cette opposition est soulignée justement par S. Mejri (2003 : 21) qui soutient que *c'est souvent l'inverse qui se constate avec les SF puisque le même principe de dissymétrie est en action dans la structuration de ce type d'unités, mais, contrairement aux unités unilexicales, avec une pluralité du signifiant*. Même si certaines unités polylexicales peuvent être polysémiques, la polysémie est loin d'être systématique. Cette situation

est imputée à des raisons qui semblent relever de l'équilibre général du système (Mejri 2003 : 22).

En y regardant de plus près, la synonymie est sélective dans le cas des unités monolexicales : chaque unité polysémique doit avoir un synonyme pour chaque signifié. Rappelons, à ce propos, le cas de *honnêtement*. Nous sommes ainsi amenés à nous interroger sur la façon dont on traite la synonymie en lexicographie.

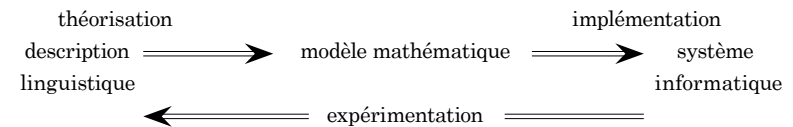
3. Étude critique de la synonymie entre UAdvP et UAdvM dans le Dictionnaire des Synonymes de Caen

3.1. Fondements théoriques du Dictionnaire des synonymes de l'Université de Caen

Le *Dictionnaire électronique des synonymes* (désormais DES) constitue la synthèse de deux dictionnaires de synonymes du XIX^e siècle — à savoir ceux de Guizot et Lafaye — et de trois dictionnaires de synonymes du XX^e siècle — Bailly, Bénac et Bertaud Du Chazaud —, auxquels on ajoute les synonymes retenus par les deux dictionnaires analogiques : *Le Grand Larousse de la langue française* et le *Grand Robert*. J. François et J.-L. Manguin (2004 : 2) précisent que les entrées aussi bien que les synonymes *proviennent de la fusion des sept fichiers contenant les relations présentes dans les dictionnaires qui ont servi de base à cette construction, et qui nous ont été fournies par l'Institut national de la langue française*. Le DES n'est donc autre

qu'un fichier dans lequel chaque ligne contient un mot-vedette suivi de la liste de ses synonymes (Manguin 2004 : 2). Les 49 000 mots-vedettes qu'il compte sont reliés les uns aux autres par un réseau de relations de synonymie. Selon François et Manguin (2004 : 1), on dénombre 19 9000 relations pour tout le dictionnaire. Ainsi ont été fusionnées toutes les listes de synonymes que présentaient les différents dictionnaires pour n'en faire qu'une seule liste, qui est devenue le DES.

Pour bien comprendre ce qu'est la synonymie dans le DES, rappelons avant toute chose le fondement théorique de ce dictionnaire. Celui-ci a pour fondement théorique la théorie de la construction dynamique du sens de Victorri et Fuchs (1996). Il constitue en effet une des applications possibles de cette théorie. Il s'agit même d'un *moyen efficace de mettre en œuvre un protocole expérimental de la théorie* (Victorri et Fuchs 1996 : 88). Pour concevoir cette théorie, Fuchs et Victorri se sont servis des outils mathématiques ; après quoi ils ont implémenté cette théorie à l'aide d'un système informatique. Les étapes de cette démarche peuvent être résumées comme suit¹ :



1. Ce schéma a été emprunté à G. Doualan (2011 : 41).

Pour résumer, nous pouvons dire que le DES n'est autre qu'une application de la théorie générale de la construction dynamique du sens, laquelle théorie est déductive dans ce sens où elle formule des hypothèses. L'ensemble de ces hypothèses compose cette théorie. Pour que celle-ci soit exploitable, il est nécessaire qu'elle soit formalisée selon un modèle mathématique. Une fois formalisée, cette théorie peut être implémentée. Le système informatique sert de moyen de validation. Les auteurs de cette théorie précisent : *Nous ne faisons pas de modélisation informatique à proprement parler, nous ne faisons que retranscrire sous forme informatique un modèle élaboré par ailleurs* (Victorri et Fuchs 1996 : 88). Pour eux, cette implémentation constitue *le seul moyen de pouvoir traiter des exemples concrets, susceptibles de valider ou d'invalidier le modèle mathématique* (Victorri et Fuchs 1996 : 88).

Étant issu du modèle des réseaux connexionnistes, lequel modèle permet d'implémenter des phénomènes de réseaux, le DES est avant tout un réseau de connexions. En effet, les mots-vedettes sont des points à partir desquels s'établissent des connexions avec d'autres points par un jeu de relations synonymiques. Ainsi ce dictionnaire se présente comme un vaste réseau dans la mesure où les mots-vedettes sont souvent polysémiques. Des liens de synonymie s'établissent donc entre les différents sens. Nous sommes ainsi amenés à examiner le traitement réservé à la synonymie dans le DES.

3.2. Le traitement de la synonymie dans le DES

L'avertissement dans l'interface du DES définit la synonymie comme suit :

Deux unités lexicales sont, en pratique, toujours partiellement synonymiques ; cela signifie que c'est seulement dans un contexte donné que l'on peut remplacer l'une par l'autre sans modifier notablement le sens de l'énoncé. Cela implique alors qu'un autre synonyme ne conviendrait pas forcément à ce contexte.

Cette définition fait donc écho aux différentes définitions de ce concept, lesquelles définitions insistent systématiquement sur la partialité de cette relation d'équivalence sémantique. Des synonymes proposés ne correspondent qu'à certains contextes. Ces contextes changent en fonction de la situation de communication, du niveau de langue, etc. C'est pour cette raison que les auteurs du DES affirment dans l'*Avertissement* que *certaines mots ont ainsi des synonymes à connotations péjorative, injurieuse ou raciste qui ne peuvent évidemment s'employer dans un contexte neutre, mais qui apparaissent néanmoins ici, puisque ce dictionnaire est censé être le reflet de la langue dans ses divers emplois*. Ainsi apparaît l'objectif des auteurs : fournir un dictionnaire de synonymes exhaustif.

En dépit de l'exhaustivité dont se réclament les auteurs, le DES ne tient pas compte des nuances que présente tel ou tel synonyme. Cette décision de sacrifier les nuances s'explique par l'analyse mathématique des relations synonymiques. Dans ce sens, les auteurs avancent que *l'originalité de [leur] dictionnaire réside dans l'obtention des différences de sens par une analyse mathématique des relations synonymiques ; c'est la raison pour laquelle [ils ont] volontairement éliminé des résultats présentés les nuances d'emploi que l'on trouve dans les dictionnaires traditionnels*.

En consultant le DES, nous pouvons nous rendre compte de sa conception de la synonymie. En effet, ce dictionnaire, comme nous l'avons déjà souligné, constitue une fusion entre plusieurs dictionnaires. Il constitue de ce fait un dictionnaire cumulatif. Bien qu'il reprenne les dictionnaires sources, le DES comporte des entrées ajoutées manuellement. Par contre, il n'a pas retenu toutes les entrées des dictionnaires sources. Il est ainsi considéré à la fois comme compilatoire et cumulatif. Il se présente en effet comme une entrée suivie de ses synonymes. Lors de la confection de ce dictionnaire, on a procédé à la sélection des données retenues. Manguin (2005 : 2) explique cette démarche en affirmant qu'il était impossible de conserver les distinctions et divisions d'articles faites dans certains dictionnaires, en raison de l'hétérogénéité des classifications opérées suivant les lexicographes. Pour des raisons informatiques, on a fait du DES une liste hétérogène des synonymes, ce qui en fait un dictionnaire cumulatif. Par conséquent, ce dictionnaire ne fournit ni explication ni définition. Il se trouve ainsi dépourvu de toute précision syntaxique. Les auteurs, conscients de cette pauvreté de la description syntaxique, la qualifient de sècheresse. Ils disent à ce propos que *cette sècheresse apparente est compensée par le nombre élevé d'entrées et par la grande rapidité d'obtention des résultats* (François et Manguin 2004 : 2). Le manque d'explications est pallié par un hyperlien avec le TLFi.

Divers constats se dégagent de cette présentation rapide :

- La synonymie est essentiellement synchronique. Le cumul des synonymes dans le DES à partir des dictionnaires appartenant à plusieurs diachronies fait que les synonymes se présentent de façon hétérogène. Ils se présentent en vrac selon le dire de Manguin (2005 : 2).

- La dimension syntaxique n'est pas prise en considération, ce qui fait que la consultation du DES donne lieu à un grand nombre de synonymes sans avoir la possibilité de distinguer un emploi d'un autre.
- La morphologie n'est pas prise en considération. Deux mots synonymes peuvent avoir la même racine prédicative.
- La relation de synonymie est, pour les auteurs de DES, une relation d'équivalence sémantique. Or la synonymie est plus précise que l'équivalence. En effet, l'équivalence, prenant un sens plus large, plus simple à manier, moins contraint, *peut donc recevoir un grand nombre de réalités* (Doualan 2011). D'ailleurs, dans ce dictionnaire, la synonymie s'étend de l'équivalence sémantique stricte à l'analogie. Doualan (2011 : 9) souligne le fait qu'*avec l'analogie, ce n'est pas seulement la notion de synonymie en elle-même qui se trouve étendue, mais aussi ce que renferment les dictionnaires dits de synonymes. Ils ne se limitent plus seulement à des synonymes, ils poussent leur compilation jusqu'à l'analogie.*
- Dans le DES, on parle d'espace sémantique. La synonymie est par conséquent une intersection entre au moins deux espaces sémantiques. Cela nous permet de comprendre le fonctionnement de ce dictionnaire : à une UAdvP correspondent plusieurs adverbes monolexicaux ou polylexicaux. Deux séquences adverbiales synonymes ont nécessairement un espace sémantique en commun. L'importance de cet espace dépend du nombre de cliques². Par exemple, à l'adverbe à tâtons correspondent les adverbes suivants : à l'aveugle, à l'aveuglette, aveuglément. Alors qu'à adverbe du type à pas comptés correspond un seul synonyme adverbial, *lentement*.

2. Ou plus précisément cliques de graphes, concept mathématique, appliqué à l'étude de la polysémie, que B. Victorri définit comme les sous-ensembles de synonymes du mot qui sont tous des synonymes entre eux deux à deux. Pour plus de précisions, le lecteur se référera à Victorri B., (2002), « Espaces sémantiques et représentation du sens », in *Textualités et nouvelles technologies*, écartS, 3.

4. La synonymie entre UAdvP et UAdvM dans la théorie des classes d'objets

4.1. Principes d'analyse

La théorie des classes d'objets est une théorie lexicaliste qui consiste à décrire le lexique d'une langue afin de confectionner un dictionnaire de traitement automatique de cette langue. On procède ainsi à une description intégrée dans la mesure où l'on ne se contente pas de décrire le sens d'une unité lexicale, mais où l'on décrit également sa morphologie et sa syntaxe combinatoire, sachant que les trois niveaux, à savoir la forme, la construction et le sens d'une unité lexicale quelconque sont intimement liés. Il en résulte que l'unité minimale d'analyse, ce n'est plus le mot, mais la phrase. L'unité n'a donc d'existence que dans une phrase élémentaire. Comme chaque phrase comporte des unités, chaque unité assure une fonction. Dans une phrase élémentaire, l'unité lexicale peut assurer l'une des trois fonctions élémentaires possibles : soit une fonction prédicative, soit une fonction argumentale, soit une fonction actualisatrice.

Abstraction faite des parties du discours traditionnellement établies, la théorie des classes d'objets distingue les unités lexicales selon leur fonction dans la phrase. De plus, la dimension sémantique est inséparable du lien qu'entretiennent les unités entre elles. Par exemple, il suffit de changer l'argument sujet pour opérer une modification profonde du contenu sémantique du prédicat. La mention systématique des synonymes dans les différents travaux du LDI met l'accent sur l'importance de cette relation entre la syntaxe, la morphologie et la sémantique. Prenons l'exemple de l'adjectif *âpre*. Cet adjectif

change de sens en fonction de son sujet comme le montrent les schémas d'arguments suivants :

âpre/N0 : <objet concret>/Sy : rugueux
 âpre/N0 : <lieu géologique>/Sy : escarpé, raboteux
 âpre/N0 : aliment<fruit>/Sy : acide, âcre
 âpre/N0 : <dispute>/Sy : violent, rude
 âpre/N0 : <propos>/Sy : véhément
 âpre/N0 : <combat>/Sy : violent, sanglant, acharné, sauvage
 âpre/N0 : <phénomène météo>/Sy : rude, rigoureux

On comprend à travers ces exemples que le schéma d'argument qui relève de la syntaxe permet d'identifier, avec plus de précision, le synonyme correspondant. Tous ces éléments relèvent de ce que G. Gross (2004) appelle *emploi*. Un emploi d'une unité est la somme de toutes ces données syntaxiques, morphologiques et sémantiques. Pour illustrer cette idée, citons les exemples suivants des adjectifs *aimable* et *agressif* que nous empruntons à G. Gross :

aimable/N0:
 hum/N1:avec hum/S: affable, avenant, /A:désagréable, /D:lg
 aimable/N0:
 <comportement>/N1:avec hum/S:accueillant, /A:désagréable, /D:lg
 aimable/N0:
 <propos>/N1:/S:amène, courtois /A:désagréable, discourtois/D:lg
 aimable/N0:
 loc/N1:/S:agréable/A:désagréable/D:lg
 agressif/N0:
 hum/N1:avec hum/S:menaçant, violent/A:doux, inoffensif/D:lg
 agressif/N0:
 <pays>/N1:/S:menaçant, belliqueux/A:pacifique/D:lg

agressif/N0:
<caractère>/N1:/S:violent/A:doux, calme/D:lg
agressif/N0:
<comportement>/N1:avec hum/S:violent, menaçant/A:doux/D:lg
agressif/N0:
<propos>/N1:/S:violent, menaçant/A:apaisant, calme/D:lg
agressif/N0:
in <couleur>/N1:/S:criard/A:/D:lg

4.2. Les synonymes dans la description des séquences adverbiales figées selon la théorie des classes d'objets

La description des adverbes simples ou figés ne déroge pas à la règle. En effet, toutes les UAdvP sont décrites comme toutes les unités lexicales : elles ont une forme, une construction et un sens. Elles assurent, en plus, l'une des trois fonctions. Elles peuvent avoir une fonction prédicative élémentaire comme dans l'exemple suivant :

(23) Léa lui a répondu **de bonne foi**.

La valeur prédicative apparaît à travers les différentes restructurations suivantes :

La réponse de Léa à lui a été de bonne foi.
Léa a été de bonne foi en lui répondant

Ces unités polylexicales adverbiales peuvent avoir également une fonction actualisatrice comme dans :

(24) Il fréquente ce bistrot de moins en moins. (L'itératif)
(25) Il conduisait à fond de train. (L'intensité)

L'actualisation à travers la valeur aspectuelle que confère l'emploi de la séquence adverbiale au prédicat. Dans l'exemple 24, l'UAdvP exprime l'itératif alors que dans l'exemple 25, à *fond de train* exprime l'intensité. Rappelons au passage que l'actualisation consiste à inscrire un prédicat dans le temps en indiquant dans certains cas sa valeur aspectuelle ou en l'inscrivant dans une échelle de gradation, à travers l'expression de l'intensité.

Les adverbiaux peuvent avoir aussi une fonction argumentale telle qu'on la voit dans l'exemple suivant :

(26) Il a réparé cette machine manuellement.

La restructuration de cette phrase permet de mettre l'accent sur l'emploi argumental de l'adverbe *manuellement* :

La réparation de la machine, il l'a faite à la main
Il a réparé la machine avec la main.

Outre la fonction primaire que joue une locution adverbiale, celle-ci peut être un classifieur dans la mesure où son emploi dans une phrase sert à rattacher le prédicat à une classe sémantique bien déterminée. Ce rôle est manifeste dans l'emploi des adverbes d'intensité. Prenons les exemples suivants :

(27) Il conduisait à **fond de train**.
(28) Il crie à **plein gosier**.

Si l'adverbial à *fond de train* est essentiellement un actualisateur dans la mesure où il indique l'intensité du prédicat, en l'occurrence la vitesse, il assure également une fonction dans ce sens qu'il rattache le prédicat *conduire* à la classe sémantique des prédicats de <déplacement humain par moyen de transport>. Bien qu'il indique la vitesse du déplacement d'un humain, il ne peut être synonyme d'un adverbial du type à *grandes enjambées* dans la mesure où ce dernier actualise des prédicats appartenant *exclusivement* à la classe sémantique des prédicats de <déplacement humain>.

De même, l'adverbial à *plein gosier* est un actualisateur de prédicat. Il indique en effet le degré d'intensité d'un cri humain. Son emploi permet de repérer tous les prédicats compatibles, ce qui permet d'élaborer une classe que l'on peut appeler la classe sémantique des prédicats des <cris humains>, laquelle classe regroupe des verbes prédictifs comme *chanter, hurler, crier, parler*.

Il est clair que dans le traitement automatique des langues, la description des unités lexicales revêt une importance capitale. Cette description doit donc déboucher sur une élaboration précise de chaque classe sémantique, tâche qui facilitera le repérage d'un synonyme adéquat. Pour mettre l'accent sur le rôle que joue l'adverbe dans l'élaboration de ces classes sémantiques prenons les exemples suivants :

- (29) Elle se défendait **bec et ongles**.
- (30) Elle s'accrochait **bec et ongles**.
- (31) Un busard avait accroché son poignet **bec et ongles**
- (32) Tout cela est réglé **au quart de tour**.
- (33) La fille a réagi **au quart de tour**.
- (34) Elle répondit **au quart de tour**, la bête blessée, la bête fidèle.

Le point commun entre ces deux adverbiaux *bec et ongles* et *au quart de tour*, c'est qu'ils appartiennent à l'hyperclasse des <intensifieurs>. Mais se contenter de dire qu'ils sont des adverbiaux d'intensité n'aide pas à trouver le synonyme adéquat. Par exemple, dans le DES, à l'adverbe de degré très correspondent quarante-trois adverbes monolexicaux et polylexicaux considérés comme synonymes de type à *l'excès, absolument, assez, bigrement, diablement, drôlement, effroyablement, énormément, excessivement, extraordinairement, follement, formidablement, fort, fortement, horriblement, infiniment, joliment, merveilleusement, pleinement, prodigieusement, remarquablement, richement, rudement, sérieusement, terriblement, etc.*

En y regardant de plus près, on constate que, s'il est vrai que ces adverbes indiquent l'intensité, ils n'intensifient pas, pour ainsi dire, toutes les classes de prédicats. Ils peuvent être classés en fonction des classes de prédicats dont ils indiquent le degré d'intensité, tâche qui réduirait la marge d'erreur. Pour s'en convaincre, il suffit de substituer un adverbe d'intensité à un autre :

- (29) *Elle se défendait **joliment**.
- (30) *Elle s'accrochait **énormément**.
- (31) * Un busard avait accroché son poignet **à l'excès**.
- (32) * Tout cela est réglé **infiniment**.
- (33) *La fille a réagi **richement**.
- (34) *Elle répondit **diablement**, la bête blessée, la bête fidèle.

L'examen de ces exemples montre bien l'importance de la description du lexique en termes de classes d'objets. Elle permet en effet d'identifier l'attirance qu'exercent certains adverbiaux actualisateurs

envers certains prédicats. Cependant, il faut remarquer que, compte tenu de la notion de globalisation qui concerne toutes les séquences polylexicales dont nous avons déjà parlé plus haut, les séquences adverbiales expriment, le plus souvent, des nuances bien précises qu'un adjectif simple ne peut exprimer, raison pour laquelle les séquences adverbiales polylexicales ne disposent vraiment pas de synonymes. En effet, des adverbes du type à *plein gosier*, *bec et ongles* expriment des nuances que d'autres adverbes considérés comme synonymes semblent incapables d'exprimer fidèlement.

Prenons les exemples à *tombeau ouvert* et à *toute vitesse* et *rapidement*. Ces trois adverbes ont un point commun : ils expriment la vitesse. Sur ce plan, ils sont présentés dans les dictionnaires comme des synonymes. Seulement, la substitution, critère syntaxique de synonymie, est loin d'être évidente. Observons les exemples suivants que nous avons pris du TLFi :

- (35) Il a expédié son travail à toute vitesse.
- (36) Le Ménénil galopait à tombeau ouvert.
- (37) Les deux voitures démarrent et s'éloignent rapidement.

La substitution ne semble pas aller de soi dans tous les cas :

- (35) Il a expédié son travail à toute vitesse/*à tombeau ouvert/
rapidement
- (36) Le Ménénil galopait à tombeau ouvert/à toute vitesse/rapidement
- (37) Les deux voitures démarrent et s'éloignent rapidement/
à tombeau ouvert/à toute vitesse.

Il apparaît à travers ces différentes restructurations que les deux adverbes à *toute vitesse* et *rapidement* sont plus proches sur le plan sémantique que de à *tombeau ouvert*. D'ailleurs, ce dernier ne disposant d'un synonyme monolexical précis, les dictionnaires tels que le TLFi ont recours à la paraphrase suivante « à une vitesse telle que l'on risque la mort ». Cette paraphrase est la preuve que la nuance exprimée par à *tombeau ouvert* ne peut être exprimée par aucune autre unité polylexicale adverbiale. Dans le cas de cet adverbial, il s'agit d'un classifieur qui ne sélectionne que des prédicats de < déplacement humain par moyen de transport : animal ou mécanique >, du type *conduire*, *galoper*, *rouler*, etc. Il faut souligner à ce propos que le nombre de prédicats sélectionnés par cet adverbial est très réduit. Ceci n'est pas le cas pour les deux autres adverbiaux.

Cependant, à *toute vitesse* et *rapidement* ne sont que partiellement synonymes. Leur proximité sémantique, dont nous avons parlé dans le paragraphe précédent, s'explique par le fait qu'à *toute vitesse* est plus polysémique que à *tombeau ouvert* dans la mesure où cette séquence polylexicale adverbiale est moins figée. Ce figement relatif apparaît à travers sa lecture compositionnelle. Si cette transparence la rend compatible avec plus de prédicats que à *tombeau ouvert*, cette séquence sélectionne toutefois moins de prédicats que *rapidement*. En d'autres termes, l'adjectif monolexical *rapidement* est plus polysémique que les deux autres dans la mesure où il peut être compatible avec plusieurs classes de prédicats. Illustrons cette idée par les exemples suivants :

- (38) Sans doute il est heureux de juger rapidement, mais il importe de bien juger.
 (39) Il commente rapidement cet événement.

Bien que *à toute vitesse* soit synonyme de *rapidement*, il ne parait pas compatible avec un prédicat du type *juger* ou *commenter* :

- (38) ?? Sans doute il est heureux de juger à toute vitesse, mais il importe de bien juger
 (39) ?? Il commente à toute vitesse cet événement.

Ces exemples montrent bien que la synonymie n'est pas une relation d'équivalence sémantique transitive ni réciproque. En effet, les adverbiaux ont tendance à exprimer des nuances que les adverbes simples n'expriment que partiellement. D'ailleurs, notre corpus nous permet de constater que plus une unité polylexicale adverbiale est figée, plus elle est appropriée à une classe bien précise de prédicats, moins elle a de synonymes et vice versa. D'ailleurs, certaines séquences adverbiales figées ne disposent d'aucun synonyme. Tel est le cas de *à tombeau ouvert* et celui de *à plein gosier*, employé dans la phrase (28).

5. Conclusion

À travers cette étude, nous avons essayé de montrer que la notion de synonymie demeure problématique notamment dans le cas des unités polylexicales adverbiales. En effet, cette classe d'unités exprime souvent des nuances que les adverbes simples ne peuvent

exprimer. Pour trouver le synonyme adéquat, il faut identifier avant tout la fonction qu'assure la séquence adverbiale. C'est-à-dire qu'il nécessaire de préciser si la séquence en question est prédicative, actualisatrice ou argumentale. Le synonyme doit, outre le contenu sémantique, assurer la même fonction. Un travail sur un large corpus est de nature à identifier les synonymes adéquats de chaque séquence adverbiale.

Bibliographie

- BAILLY (R.), 1947, *Dictionnaire des synonymes*, Paris, Larousse.
- BÉNAC (H.), 1956, *Dictionnaire des synonymes*, Paris, Hachette.
- BERLAN (F.) et PRUVOST (J.), dir. 2008, « Synonymie et lexicographie (XVI^e-XXI^e siècles) », in *Les Cahiers de lexicologie*, Paris, n° 92-2008, p. 5.
- BERLAN (F.) et BOUVEROT (D.), dir. 2007, « La synonymie en questions : échanges entre les époques », in *Le français moderne*, Paris, 2007-1.
- DOUALAN (G.) 2011, « Introduction à une approche instrumentée de la synonymie, l'exemple du Dictionnaire électronique des synonymes du CRISCO », in *Cahiers du CRISCO*, Rouen, n° 32, septembre 2011, p. 9.
- FRANÇOIS (J.) et MANGUIN (J.-L.), 2004, « Le dictionnaire électronique des synonymes du CRISCO », in *Cahiers du CRISCO*, Rouen, n° 17.
- FUCHS (C.) 1994, *Paraphrase et énonciation*, Paris, Ophrys.
- GIRY-SCHNEIDER (J.) 2005, « Les adjectifs intensifs : syntaxe et sémantique », in *Cahiers de lexicologie*, Paris, 86/1, pp. 163-178.
- GROSS (G.) 2009, *Sémantique de la cause*, Leuven-Paris, Peeters.
- GUILBERT (L.) et alii, 1971, *Grand Larousse de la langue française*, Paris, Larousse.
- IBRAHIM (A.-H.) et ZALESSKY (M.) 1989, « Lexiques », *Le français dans le monde*, numéro spécial, Paris, p. 206.
- JULIA (C.) 1990, « Fixer le sens la sémantique spontanée des gloses de spécification du sens, la sémantique », in *Langages, La synonymie*, Paris, n° 128, pp. 91-104. pp. 13-22.
- MANGUIN (J.-L.) 2004, « Transitivity partielle de la synonymie : application aux dictionnaires de synonymes », in *Corela*, Poitiers, 2/2 (corela.edel.univ-poitiers.fr).
- MANGUIN (J.-L.), TIEDEMANN (J.) et VAN DER PLAS (L.) 2008, « Extraction de synonymes à partir d'un corpus multilingue aligné », revue électronique *Texte et corpus*, n° 3, *Actes des Journées de la linguistique de corpus 2007*, pp. 151-161.
- MANGUIN (J.-L.), 2005, « Les dictionnaires en ligne : nouvelles diffusions, nouveaux objectifs », colloque SIF 2005, Paris, 12-13 décembre 2005 (<http://edutice.archives-ouvertes.fr/edutice-00001422>).
- MARTIN (R.) 1990, « La définition "naturelle" », in J. Chaurand et F. Mazière, eds, *La définition*, Paris, Larousse, pp. 86-96.
- MARTIN (R.) 1983, *Pour une logique du sens*, Paris, PUF, pp. 209-221.
- MEJRI (S.) 2009, « Le mot, la problématique théorique », in *Le français moderne*, Paris, n° 1, pp. 68-82.
- MEJRI (S.) 2006, « Polylexicalité, monolexicalité et double articulation : la problématique du mot », in *Cahiers de lexicologie*, Paris, n° 89, 2006-2, pp.
- MEJRI (S.), 2005, « Introduction: polysémie et polylexicalité », in S. Mejri et J. François, dir., *Syntaxe et sémantique: Polysémie et polylexicalité*, n° 5, Caen, pp. 13-30.
- NAZARENKO (A.) 2005, « Sur quelle sémantique reposent les méthodes automatiques d'accès partir d'un corpus multilingue aligné », revue électronique *Sémantique et corpus*, A. Condamines éd., Paris, Lavoisier, pp. 211-239.
- OUESLATI (L.) 2012, « Relations transphrastiques, codage et adverbe », dans *Relations, connexions, dépendances, Hommage au professeur Claude Guimier*, N. Le Querler, F. Neveu et E. Roussel, Rennes, Presses Universitaires de Rennes, pp. 85-100.
- OUESLATI (L.) 2011, « L'actualisation des prédicats : le cas des adverbes indicateurs d'aspect », dans S. Mejri et J. Bacha, *Le traitement du lexique, catégorisation et actualisation*, Sousse, Presses universitaires de Sousse, pp. 256-280.
- OUESLATI (L.) 2006, *Les constructions adverbiales en français contemporain*, thèse de doctorat soutenue à Paris 13.
- PETIT (G.) 2008, « Le Dictionnaire des Synonymes de Condillac, de la synonymie à la cohyonymie », in *Cahiers de Lexicologie*, Paris, n° 92, pp. 87-120.
- POTTIER (B.) 1992, *Sémantique générale*, Paris, PUF.
- PRANDI (M.) 1987, *Sémantique du contresens. Essai sur la forme interne du contenu des phrases*, Paris, Minuit.
- REY (A.) 1997, « Polysémie du terme définition », in *La définition*, I. Chamrand et F. Mazière eds, Paris, Larousse, pp. 13-22.
- REY-DEBOVE (J.) 2009, « La synonymie ou les échanges de signes comme fondement de la sémantique », in *Langages*, Paris, n° 128, pp. 91-104.
- TAMBA-MECZ (I.) 1981, *Le sens figuré. Vers une théorie de l'énonciation figurative*, Paris, PUF.
- VENANT (F.) 2009, « Géométriser le sens lexical », *Quatrième conférence internationale sur la Théorie Sens-Texte (MTT'09)*, Montréal, en ligne : <http://hal.archives-ouvertes.fr/hal-00429219/> (dernière consultation : 6/5/2013).

VICTORRI (B.) et FUCHS (C.) 1996, *La polysémie : construction dynamique du sens*, Paris, Hermès.

VICTORRI (B.) 1994, « La construction dynamique du sens », in *Passions des formes*, M. Porte éd., Lyon, ENS, pp. 733-747.

La traduction linguistique vers l'arabe : la métalangue polylexicale comme exemple

Abdellatif CHEKIR
Université de Gabès

Résumé

Nous constatons aujourd'hui un foisonnement de termes polylexicaux dans la linguistique française pour véhiculer des phénomènes et des concepts de plus en plus complexes ; ce qui traduit une évolution manifeste de ce domaine de connaissance. Toutefois ces termes présentent des structures variées en français et posent énormément de difficultés au moment de leur transfert en arabe. Certes, certaines séquences restent proches du patron de départ, mais d'autres s'en écartent sur le plan morphosyntaxique. D'autres encore semblent inadéquates parce qu'elles ne reflètent pas le fonctionnement du phénomène de départ et plus particulièrement celles qui désignent des phénomènes idiosyncrasiques. Toutefois la traduction des ouvrages de linguistique est d'un grand apport : elle permet, à travers ce regard croisé, d'affiner la description de la langue arabe et même de découvrir des phénomènes insoupçonnés dans notre langue.

Mots-clés :

métalangue, terminologie, équivalence, polylexicalité, contrastif, traduction.

1. Introduction

Nous assistons de nos jours à un grand mouvement de traduction des textes fondamentaux dans le domaine de la linguistique. Cette action vise à transférer les connaissances, à les adapter à la langue arabe et *a posteriori* à élaborer un dictionnaire de linguistique en arabe. Or, cette traduction achoppe sur plusieurs écueils et plus particulièrement au niveau de la métalangue grammaticale puisqu'il s'agit d'un transfert complexe qui doit tenir compte et du terme et de la charge conceptuelle qu'il véhicule. Certes, la linguistique recourt à des termes simples, mais nous constatons un usage croissant des termes polylexicaux, ce qui constitue un des indices de l'évolution des sciences du langage. L'objectif est de contribuer à la réflexion épistémologique sur les difficultés de transfert du savoir linguistique de L1 vers L2 ainsi que sur la nature des métalangues grammaticales à travers un regard croisé sur deux systèmes linguistiques aussi différents que l'arabe et le français.

Nous allons étudier cette problématique à partir d'un corpus de termes relevés dans les ouvrages de linguistique suivants :

- *Comprendre la linguistique*, de Robert Martin, traduit par Abdelkader Mhiri ;
- *Pour une logique du sens*, de Robert Martin, traduit par Taieb Baccouche et Salah Mejri ;
- *Les expressions figées en français. Noms composés et autres locutions*, de Gaston Gross, traduit par Salah Mejri et Béchir Ouerhani ;
- *Le dictionnaire des sciences du langage*, de Franck Neveu, traduit par Salah Mejri.

Notre travail s'articule autour de trois axes. Nous procéderons tout d'abord à une typologie des termes polylexicaux ; ce qui nous permettra par la suite d'observer le degré de fidélité des termes équivalents en arabe et les problèmes qu'ils soulèvent. Nous réfléchirons enfin sur l'apport de la traduction et, plus particulièrement, la traduction des termes polylexicaux dans le développement de la linguistique arabe.

2. Typologie des termes polylexicaux

Les séquences polylexicales métalinguistiques sont en concurrence avec les termes simples, mais elles sont de plus en plus fréquentes dans les travaux de linguistique. Elles constituent un des procédés de néologie disponibles dans la langue qui répond au foisonnement des termes forgés par les différentes théories linguistiques modernes. Elles traduisent une tendance de la linguistique moderne vers une démarche descriptive et véhiculent des concepts et des phénomènes linguistiques de plus en plus complexes. Elles ont la même structure interne que les noms composés utilisés dans le langage courant en arabe et en français. Elles se présentent sous les formes suivantes en français :

- Nom + adjectif : force allocutaire, tournure perceptive, prédication secondaire, verbes symétriques, tournure factitive, logique plurivalente, idée regardante, etc.
- Nom + déterminant + nom : acte de langage, adjectif de relation, classes d'objets, ligne de pertinence, modalité d'action, relatif de liaison, image d'univers, champ de dispersion, champ d'application, etc.
- Nom + nom : lexique grammairal, article zéro, etc.

- Nom + *en* + désinence : forme en *-ant*, forme en *-rais*, etc.
- Deux éléments de différentes classes séparés par un trait d'union : *question-écho*, *pseudo-suffixal*, *supra-segmental*, *hétéro-univers*, *sous-classe*, *sous-jacent*, *sous-codage*, *non-dit*, *non-insertion*, etc.

Nous constatons que ces termes polylexicaux n'épuisent pas toutes les possibilités des séquences nominales et les procédés de la composition disponibles en général dans la langue française. Toutefois les types les plus productifs sont ceux qui adoptent les patrons nom + adjectif et nom + déterminant + nom.

Ces termes polylexicaux sont en général des séquences endocentriques composées d'un substantif-tête qui constitue le pivot et des éléments de détermination ou des modificateurs. En effet, dans les séquences : *adjectif qualificatif* et *adjectif de relation* le substantif-tête est adjectif alors que les éléments de détermination sont *qualificatif* et *de détermination*. Ces deux termes polylexicaux renvoient ainsi à des types particuliers d'adjectifs qui se distinguent par leur fonctionnement. Cependant, nous n'avons pas relevé dans notre corpus des séquences exocentriques non construites autour d'un substantif-tête et qui se caractérisent par leur opacité comme dans les expressions figées du type de *cordons bleus* qui n'est pas un cordon ou de *paniers percés* qui n'est pas un panier. Autrement dit, la métalangue grammaticale polylexicale ne dispose pas d'expressions figées pour dénommer les phénomènes linguistiques comme dans la langue commune.

Nous pouvons également constater que sur le plan sémantique, *les termes complexes polylexicaux présentent deux éléments de nomination principaux* (Neveu 2005 : 114). En effet, les termes *tour*

attributive, *tour* *perceptive*, *tour* *factitive* sont formés chacun de deux éléments de nomination et véhiculent ainsi plusieurs traits conceptuels (*op. cit.* : 110). Cependant un des éléments permet de les distinguer les uns par rapport aux autres. Nous sommes, en effet, en présence de *tour* *ures* de différents types qui sont discriminées par les classificateurs : *attributive*, *perceptive*, *factitive*.

Mais malgré l'existence de plusieurs traits de dénomination, nous pouvons constater que ces expressions polylexicales fonctionnent comme des unités unilexicales : *L'unilexicalité est une caractéristique syntactico-sémantique propre aux unités lexicales, qu'elles soient polylexicales ou monolexicales, par laquelle les unités fonctionnent d'une manière globale comme une seule unité lexicale dont l'autonomie a pour corollaire la globalité de sens et de fonctionnement syntaxique* (Mejri 2006 : 214). En effet, ces séquences polylexicales métalinguistiques se distinguent par leur sens global et sont ainsi considérées comme une seule unité lexicale puisqu'elles renvoient à un seul phénomène linguistique. C'est l'exemple de *subjonctif présent* ou *passé* qui réfèrent chacun à un tiroir verbal bien particulier. D'ailleurs ces expressions polylexicales peuvent parfois avoir pour équivalent une unité monolexicale. C'est le cas de *conditionnel* qui peut être désigné par *forme en -rais*, ou *aspect* qui équivaut dans des cas particuliers à *modalité d'action*.

Il faut tenir compte de tous ces paramètres lors du transfert de ces unités terminologiques en arabe.

2. Fidélité et écart

Le français et l'arabe appartiennent à deux familles de langues différentes : le français est une langue indo-européenne alors que l'arabe est une langue chamito-sémitique. Ces deux familles se distinguent par plusieurs caractéristiques et plus particulièrement au niveau morphosyntaxique. Le français est une langue concaténative, elle recourt, à l'instar des langues indo-européennes, à la dérivation par affixation et à la composition comme procédés de créativité lexicale. Alors que l'arabe est une langue non concaténative : elle use de procédés de dérivation et de composition spécifiques pour créer des unités nouvelles. En effet, la racine consonantique est versée dans des moules selon des schèmes qui ont un sens préétabli pour dériver toutes les catégories verbales ou nominales. Toutefois concernant la composition, la langue arabe offre des procédés parfois similaires et d'autres fois différents par rapport à la langue française. Il faut également signaler que la langue arabe dispose d'un appareillage métalinguistique consacré par la tradition grammaticale dans lequel les traducteurs linguistiques puisent souvent pour transférer des termes monolexicaux mais rarement dans le cas des termes polylexicaux utilisés fréquemment pour désigner des concepts inédits. Mais, dans l'acte de transfert des expressions polylexicales en arabe, certaines restent fidèles alors que d'autres s'écartent du modèle de départ.

2.1. Expressions fidèles

Dans leur souci de trouver un équivalent arabe à chaque terme, les traducteurs-linguistes élaborent des unités polylexicales parallèles, qui respectent le patron du terme de départ.

C'est le cas de la structure : nom + adjectif :

ʔafʕa:l mutana:ðira (verbes symétriques)
 ʔafʕa:l taxji:rijja (verbes transformationnels)
 tarki:b ʔiaba:ri (tournure attributive)

Ces équivalents arabes sont formés d'un verbe et d'un nom à l'image des termes français.

Les traducteurs-linguistes empruntent même un procédé de composition étranger à la langue arabe à savoir les deux éléments séparés par un trait d'union. C'est l'exemple de :

aa:ri3-lisa:ni (extra-linguistique);
 da:ail- lisa:ni (intra-linguistique);
 fibh -ʔilħa:qi (pseudo-suffixal);
 suʔa:l-sada (question-écho);
 maʕa-taʔli:fi (para-synthétique).

Cette adoption d'un procédé de composition français dans le transfert en arabe permet au traducteur de rester le plus proche possible du modèle de départ.

Toutefois nous remarquons que certains termes sont hybrides : ils résultent de l'association d'un terme emprunté au français et d'un adjectif arabe. C'est l'exemple de :

tajmama qawijja (thématisation forte) ;
tajmama ḍaʔi:fa (thématisation faible).

Ces termes sont composés de l'élément français *thème* à partir duquel on dérive le déverbal (masdar) *tajmama* puis on lui ajoute le classifieur *qawijja* ou *ḍaʔi:fa* .

Contrairement à l'époque des premières traductions, nous constatons que le nombre d'emprunts est très réduit. Mais si le terme *thème* a été emprunté c'est parce qu'il se caractérise par sa légèreté, son économie et sa facilité d'adaptation au moule arabe. D'ailleurs *thème* a servi à dériver d'autres termes tels que :

Tajmi: (thématique) ;
Tajmijjan (thématiquement).

Ce qui permet de former un paradigme de termes de différentes catégories utilisables, selon les exigences, dans les différents moments du développement de l'argumentation.

Dans tous ces cas de figure, nous sommes en présence de termes qui résultent de la traduction littérale, du calque. En effet, les traducteurs s'efforcent de reproduire les patrons des termes de départ surtout lorsqu'il s'agit de phénomènes inédits, introduits récemment dans la langue arabe à la faveur de la traduction des ouvrages fondateurs de la linguistique.

Mais nous remarquons que certaines unités polylexicales trahissent le modèle de départ.

2.2. Expressions infidèles

Dans le transfert vers l'arabe, nous remarquons parfois un changement de la classe grammaticale des éléments ou, parfois, un changement du nombre de lexèmes. Mais dans tous ces cas de figure, les traducteurs cherchent à adapter les termes au moule de la langue arabe sans pour autant défigurer leur charge conceptuelle.

Certains équivalents arabes se caractérisent par un écart par rapport au modèle de départ par la disparition d'un élément. La structure nom + préposition + nom qui devient nom + nom avec la disparition de la préposition. Nous pouvons le vérifier à travers les exemples suivants :

ʔihtima:lu lwuzu:di (présomption d'existence)
ʔiftira:ḍu ssidq (présomption de sincérité)
sifatul kam (adjectif de relation)
sifatul ʔala:qa (adjectif de relation)
ʔasna:ful ʔajja:ʔ (classe d'objets)

Il s'agit ici d'une traduction conforme aux spécificités de l'arabe, langue dans laquelle le complément du nom est d'un emploi direct, il n'est pas précédé d'une préposition comme en français.

L'écart s'exprime également par l'augmentation du nombre de lexèmes comme dans les termes suivants :

aita:b vajr muba:fir hor (discours indirect libre)
 mana:tiq mutaʿaddidatul qi:ma (logiques plurivalentes)
 ʿala:matu ʿiʿra:b ʿal fiʿl (désinence verbale)
 qawwa mutaḍammīna fil fiʿl (force allocutaire)
 ʿiftira :ḍ moʿbaq la: ʿasa:lahu (présupposé non rempli)

À travers ces exemples nous remarquons qu'un élément de l'expression de départ obtenu par affixation a été fragmenté en deux lexèmes en L2. Si nous observons la séquence discours indirect libre et son équivalent aita:b vajr muba:fir hor, nous constatons que l'augmentation a concerné le correspondant du lexème indirect qui est formé du préfixe *in* et de la base *direct*. Or ce procédé de dérivation spécifique au français n'est pas disponible dans la langue arabe. C'est pour cette raison qu'il a été traduit par la préposition exprimant la négation et l'adjectif muba:fir. Et c'est pratiquement la même démarche qui a été suivie pour traiter les autres exemples.

Dans d'autres cas très rares, nous assistons au procédé contraire, à savoir la réduction du nombre des unités lors du transfert vers L2. C'est l'exemple de :

naht (mot-valise)
 tahti (sous-jacent)

Certaines expressions se caractérisent par un changement de la classe grammaticale du lexème. Cet écart est perceptible à travers les expressions suivantes :

su:ra maʿza:zijja (figure de trope)
 muhi:t moʿtaqadi (univers de croyance)
 musa:ʿid si:xi (auxiliaire de mode)
 taʿoi:r maʿnawi (effet de sens)

Dans ce cas de figure, nous remarquons que le complément déterminatif devient un adjectif. Par contre dans les exemples suivants la préposition devient un adjectif :

tasni:f farʿi (sous-catégorisation)
 sinf farʿi (sous-classe)
 taqni:n taḥti (sous-codage).

Comme nous pouvons le constater, à travers ces occurrences, le contenu de la préposition *sous* est repris par l'adjectif placé après le nom. Mais nous constatons que dans les deux premiers exemples *sous* a pour équivalent l'adjectif *farʿi*, alors que dans le dernier exemple il a été traduit par *taḥti*. Toutefois, la traduction littérale avec une préposition préfixée est possible. Mais si les traducteurs ont opté pour la forme avec adjectif postposé, c'est qu'elle permet aisément de dériver le verbe et l'adjectif comme l'a expliqué Béchir Ouehani (à paraître) :

verbe: qannana taḥtijjan (sous-coder) ;
 adjectif: muqannan taḥti (sous-codé)

2.3. Termes problématiques

Certains équivalents semblent malheureux parce qu'ils présentent une combinaison un peu particulière. Ce qui peut être perceptible à travers les expressions suivantes :

sifa wa:sifa (adjectif qualificatif)
 ma:ḍi sa:baq (passé antérieur)
 moʿtaqbal sa:biq (futur antérieur)

À travers les deux premières séquences nous constatons que les éléments constitutifs sont pléonastiques. En effet, nous pouvons noter que l'adjectif *wa:sifa* équivalent de qualificatif n'est en fait qu'un dérivé du nom *sifa*. Mais si les traducteurs ont opté pour ce choix, c'est pour le distinguer des autres types d'adjectif comme l'adjectif de relation traduit littéralement par *sifatul ʕala:qa*. Il en est de même pour *ma:ði sa:baq* car les deux éléments constitutifs sont presque des synonymes. Ils indiquent ce qui est passé. Par contre, dans le dernier exemple, les deux éléments semblent antonymiques. En effet, *mostaqbal sa:biq* est un calque de futur antérieur. *mostaqbal* signifie « futur » et *sa:biq* « passé ».

Dans ces deux derniers cas, la difficulté provient du fait que nous sommes en présence de phénomènes spécifiques de la langue française, non disponibles dans la langue arabe.

Parfois la difficulté provient du fait que l'équivalent ne constitue pas une traduction littérale car le terme est désigné par une partie de son fonctionnement. Les traductions de certains tiroirs verbaux l'illustrent bien :

ma:ði tta7li:fi (passé simple) littéralement *le passé global* ;
ma:ði tta:m (plus-que-parfait) littéralement *le passé accompli*.

Dans ces deux exemples, la désignation reprend une partie du fonctionnement du tiroir verbal. Dans le cas du passé simple, les traducteurs ont retenu la valeur de globalité alors que dans le cas du plus-que-parfait ils se sont focalisés sur l'aspect accompli. Comme

nous l'avons expliqué précédemment la difficulté provient du fait que ces tiroirs verbaux ne sont pas disponibles dans la langue arabe.

Quelques exemples montrent par ailleurs une confusion entre les phénomènes linguistiques. Tel est le cas des termes :

si:yatu ððami:r (diathèse pronominale)
si:yatul la:ððami:r (diathèse impersonnelle)

À travers ces deux séquences, nous pouvons noter une confusion entre pronom et personne qui ont été traduits par *ðami:r* ce qui peut induire en erreur parce que nous pouvons comprendre que la diathèse pronominale et la diathèse impersonnelle sont deux phénomènes antithétiques. Or les termes *pronom* et *personne* désignent en français deux phénomènes différents qui n'ont pas la même fonction. D'ailleurs, *pronom* et *ðami:r* n'ont pas le même fonctionnement dans les deux langues car conformément aux spécificités de la langue arabe, l'expression du pronom devant le verbe n'est pas obligatoire. C'est qui est exprimé par le terme *ðami:r* qui signifie « ce qui est caché, ce qui n'est pas exprimé ».

Comme nous pouvons le constater à travers ces termes problématiques, nous sommes en présence de phénomènes idiosyncrasiques qu'il est difficile de transférer d'une langue à l'autre. Mais malgré ces cas problématiques, la traduction des termes polylexicaux qui sont le fruit de l'évolution des sciences du langage, la traduction peut enrichir la réflexion sur la langue cible.

3. L'apport de la traduction dans l'affinement de la description de la langue arabe

La langue arabe a été langue grammatisée, dotée de grammaires et de dictionnaires depuis le moyen âge. Cependant ce regard croisé nous permet de découvrir des phénomènes jusque-là insoupçonnés dans notre langue et, partant, de décrire avec plus de précision certains phénomènes linguistiques. En effet, nous constatons un affinement de la description de la langue arabe très manifeste dans les écrits récents qui se démarquent par rapport à la description traditionnelle de la langue.

3.1. Distinction entre conjonction de coordination et préposition (ʔa:da:t ʔatf et harf maʔna)

La tradition grammaticale arabe logeait à la même enseigne la préposition et la conjonction et les désignait par le terme *harf* (particule), l'une des trois parties du discours en plus du nom et du verbe. La traduction linguistique qui a permis de confronter les deux langues et de décrire le fonctionnement des deux phénomènes linguistiques en arabe a fini par les distinguer en *ʔa:da:t ʔatf* (conjonction de coordination) et *harf maʔna* (préposition).

3.2. La distinction entre proposition et phrase

La tradition grammaticale ne distingue pas la proposition de la phrase : les deux sont logées à la même enseigne et sont désignées par le terme *phrase* (ʔomla). Or la traduction a permis de confronter les

deux systèmes linguistiques et de discriminer les deux phénomènes. ʔomla s'est spécialisé dans la désignation de phrase alors qu'on a créé le terme ʔumajla pour dénommer la proposition. Nous parlons ainsi de ʔumajla *tatmi:mijja* (proposition complétive), de ʔumajla *mawsula* (proposition relative), ʔumajla *raʔi:sjja* (proposition principale).

3.3. La locution verbale (ʔiba:ra fiʔlijja)

Le concept de *ʔiba:ra fiʔlijja* (locution verbale) n'était pas opératoire en arabe. Les grammairiens ne parlaient que de « verbe simple » parce qu'ils ne concevaient aucune forme composée du verbe. Cependant, la traduction des textes linguistiques français et plus particulièrement des *Expressions figées* de Gaston Gross (2006) nous a permis de constater l'existence de cette séquence verbale en arabe. Certes, les grammairiens arabes avaient pris conscience de l'existence des expressions figées en arabe, mais aucun n'a parlé des locutions en général et de la locution verbale en particulier.

3.4. Verbe auxiliaire fiʔl musaʔid

Comme nous l'avons vu concernant la locution verbale, la grammaire arabe n'avait pas soupçonné l'existence de l'auxiliaire. Certes, la tradition grammaticale a parlé des différents types de verbes : les abrogatifs, les verbes d'aspect, etc., mais aucune distinction n'était faite entre le verbe et l'auxiliaire. C'est à travers ce regard croisé qu'offre la traduction que nous avons constaté l'existence de l'auxiliaire et que nous l'avons distingué du verbe support, du verbe prédicatif, du verbe causatif, etc., puisqu'ils actualisent le prédicat de façons différentes.

3.5. Le verbe support

La notion de « verbe support » se présente comme un outil d'analyse dans le cadre de l'étude de la prédication d'une manière générale et de la distinction entre prédication verbale et non verbale plus précisément (Ouerhani 2004 : 114). Ce phénomène a été repéré par Maurice Gross (1981 : 17) et Gaston Gross (1996 : 72) dans leurs travaux sur la prédication. Certains linguistes arabes ayant eu connaissance de ces travaux et de la théorie lexicale-grammaire, ont essayé de l'introduire en arabe. Mais c'est Béchir Ouerhani qui lui a consacré tout un ouvrage et a élaboré un corpus de plus de deux mille occurrences. Cependant, les linguistes ne sont pas d'accord sur l'étiquette à adopter pour sa dénomination. Certains parlent de fiʕl reki:za de fiʕl ʕima:da ou de fiʕl ha:mi, alors que Béchir Ouerhani recourt à fiʕl na:qil. Bref ce phénomène est en passe d'être adapté par les linguistes arabes pour devenir un outil consacré grâce à la traduction qui a permis de le déceler.

Ce sont autant d'exemples emblématiques de l'affinement de la description de la langue arabe favorisée par la traduction des textes fondamentaux de la linguistique et l'adaptation des théories linguistiques récentes et des nouvelles méthodes d'approche de la langue par l'arabe.

4. Conclusion

La traduction des textes linguistiques permet de transférer les connaissances vers l'arabe et, partant, d'enrichir la réflexion sur la langue arabe surtout lorsqu'il s'agit de traductions ciblées. La traduction permet également le renouvellement et l'affinement de la métalangue arabe ; ce qui lui permet de véhiculer les idées linguistiques avec beaucoup de précision. Mais les termes restent toujours approximatifs puisqu'ils ne reprennent jamais tout le contenu conceptuel du terme de départ. La traduction linguistique permet également de procéder à un regard croisé entre les deux langues et favorise la fonction heuristique qui nous permet de découvrir des phénomènes jusque-là insoupçonnés dans notre langue.

Bibliographie

GROSS (G.) 1996, *Les expressions figées en français : noms composés et autres locutions*, Paris, Ophrys.

GROSS (M.) 1981, « Les bases empiriques de la notion de prédicat sémantique », in *Langages* n° 63, Paris, p. 7-52.

LAZARD (G.) 1999, « Pour une terminologie rigoureuse : quelques principes et propositions », in A. Le Maréchal (dir.), *La Terminologie linguistique, Mémoires de la Société linguistique de Paris*, Paris, n.s. VI, pp. 11-49.

MEJRI (S.) 2006, « Polylexicalité, monolexicalité et double articulation : la problématique du mot », in *Cahiers de lexicologie*, 89.

MEJRI (S.), THOIRON (Ph.) éd.s 2005, *La terminologie, entre traduction et bilinguisme, Journées scientifiques de formation et d'animation régionale*, Hammamet (Tunisie) le 14 octobre 2004, Agence Universitaire de la Francophonie, Réseau Lexicologie, Terminologie, Traduction & Rencontres Linguistiques méditerranéennes, Tunis, Artypo.

NEVEU (F.) 2005, « Sur l'usage des termes complexes dans le discours de la science du langage. Préliminaire à une étude comparée de la terminologie linguistique », in S. Mejri et Ph. Thoiron, *op. cit.*, pp. 107-120.

OUERHANI (B.) à paraître, *La terminologie linguistique : problématique du transfert des termes polylexicaux (du français vers l'arabe)*.

OUERHANI (B.) 2005, « La terminologie linguistique entre innovation lexicale et renouvellement des analyses », in S. Mejri et Ph. Thoiron, *op. cit.*, pp. 121-138.

SWIGGERS (P.) 1999, « Pour une systématique de la terminologie linguistique : considérations historiographiques, méthodologiques et épistémologiques », in A. Le-maréchal (dir.), *La Terminologie linguistique, Mémoires de la Société linguistique de Paris*, Paris, n.s. VI, pp. 11-49.

Aux frontières du lexique

Bases lexicales multilingues : traitement des acronymes

Ying ZHANG
Université Joseph Fourier
Mathieu MANGEOT
Université de Savoie
*Groupe d'étude en traduction et traitement
automatisé des langues et de la parole*

Résumé

La gestion des terminologies pose encore des problèmes, en particulier pour la gestion des situations compliquées comme les acronymes. Dans cet article, nous proposons une solution en reliant à un seul référent plusieurs termes différents via les notions de pivot et de prolexème. Ces notions permettent par exemple de faire le lien entre plusieurs termes qui désignent un même et unique référent : Nations Unies, ONU, Organisation des Nations Unies et onusien. Il existe, une plate-forme générique de gestion de bases lexicales, Jibiki, permettant de gérer n'importe quel type de structure (macro et microstructure). Nous avons implémenté une nouvelle macrostructure de ProAxie dans la plate-forme Jibiki pour réaliser les gestions des acronymes.

Mots-clés :

base lexicale multilingue, macrostructure, Jibiki, acronyme, *Common Dictionary Markup*, entrepôt de données linguistiques.

1. Introduction

Cet article concerne la gestion de terminologies multilingues. Le problème abordé dans cet article est celui de l'association de plusieurs termes d'une même langue à un même référent. Par exemple, un terme et son acronyme désignent le même référent. Dans un contexte multilingue, la difficulté est d'établir une correspondance entre ces termes. Il peut, par exemple, y avoir dans une même langue plusieurs acronymes pour un même terme, ou aucun, à l'inverse. Quelles solutions mettre en place de façon à choisir, pour un terme donné dans une langue donnée, le meilleur équivalent dans une langue cible ? Cette recherche est motivée par un besoin réel d'une entreprise dans la gestion de sa terminologie multilingue.

L'intérêt de cette recherche est de définir un cadre théorique composé d'une nouvelle macrostructure basée sur des concepts existants et sur la définition de nouveaux concepts. Ce cadre sera validé ensuite par une expérimentation pratique à l'aide d'un outil générique de gestion de bases lexicales.

Cet article est organisé de la façon suivante. Dans la section 2, nous présentons les problèmes et les difficultés de gestion des terminologies riches. Dans la section 3, nous présentons les macrostructures préconisées pour les données. La section 4 présente les outils utilisés et l'implémentation de la macrostructure. La section 5 présente les résultats de l'implémentation et de l'utilisation. Enfin, nous concluons et donnons quelques perspectives pour la gestion des terminologies riches.

2. Problématique

Les liens entre les termes sont compliqués. Il existe la possibilité de lier plusieurs termes différents à un seul référent : « Jean-Paul II » et « Karol Jozef Wojtyla » en français, ou en anglais « John Paul II » et « Karol Jozef Wojtyla ». De même, certains liens évoluent avec le temps : le pape désignait « Jean-Paul II » en 2004 et « Benoît XVI » en 2012. Des pays parlant la même langue (p. ex : France et Suisse romande) peuvent également utiliser des mots différents pour le même concept. Par exemple, « chien renifleur » et « chien drogue ». Inversement, le même terme peut désigner des concepts différents : dans la province de langue allemande de Bolzano en Italie, le « Landeshauptmann » est le président du conseil provincial, avec des compétences beaucoup plus limitées que le « Landeshauptmann » autrichien, qui est à la tête de l'un des États (Land) de la fédération autrichienne.

Pour la gestion des acronymes, les liens riches sont plus répandus et plus complexes. Dans l'article introduisant la notion de *prolexème* (Tran 2006), il a été discuté qu'il existe des lexies ayant des acronymes dans certaines langues, mais pas dans d'autres. Par exemple, en français, il y a « Organisation des Nations unies »¹, on peut aussi dire « Nations unies », « ONU » ou « onusien ». En anglais, on a « United Nations » et son acronyme « UN ». En chinois, on a « 联合国 » qui est la seule lexie pour ce sens, et il n'y a pas d'acronyme. Quand on construit des ressources lexicales, comment créer les liens entre les acronymes et les mots de définition ?

Dans les chapitres suivants, nous explorerons diverses pistes pour résoudre les problèmes décrits ci-dessus.

3. Données : choix de la macrostructure

Lors de toute discussion scientifique, il est primordial de bien s'entendre sur les termes utilisés. C'est pourquoi nous commencerons par définir les termes et concepts principaux que nous utiliserons par la suite.

Un dictionnaire est composé d'un ou plusieurs volumes reliés entre eux par des liens qui sont le plus souvent des liens de traduction. Un volume est un ensemble d'articles comportant des mots-vedettes de la même langue. Un article comporte au moins un mot-vedette et le plus souvent d'autres informations (prononciation, classe grammaticale, définition, exemples, etc.). La structure des articles est appelée *microstructure*. L'organisation des volumes qui compose la structure d'un dictionnaire est appelée *macrostructure*. La macrostructure la plus simple est celle d'un dictionnaire monolingue ne comportant qu'un seul volume. Pour les dictionnaires bilingues langue A \leftrightarrow langue B, on peut trouver des macrostructures avec deux volumes : un volume langue A \rightarrow langue B et un volume miroir langue B \rightarrow langue A. Ces macrostructures constituent l'essentiel des dictionnaires imprimés. L'avènement de l'outil électronique permet de s'abstraire des contraintes liées à l'impression, notamment la représentation restreinte à deux dimensions. Il est alors possible de concevoir des

macrostructures plus complexes utilisant par exemple des volumes pivots. Le dictionnaire devient alors une base lexicale à plusieurs dimensions d'où il est ensuite possible d'extraire des vues spécifiques permettant de retrouver le format initial des dictionnaires imprimés.

Nous détaillerons trois macrostructures de ce type par la suite.

3.1. Macrostructure pivot

Le projet Papillon lancé en 2000 (Tomokyo *et al.* 2001), a pour but de construire une ressource lexicale pour plusieurs langues dont au moins l'anglais, le français et le japonais. Les macrostructures bilingues traditionnelles obligeant à construire un dictionnaire par couple de langues, le nombre de dictionnaires croît de manière factorielle par rapport au nombre de langues en présence. Cette solution devient rapidement ingérable. Il fallait donc en trouver une nouvelle, un dictionnaire multilingue à structure pivot : un volume monolingue pour chaque langue et un volume pivot (ou volume interlingue) au centre regroupant les liens entre les articles (Sérasset et Mangeot 2001). La microstructure des articles monolingues reprend le concept de lexie défini dans la lexicographie explicative et combinatoire (Mel'čuk, Clas et Polguère 1995) issue de la théorie Sens-Texte. Chaque article décrit une lexie. Une lexie est une unité lexicale (ou sens de mot) qui est représentée soit par un lexème (regroupement de mots-formes), soit par une locution nominale.

Chaque lexie est reliée par un lien interlingue à une axie (ou acception interlingue). Les axes sont contenues dans le volume pivot.

Chaque axie regroupe les équivalents dans plusieurs langues d'une même lexie (ou sens de mot).

Les concepts d'axie et de structure pivot ont été définis pour le projet Papillon et ensuite repris dans la norme *Lexical Markup Framework* (Francopoulo *et al.* 2009).

3.2. Macrostructure Pivax

La macrostructure pivot vue précédemment permet de résoudre le problème posé par une situation multilingue. Au niveau monolingue, par contre, on doit se contenter d'un seul volume pour chaque langue, ce qui constitue une limitation importante. Dans sa thèse, Hong-Thai Nguyen (2009) avait le projet de construire une base lexicale avec des ressources issues de systèmes de traduction. Chaque système de traduction utilise un format qui lui est propre pour ses dictionnaires, lui permettant de stocker des variables spécifiques. Si l'on veut regrouper ces dictionnaires, il n'est pas souhaitable de fusionner tous les dictionnaires de chaque langue. On peut par contre regrouper les lexies identiques d'une même langue dans un même objet. C'est le rôle des axèmes, ou acceptions monolingues.

Pour chaque langue ou espace lexical, on aura alors un volume pour chaque système de traduction présent (Systran, Ariane, etc.) et un seul volume d'axèmes. Chaque lexie de chaque volume est reliée à un axème. Chaque axème est à son tour relié à une axie dans le volume pivot central.

Cette structure ainsi constituée est une structure pivax ou structure pivot à étages (Mangeot et Nguyen 2009). Elle peut être utilisée également pour gérer différentes versions d'un même volume dans un espace lexical.

3.3. Macrostructure ProAxie

Cette macrostructure a pour but de résoudre le problème du lien à établir entre plusieurs termes qui désignent un même et unique référent. Pour les gestions des acronymes, les liens riches sont plus répandus et plus complexes. Dans l'article introduisant la notion de prolexème, il a été discuté qu'il existe des lexies ayant des acronymes dans certaines langues, mais pas dans d'autres. Donc, la macrostructure de la gestion des acronymes devra nous permettre de

- gérer les liens plus complexes et
- gérer les multiples volumes dans la même langue.

Pour implémenter la gestion des acronymes, nous avons proposé une nouvelle macrostructure avec deux notions : proaxie et prolexème (Tran 2006). Nous avons également besoin des concepts d'axie (Tomokyo *et al.* 2001) et de lexie (Mel'čuk, Clas et Polguère 1995).

3.3.1. Prolexème

Il y a un seul volume de prolexèmes pour chaque langue. Dans ce volume, les prolexèmes regroupent les lexies. Les liens bidirectionnels entre les lexies et leurs prolexèmes sont marqués avec une

étiquette (alias, acronyme, définition, etc.). Un sens lexical est une entrée de prolexème qui lie les différentes lexies de ce même sens. Par exemple, l'entrée de prolexème est « fr.organisation_des_nations_unies.1 » qui lie l'entrée des lexies :

- « ONU » avec une étiquette « acronyme »
- « Nations unies » avec une étiquette « alias »
- « onusien » avec une étiquette « dérivée »
- « Organisation des Nations unies » avec une étiquette « définition ».

3.3.2. Proaxie

Il y a un seul volume de proaxies dans un dictionnaire. Les proaxies regroupent les prolexèmes de langues différentes dans un même sens. Par exemple, dans un dictionnaire trilingue : français, anglais et chinois. L'entrée de proaxie « proaxie.united_nations.1 » relie

- l'entrée « fra.organisation_des_nations_unies.1 » du volume des prolexèmes français,
- l'entrée « eng.united_nations.1 » du volume des prolexèmes anglais,
- l'entrée « zho.联合国.1 » du volume des prolexèmes chinois.

Les liens entre l'entrée de proaxie et les entrées de prolexèmes sont bidirectionnels.

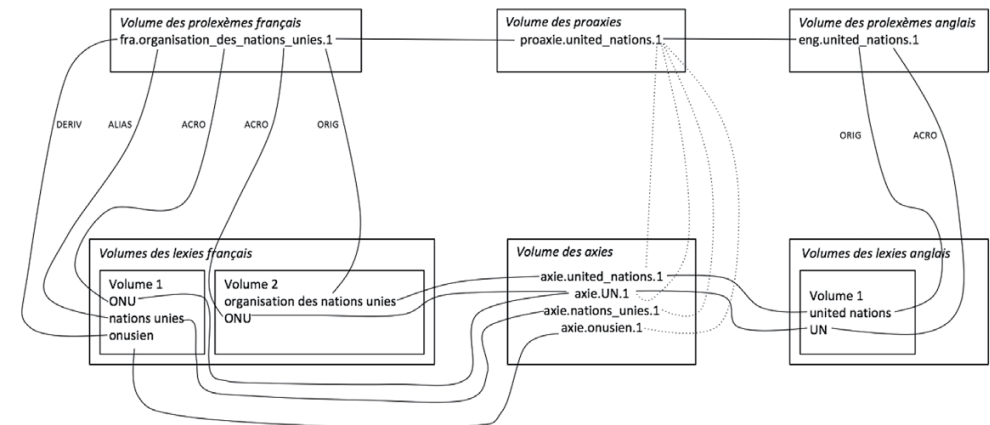
3.3.3. Conception globale

Dans cette macrostructure, nous avons deux couches : une couche basique et une couche « Pro ». Dans la couche basique, nous

avons deux types de volume : volume des lexies et volume des axes. Dans la couche « Pro », nous avons également deux types de volume : volume des prolexèmes et volume des proaxies.

Grâce au volume d'axes, nous pouvons faire les liens entre les lexies qui correspondent exactement, comme l'acronyme français « ONU » relié avec l'acronyme anglais « UN ». Grâce à la couche « Pro », nous pouvons proposer en traduction des lexies des langues cibles de même sens. Voir la figure 1 - « Exemple de ProAxie ».

FIGURE 1 : EXEMPLE DE PROAXIE



La notion d'étiquette a pour but de proposer les meilleures traductions. Par exemple, le japonais « 国際連合 » est la lexie de même sens qu'« Organisation des Nations unies », son acronyme est « 国連 » comme l'indique la table 1 - « Exemple de quatre langues pour le sens « Organisation des Nations unies » ». Cet acronyme utilise le premier

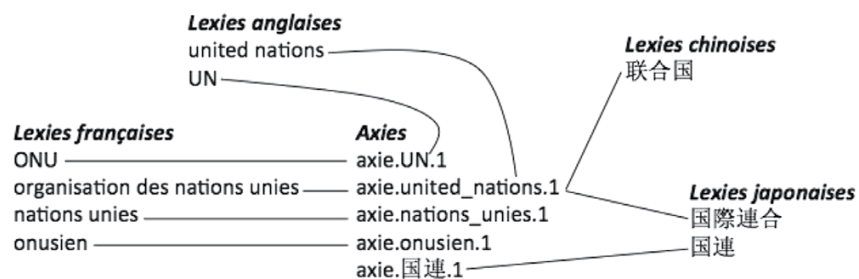
1. Nous avons repris exactement la terminologie française de Prolexbase et les concepts.

et le troisième caractère d'idéogramme ou kanji, ce qui est différent des initiales de la lexie de définition. Il existe peut-être une langue qui a deux acronymes, l'un correspondant à l'acronyme des initiales, l'autre correspondant à une sélection de caractères ou de mots. Donc, nous avons décidé de ne pas lier ces deux types acronymes avec une même axie. Voir la figure 2 - « Liens entre les lexies et les axes ».

**TABLE 1. EXEMPLE EN QUATRE LANGUES
POUR LE SENS « ORGANISATION DES NATIONS UNIES »**

LANGUE	ANGLAIS	FRANÇAIS	CHINOIS	JAPONAIS
Définition	United Nations	Organisation des Nations Unies	联合国	国際連合
Acronyme	UN	ONU		国連
Alias		Nations Unies		
Dérivée		onusien		

FIGURE 2 - LIENS ENTRE LES LEXIES ET LES AXIES



Lorsqu'on cherche les liens de la lexie « ONU » du français vers l'anglais, vers le japonais et vers le chinois, nous proposons trois niveaux de traduction classés selon la précision obtenue :

- Vers l'anglais : « ONU » → « UN ». Le système trouve une lexie directe par le volume des axes. C'est le premier niveau de traduction et le plus précis.
- Vers le japonais : « ONU » → « 国連 ». Le système cherche le lien dans le volume des prolexèmes français avec l'étiquette « Acro ». Puis il trouve le lien dans les proaxies, ensuite il suit le lien de prolexème japonais, enfin il arrive au volume des lexies japonaises, et il trouve une lexie avec l'étiquette « Acro ». Donc la lexie proposée du deuxième niveau de langue cible est cet acronyme. Le deuxième niveau de traduction comprend toujours le premier niveau de traduction. C'est-à-dire qu'« ONU » et « UN » ont la même étiquette « Acro », donc le lien « ONU » → « UN » correspond également au deuxième niveau de traduction.
- Vers le chinois : « ONU » → « 联合国 ». Le système trouve les lexies par prolexème et proaxie sans étiquette. Ces lexies proposées constituent le troisième niveau, le moins précis. Le troisième niveau de traduction comprend les niveaux précédents.

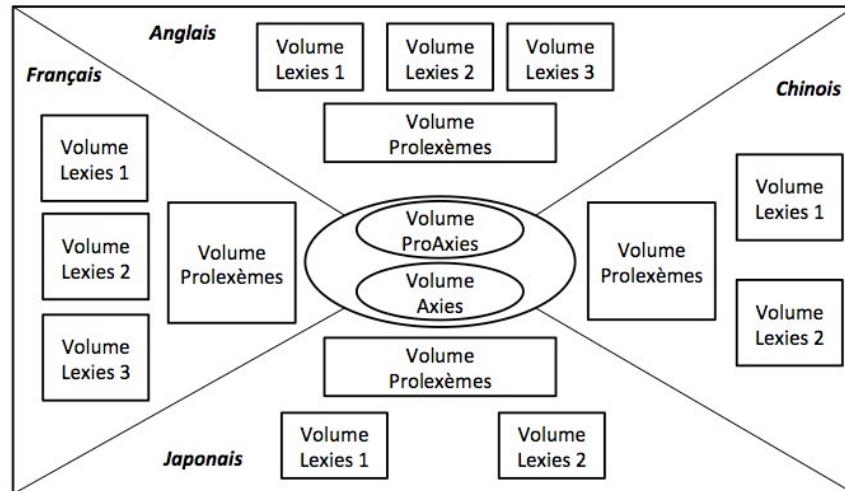
La quantité de lexies de résultat est augmentée suivant les niveaux de traduction du premier vers le troisième. Nous avons également utilisé l'exemple de traduction du terme « ONU » pour l'expliquer précisément, voir la table 2 - « Trois niveaux de traduction : les traductions du terme « ONU » ».

**TABLE 2. TROIS NIVEAUX DE TRADUCTION :
LES TRADUCTIONS DU TERME « ONU »**

	PREMIER NIVEAU	DEUXIÈME NIVEAU	TROISIÈME NIVEAU
Anglais	UN	UN	UN, United Nations
Japonais		国連	国連, 国際連合
Chinois			联合国

Dans certaines situations, une base lexicale (un dictionnaire) a plusieurs volumes pour une seule langue. Par exemple, lorsqu'il y a plusieurs versions d'édition ou que la ressource lexicale est créée par un système de traduction automatique, on trouvera par exemple un volume provenant de Systran, un volume de Google, un volume d'IATE, etc. Notre macrostructure permet de gérer plusieurs volumes dans une même langue. Voir la figure 3 « Macrostructure de ProAxie ». Pour une langue donnée, il existe un ou plusieurs volumes de lexies, mais un seul volume de prolexèmes. Pour un dictionnaire, il y a un seul volume de proaxies et un seul volume d'axies. Les entrées des lexies sont liées aux entrées de prolexèmes et sont aussi reliées aux entrées d'axies. De plus, les entrées de prolexèmes sont reliées avec les entrées de proaxies, et vice-versa.

FIGURE 3 : MACROSTRUCTURE DE PROAXIE



4 Outils nécessaires : plateformes de manipulation

4.1. Plateforme Jibiki version 1

Pour implémenter la macrostructure de ProAxie, nous avons utilisé la plate-forme Jibiki. Elle permet la construction de sites web de communautés dédiées à la construction de bases lexicales multilingues. Cette plate-forme a été développée principalement par Mathieu Mangeot et Gilles Sérasset (Mangeot 2006). Cette plate-forme est utilisée dans divers projets (projet LexALP, projet Papillon, projet GDEF, etc.). Le code est disponible en source ouverte et téléchargeable gratuitement sur ligforge.imag.fr par SVN. Avec cette plate-forme, on peut faire les manipulations d'import, export, édition, modification et recherche dans des bases lexicales. On peut aussi gérer les contributions.

Jibiki est une plate-forme générique, elle permet de traiter presque toutes les ressources lexicales de type XML en utilisant différentes microstructures et macrostructures. Nous avons déjà parlé de la macrostructure dans la section 3. La microstructure est la structure des entrées. On utilise des pointeurs CDM (*Common Dictionary Markup*) (Mangeot 2002) pour gérer n'importe quel type de microstructure sans la modifier. Les pointeurs sont utilisés également pour indexer des parties d'information spécifiques et permettre ensuite une recherche multicritère. Cette structure est stockée dans un fichier de métadonnées sous forme XML. Les fichiers de métadonnées ont pour but de faciliter l'import de ressources lexicales. Il y a un fichier de métadonnées pour chaque ressource. Dans ce fichier, sont décrites les

informations sur cette ressource : les langues source et cibles, l'auteur, les noms et fichiers de volumes, etc. Pour chaque volume, il existe un fichier de métadonnées. Dans ces fichiers, sont décrites toutes les informations des volumes, y compris les pointeurs CDM. Pour chaque pointeur CDM, on indique le chemin XPath vers l'élément correspondant dans la microstructure XML. Voir la figure 4 - « Exemple de CDM ». Les liens de traduction sont à ce stade traités comme des pointeurs CDM classiques (voir la figure 4, cdm-translation-ref).

FIGURE 4 : EXEMPLE DE CDM

```
<cdm-elements>
<cdm-volume xpath="/g:volume"/>
<cdm-entry xpath="/g:volume/g:article"/>
<cdm-entry-id xpath="/g:volume/g:article/@g:id" />
<cdm-headword xpath="/g:volume/g:article/g:vedette/g:mot/text()" d:lang="fra" />
<cdm-pronunciation xpath="/g:prononciation/text()" d:lang="fra" />
<cdm-pos xpath="/g:cat-gram/text()" d:lang="fra" />
<cdm-definition xpath="/g:volume/g:article/g:vedette/g:mot/text()" />
<cdm-translation-ref xpath="/g:mot-princ/g:value/text()" d:lang="fra"/>
</cdm-elements>
```

La version 1 de Jibiki présentait plusieurs limitations :

- Les fichiers de métadonnées utilisent des chemins Xpath pour décrire les fichiers XML. La création des métadonnées peut être difficile, en particulier pour un non-informaticien.
- Les liens de traduction sont traités avec des pointeurs CDM, comme des éléments d'information classiques. Ces liens sont simples. Il n'y a pas de possibilité de décrire des liens entre plusieurs volumes différents. Il n'est pas non plus possible d'ajouter des attributs (poids, étiquette, type, volume cible, etc.) sur les liens.
- Jibiki est utilisé pour plusieurs macrostructures. Pour chaque macrostructure, il est nécessaire de recoder une partie du programme pour réaliser les différents types de liens.

4.2. Gestion des données et métadonnées : iPoLex

Une fois la première version de Jibiki réalisée, nous nous sommes rendu compte d'un autre problème soulevé par les ressources : que faire des fichiers de données brutes au format textuel avant importation et après exportation de la plate-forme ? De plus, la rédaction de fichiers XML de métadonnées décrivant chaque ressource est très délicate et réservée à des informaticiens avertis. La gestion des différentes versions d'une même ressource pose également problème. Enfin, les ressources au format textuel peuvent être utilisées directement par d'autres applications. Il convient donc d'organiser leur stockage dans de bonnes conditions.

L'idée s'est alors imposée de concevoir un entrepôt de données lexicales permettant de stocker les données brutes au format textuel ainsi que de définir les métadonnées d'une nouvelle ressource à l'aide d'une interface. Ce projet, nommé iPoLex avait comme objectif d'être simple et rustique.

L'entrepôt est accessible par le Web. Les interfaces sont programmées en PHP sans connexion à une base de données. Toutes les informations sont stockées dans des fichiers texte. La page d'accueil affiche la liste des ressources disponibles et permet d'effectuer différents tris (format, langue source, langues cibles, etc.).

L'ajout d'une nouvelle ressource se fait en trois étapes : d'abord la description de la ressource avec ses métadonnées (langues source et cibles, domaine, auteur, date de création, format, taille des fichiers texte, etc.)

puis, pour chaque volume de la ressource, la description de ses métadonnées (langue source, nombre de mots-vedettes, version, etc.) ainsi que celle des pointeurs CDM permettant d'indexer le volume après importation dans la base Jibiki. La description des ressources se termine par la création d'un répertoire sur le serveur et la génération de fichiers de métadonnées.

La dernière étape est celle du téléversement des fichiers de données sur le serveur. Pour cela, l'entrepôt peut se monter comme répertoire distant grâce au protocole WebDAV. Une fois les fichiers de données téléversés sur le serveur, l'import dans une plate-forme Jibiki se fait très facilement en indiquant dans l'interface de Jibiki l'URL du fichier de métadonnées de la ressource à importer.

Nous prévoyons d'étendre ce concept d'entrepôt de données à celui d'entrepôt de services. En effet, pour la plupart des ressources stockées dans l'entrepôt, il existe des services associés, comme des moulinettes de conversion d'un format à un autre. Ces scripts sont en général prévus pour une version particulière des données. S'ils ne sont pas documentés précisément et stockés avec les données, il est très difficile, voire impossible, de réexécuter les processus divers associés aux données.

4.3. Gestion des liens riches : extension Jibiki-Pivax

La gestion des liens riches correspond aux liens avec des attributs, comme le volume cible, le poids, le type (axi, final), la langue, une étiquette libre, etc. Pour réaliser l'implémentation de liens riches, nous avons séparé le module de traitement des liens de celui des autres pointeurs CDM.

La réalisation de programme est basée sur deux algorithmes. Le premier est l'algorithme de collecte des liens, le deuxième est l'algorithme de construction du résultat. Le premier algorithme permet de chercher tous les liens possibles dans l'ensemble des liens riches de tous les volumes pour une entrée recherchée. Le deuxième algorithme nous permet de réaliser le parcours des liens riches. Il s'agit des étapes suivantes :

- chercher le lien vers les axies puis vers les lexies cibles,
- chercher le lien vers les prolexèmes de langue source puis vers les proaxies, ensuite vers les prolexèmes de langue cible, à la fin vers les lexies cibles,
- traiter l'étiquette,
- trier et afficher.

5. Résultats préliminaires

Nous avons séparé les trois niveaux de traduction pour afficher les résultats de recherche dans Jibiki : 1^{er} – traduction directe par axie, 2^e – traduction par prolexème et proaxie avec la même étiquette, 3^e – traduction par prolexème et proaxie sans étiquette. Pour faciliter la lecture, nous avons décidé :

- d'afficher l'étiquette, la langue et le mot-vedette dans le 1^{er} et le 2^e niveaux,
- d'afficher tous les détails (phrases d'exemple, définition, POS, etc.) dans le 3^e niveau, y compris les lexies du même prolexème de la langue source,
- de ne pas afficher la traduction dans le 2^e niveau si elle a déjà été trouvée et est déjà affichée dans le 1^{er} niveau.

5.1. Scénario 1 : terme « UN » de l'anglais vers toutes les langues

TABLE 3. TROIS NIVEAUX DE TRADUCTION ET AFFICHAGE : TERME « UN » DE L'ANGLAIS VERS TOUTES LES LANGUES

NIVEAU	LEXIES TROUVÉES PAR THÉORIES				LEXIES AFFICHÉES PAR INTERFACE			
	Fra	Eng	Zho	Jpn	Fra	Eng	Zho	Jpn
1	ONU				ONU			
2	ONU			国連				国連
3	ONU, Nations unies, onusien, Organisation des Nations unies		联合国	国際連合, 国連	ONU, Nations unies, onusien, Organisation des Nations unies	UN, United Nations	联合国	国際連合, 国連

FIGURE 5 : TERME « UN » DE L'ANGLAIS VERS TOUTES LES LANGUES

FIGURE 6 : AFFICHAGE COMPLET – TERME « UN » DE L'ANGLAIS VERS TOUTES LES LANGUES

5.2. Scénario 2 : terme « 国連 » du japonais vers toutes les langues

TABLE 4. TROIS NIVEAUX DE TRADUCTION ET AFFICHAGE : TERME « 国連 » DU JAPONAIS VERS TOUTES LES LANGUES

NIVEAU	LEXIES TROUVÉES PAR THÉORIES				LEXIES AFFICHÉES PAR INTERFACE			
	Fra	Eng	Zho	Jpn	Fra	Eng	Zho	Jpn
1								
2	ONU	UN			ONU	UN		
3	ONU, Nations unies, onusien, Organisation des Nations unies	UN, United Nations	联合国		ONU, Nations unies, onusien, Organisation des Nations unies	UN, United Nations	联合国	国際連合, 国連

FIGURE 7 : TERME « 国連 » DU JAPONAIS VERS TOUTES LES LANGUES



FIGURE 8 : TERME « ONUSIEN » DU FRANÇAIS VERS TOUTES LES LANGUES



5.3. Scénario 3 : terme « onusien » du français vers toutes les langues

TABLE 5. TROIS NIVEAUX DE TRADUCTION ET AFFICHAGE : TERME « ONUSIEN » DU FRANÇAIS VERS TOUTES LES LANGUES

NIVEAU	LEXIES TROUVÉES PAR THÉORIES				LEXIES AFFICHÉES PAR INTERFACE			
	Fra	Eng	Zho	Jpn	Fra	Eng	Zho	Jpn
1								
2								
3		UN, United Nations	联合国	国際連合, 国連	ONU, Nations unies, onusien, Organisation de nations unies	UN, United Nations	联合国	国際連合, 国連

6. Conclusion et perspectives

La gestion des terminologies avec liens riches est présentée avec un exemple d'acronyme (ONU) de nom propre dans cet article. Nous avons repris les concepts de lexie, d'axie, de prolexème, et établi le concept de proaxie pour produire la macrostructure de ProAxie. Dans cette macrostructure, une étiquette est utilisée pour représenter les relations de lexie. Nous avons implémenté la solution de la macrostructure ProAxie dans la plateforme Jibiki en utilisant la nouvelle extension Jibiki-Pivax, et créé trois niveaux de traduction en théorie et en affichage.

Pour le futur, nous souhaitons faire évoluer la macrostructure de ProAxie pour prendre en compte les synonymies de genre, et transposer le concept de prolexème pour que cette solution puisse être utilisée dans un autre domaine linguistique. Par exemple, pour une ressource lexicale comprenant des textos, en français « A+ » correspond à « À plus » avec une étiquette « texto », et en anglais « L8R » correspond à « later » avec l'étiquette « texto ».

Nous prévoyons de prendre en compte également les quatre variations du diasystème basé essentiellement sur ce que Eugenio Coseriu propose : diachronique (variété dans le temps), diaphasique (variété concernant les finalités de l'emploi), diatopique (variété dans l'espace) et diastratique (variété relative à la stratification socio-culturelle). Nous voudrions plus tard enrichir la notion d'étiquette.

Bibliographie

- FRANCOPOULO (G.), BEL (N.), MONTE (G.), and CALZOLARI (N.), MONACHINI (M.), PET (M.), SORIA (Cl.) 2009, "Multilingual resources for NLP in the lexical markup framework (LMF)", in *Journal of Language Resources and Evaluation*, March 2009, vol. 43, pp. 57-70.
- MANGEOT (M.), NGUYEN (H.-T.) 2009, "Building lexical resources: towards programmable contributive platforms", in *Proc. IEEE-RIVF*, 2009, pp. 84-92.
- MANGEOT (M.) 2006, "Dictionary Building with the Jibiki Platform", in *Proc. of EURALEX*, 2006, pp. 1666-1669 (www.lrec-conf.org/lrec2006).
- MANGEOT (M.) 2002, "An XML Markup Language Framework for Lexical Databases Environments: the Dictionary Markup Language", in *Proc. of Papillon 2002 Workshop* (www.papillon-dictionary.org/static/info_media/1620673.pdf).
- MELČUK (I.), CLAS (A.), POLGUÈRE (A.) 1995, *Introduction à la lexicologie explicative et combinatoire*, Louvain-la-Neuve, De Boeck.
- NGUYEN (H.-T.) 2009, *Des systèmes de TA homogènes aux systèmes de TAO hétérogènes*, thèse de doctorat d'informatique, Université de Grenoble, 236 p.
- SÉRASSET (G.), MANGEOT (M.) 2001, "Papillon Lexical Database Project: Monolingual Dictionaries and Interlingual Links", in *Proc. Of NLPRS*, pp. 119-125 (www.papillon-dictionary.org/static/info_media/1620822.pdf).
- TOMOKIYO (M.), MANGEOT (M.), PLANAS (E.) 2001, "Papillon: a Project of Lexical Database for English, French and Japanese, using Interlingual Links", in *Actes de JST 2001* (www.papillon-dictionary.org/static/info_media/120771.pdf).
- TRAN (M.) 2006, *Prolexbase Un dictionnaire relationnel multilingue de noms propres: conception, implémentation et gestion en ligne*, thèse de doctorat d'informatique, Université de Tours, pp. 55-57.

Le traitement des interjections dans le dictionnaire bilingue français-arabe

Fethi JMIL
*Unité de Recherche Le vocabulaire arabe
entre lexique et dictionnaire
Université de la Manouba*

Résumé

Parmi les catégories lexicales traditionnelles connues dans la tradition grammaticale française et dans les écrits linguistiques modernes (le verbe, le nom, l'adjectif, l'adverbe, et les particules), la dernière catégorie a des caractéristiques particulières et soulève plusieurs problèmes. Nous nous proposons d'étudier le traitement de l'une de particules dans le dictionnaire bilingue français-arabe : l'interjection. Cette étude abordera deux types de problèmes : les problèmes théoriques qui concernent la définition, l'identification, la traductibilité de l'interjection, et les problèmes pratiques observés dans le dictionnaire bilingue *al-Manhal*, comme les stratégies de traduction, l'étiquetage, la néologie, la polysémie, etc.

Mots-clés :

interjection, traduction, lexicographie, dictionnaire bilingue.

1. Introduction

Parmi les catégories lexicales traditionnelles connues dans la tradition grammaticale française et dans les écrits linguistiques modernes (le verbe, le nom, l'adjectif, l'adverbe, et les particules), la dernière catégorie présente des caractéristiques particulières et soulève plusieurs problèmes pour le traducteur aussi bien que pour le lexicographe.

Nous proposons d'étudier, dans cette communication, le traitement de l'une des particules dans le dictionnaire bilingue français-arabe *al-Manhal* : l'interjection. Il s'agit d'une particule invariable et isolée dans le système. Se définissant par l'absence de rôle syntaxique, celle-ci constitue une intervention (ou une intrusion) directe du locuteur dans le discours. De point de vue sémantique, elle exprime une réaction affective vive (approbation, désapprobation, doute, colère, ironie, appel, etc.) dont le contenu est évoqué par l'intonation que lui confère le sujet parlant (Dubois, 2007 : 253).

2. Problèmes théoriques

Une bonne activité lexicographique doit être orientée, d'une façon explicite ou implicite, par une perspective théorique linguistique. C'est pourquoi nous proposons d'inaugurer cette étude par une brève discussion sur les problèmes théoriques que pose cette sous-catégorie lexicale dite « interjection ». Cette discussion évoquera les problèmes pratiques, c'est-à-dire les problèmes que le traducteur peut

affronter au cours de son activité, et les difficultés que le lexicographe peut rencontrer en élaborant un dictionnaire bilingue.

2. 1. L'interjection : définition, identification et estimation

Il faut indiquer d'abord le problème de définition et d'identification de l'interjection dans les écrits linguistiques occidentaux. C'est que, malgré les critères de natures sémantique, formelle, pragmatique (Ameka 2006 : 743 a - 744 a) et cognitive (Cuenca 2000 : 30) élaborés pour définir les interjections et les identifier, la distinction entre cette sous-catégorie lexicale (ou grammaticale) et les autres catégories et sous-catégories reste parfois douteuse. Cela peut s'expliquer par le fait que quelques interjections sont à l'origine des verbes, des noms ou des adverbes, etc. (Thawabteh 2010 : 501) qui sont utilisés par le sujet parlant pour exprimer une réaction affective vive, « *c'est le cas de certains syntagmes nominaux* (Ciel ! Mon œil !), *de certains adjectifs* (Joli !), *d'expressions verbales* (Tiens ! Tu vois !) » (Dubois 2007 : 253). Mais certains linguistes vont plus loin : ils mettent en doute la nature langagière des interjections, en les considérant comme des « *cris* » ou des « *bruits* » que l'on peut rapprocher des gestes plutôt que de les considérer comme des expressions linguistiques (Goffman 1981 : 108 ; Richet 2001 : 2). Ce point de vue, qui minimise l'interjection, concerne les interjections dites « primaires ». Ce type est fondé sur les cris, les interjections vocales et quelques onomatopées, que certains linguistes considèrent donc comme des réalités langagières universelles dont la traduction est inutile (Richet 2001 : 3). Au contraire, les interjections dites « secondaires » sont fondées sur des mots normaux devenus

interjections (Richet 2001 : 3), et sont très liées à la culture d'un pays. Cette distinction nous paraît essentielle pour appréhender la complexité de cette classe et la difficulté de la définir. L'interjection est donc une classe hétérogène : elle comporte des mots onomatopéïques, des mots simples et des éléments langagiers composés (Cuenca : 34).

Bref, malheureusement, dans les études de linguistique occidentale, l'interjection reste la sous-catégorie la plus négligée parmi les catégories et les sous-catégories lexicales (Ameka 1992). Nous allons voir que son statut n'est pas plus favorable dans la grammaire arabe.

2.2. L'interjection dans la littérature grammaticale arabe

Il est vrai que l'on trouve en arabe des catégories lexicales (ou des sous-catégories) qui correspondent presque parfaitement aux catégories lexicales en français (*'ism* = nom, *fi'l* = verbe, *Sifah* = adjectif, *hâl* et *darf* = adverbe, *'adât* = particule). Mais, les particules sont nombreuses (elles comprennent les prépositions, les pronoms, les conjonctions, les déterminants, et les interjections) (Ben Mrad 2010 : 86), et l'on ne trouve pas, dans les particules de l'arabe, une sous-catégorie qui puisse correspondre à l'interjection. C'est pourquoi le linguiste égyptien Tammâm Hassân (Hassan : 89) a emprunté, pour la dénommer, le terme classique de « *khâlifah* » qu'il a puisé chez al-Farra' (144-208h. /761- 822c.), un ancien philologue arabe. Son argument est que les « *khawâlif* » (pl. de *khâlifah*, lit. Qui remplace qqn qui vient après, qui est différent...) expriment une certaine position affective (Hassân 2001 : 88- 90 et 113-118). On peut trouver aussi dans les particules

appelées en arabe « *'ism al-fi'l* » (lit. « nom de verbe ») et « *'ism Sawt* » (lit. « nom de son ») quelques ressemblances avec les interjections.

Mais le problème est que les lexicographes arabes n'ont pas toujours enregistré – et n'ont donc pas « *systématisé* » – les interjections utilisées dans le discours quotidien. Cette négligence est très remarquée dans les dictionnaires arabes contemporains qui ne manifestent pas d'intérêt pour les interjections issues de la langue parlée. Pourtant, c'est ce niveau de langue, caractérisé par la diglossie, qui fournit une grande diversité d'intonations exprimant des positions et des réactions affectives dans le discours. En outre, certaines interjections constituent des cases vides dans le lexique arabe. Ceci constitue un vrai problème pour le traducteur et pour l'auteur du dictionnaire bilingue parce que l'arabe ne leur offre pas souvent des équivalents pour traduire les différentes interjections qu'ils rencontrent dans la langue source.

L'étude linguistique des interjections de point de vue contrastif pose le problème de l'universalité et de la traductibilité des interjections.

2.3. Universalité et traductibilité

Si l'on revient aux deux types d'interjections, primaires et secondaires, peut-on parler d'une traductibilité pour le premier type, puisque les interjections primaires sont universelles ? Et peut-on parler d'une difficulté ou d'une impossibilité de traduction pour les interjections secondaires, puisqu'elles constituent des spécificités culturelles propres à un pays ? Ces deux questions paraissent légitimes au premier coup d'œil. Mais deux constats viennent relativiser

l'universalité des interjections primaires. Le premier est que les communautés linguistiques changent souvent leur façon de s'exprimer. Le deuxième, qui semble découler du premier, est que les interjections primaires attestent de la synonymie et de la polysémie. Le fait est que nous disposons souvent de plusieurs manières pour exprimer une même réaction affective vive, et nous constatons que le sujet parlant attribue, dans plusieurs cas, plus d'un sens à une même interjection. Dans le *Petit Robert*, on lit dans l'article « Ah ! » :

interj. (xvi^e ; a, 1050 ; lat. *a(h)*, onomat.) 1. Interjection expressive, marquant un sentiment vif (plaisir, douleur, admiration, impatience, etc.) ; 2. Interjection d'insistance, de renforcement. « Ah ! *Que j'en ai suivi de ces petites vieilles !* » (Baudel.). Ah ! *J'y pense, pouvez-vous venir demain ?* 3. (Redoublé). Marque la surprise, la perplexité. Ah ! ah ! *c'est ennuyeux.* 4. Sert à transcrire le rire. Ah ! ah ! *Elle est bien bonne !* 5. (En loc. exclamat.) Ah ça ! Ah mais ! Ah bien oui ! Ah bon ! » (*Le Petit Robert* 1982 : 39).

L'interjection *ah !* est donc polysémique. En outre, on trouve d'autres interjections, comme *aïe !*, *ouïe !* ou *ouille !* et *ouf !*, qui peuvent exprimer un des sentiments vifs qu'exprime *ah !*, à savoir, la douleur.

Aïe ! Interj. (1473) ; onomat.). Interjection exprimant la douleur, et par ext. une surprise désagréable, un ennui. (*Le Petit Robert* 1982 : 40)

Ces interjections ne sont peut-être pas interchangeable dans le discours, mais elles sont bel et bien synonymes dans quelques significations.

Dans la langue arabe, nous observons le même phénomène de polysémie et de synonymie. Mais, outre les problèmes de diglossie, de faiblesse d'enregistrement de données langagières et de non-systématisation déjà évoqués, le traducteur (ou le lexicographe élaborant un

dictionnaire bilingue français-arabe) se trouve embarrassé quand il cherche l'équivalent ou les équivalents. En outre, l'accompagnement des gestes et des intonations avec les interjections (Ameka 2006 : 743b) peut remettre en cause tout choix d'équivalent.

En effet, un dictionnaire bilingue constitue un corpus linguistique qui nous aide à observer les solutions auxquelles un traducteur peut recourir pour résoudre les problèmes théoriques évoqués. Voyons maintenant quels sont les problèmes pratiques que le lexicographe doit gérer lors de l'élaboration d'un dictionnaire bilingue, qui est, en somme, une activité de traduction.

3. Problèmes pratiques

Nous allons étudier dans cette partie le traitement des interjections dans un dictionnaire bilingue français-arabe : *Al-Manhal* du Libanais Souheil Idris (Idris 1997), de manière à identifier les moyens linguistiques que l'auteur – en tant que traducteur et lexicographe – a utilisés pour combler les cases vides et les cases partiellement pleines que représentent les interjections d'une langue source, en l'occurrence le français, pour une langue cible, nommément l'arabe.

Notre étude vise à évoquer quelques problèmes sans prétendre à l'exhaustivité. La distinction entre ces problèmes est d'ordre méthodologique parce qu'ils sont bien liés les uns aux autres.

3.1. Problème d'exhaustivité

Nous remarquons d'abord que Souheil Idris a négligé dans *Al-Manhal* plusieurs interjections françaises. Citons par exemple : *à-Dieu-va(t) !*; *diable !*; *fouchtra !*, etc. Cela nous paraît fâcheux dans un dictionnaire bilingue parce que l'utilisateur ne peut pas trouver dans les procédures morphologiques une solution pour comprendre l'interjection, du type primaire au moins. Nous avons déjà discuté la nature universelle de l'interjection.

3.2. Problème de choix

Le lexicographe choisit, comme un traducteur, un équivalent qui lui paraît adéquat pour traduire l'unité lexicale de la langue source. Il a, pour ce faire, recours aux procédures de traduction.

Richet (2001 : 10) parle, dans son article consacré à la traduction des interjections entre le français et l'anglais, des sept types de traduction suivants : (i) une traduction strictement littérale (même interjection de surface) ; (ii) une traduction par une interjection différente, mais à l'intérieur d'un même type ; (iii) une traduction de type différent ; (iv) lorsque l'interjection initiale est traduite par autre chose qu'une interjection ; (v) lorsque l'interjection n'est pas traduite du tout ; (vi) et (vii) correspondent à l'introduction d'une interjection dans le texte d'arrivée, lorsque celle-ci se fonde sur une base formelle [type vi] et lorsqu'il s'agit d'une « pure création » [type vii].

Cuenca (2006 : 27), quant à elle, distingue six stratégies pour traduire les interjections :

- Stratégie a : traduction littérale.
- Stratégie b : traduction par une interjection qui diffère formellement, mais qui a le même sens.

- Stratégie c : traduction par une structure non interjective qui a un sens similaire.
- Stratégie d : traduction par une interjection qui a un sens différent.
- Stratégie e : Omission de l'interjection.
- Stratégie f : Addition des éléments.

Malheureusement, nous ne pouvons pas tirer profit de ces types et stratégies parce qu'ils ne conviennent pas exactement au corpus que nous étudions. C'est que les interjections dans un dictionnaire bilingue sont hors contexte. Nous allons donc décrire les différentes procédures que Souheil Idris a utilisées pour rendre compte des interjections dans son dictionnaire bilingue. Ce tableau présente les diverses procédures de traduction relevées, avec quelques exemples :

PROCÉDURES	EXEMPLES
1. ÉQUIVALENT ARABE : <i>a- Un seul équivalent</i>	a- Bof !: <i>tabban lahu!</i> [insulte : qu'il périsse !], p. 175 Zut !: <i>'uffin !</i> , p. 1289 Bernique !: <i>yâ lal-khaybah</i> [quel retournement !], p. 144
<i>b- Plus qu'un équivalent</i>	b- Bravo !: <i>marHâ !</i> [hourra !], <i>mumtâz</i> [excellent], <i>'ahsanta</i> [tu as bien fait], p. 174 Ah !: <i>'âh !, wâhan !, 'ay !, way !</i> (<i>Sawt yadullu 'alâ al-faraH 'aw al-'alam 'aw al-'ijâb 'aw nafadh aS-Sabr</i> [son qui exprime la joie ou la douleur ou l'admiration ou l'impatience]), p. 49
2. ÉQUIVALENT AR. + PARAPHRASE	Bah !: <i>bâh !, 'ajaban !</i> [mot signifiant l'étonnement], p. 123 Hourra, Hurrah !: <i>hurrâh !, marHâ !</i> , p. 622
3. EMPRUNT EXTERNE + PARAPHRASE	Hou !: <i>hû !</i> (<i>sawt yadullu 'alâ 'al'istiyyâ'</i> [son qui exprime le mécontentement]), p. 622 Olé !: <i>'ûlih ! (nidâ' lit-ta'jî'</i> [vocation pour encourager], p. 841
4. EMPRUNT INTERNE AUX DIALECTES	Hou !: <i>hû !</i> (<i>sawt yadullu 'alâ 'al'istiyyâ'</i> [son qui exprime le mécontentement]), p. 622 Olé !: <i>'ûlih ! (nidâ' lit-ta'jî'</i> [vocation pour encourager], p. 841
5. ÉQUIVALENT AR. + EMPRUNT EXTERNE	Han !: <i>hâh !</i> (son pour exprimer la fatigue), p. 603 Hum !: <i>'Him ('aw 'yu'laq lid-dalâlah 'alâ a'-akk 'aw at-taraddud</i> [son exprimant le doute et l'incertitude], p. 623
6. TRADUCTION D'UN CONTEXTE PRÉCIS	Holâ !: <i>yâ ! ('adât nidâ'</i> [particule de vocation]), <i>holâ !</i> pas si vite : « <i>mahlan! lâ tu'ajjil hakadâ</i> [Attends ! Ne te hâte pas comme ça], p. 617 Ah !: <i>'âh !, wâhan !, 'ay !, way !</i> (...); Oh ! Quel tableau !: <i>yâ lahâ min lawHah</i> ; oh ! Quel bonheur !: <i>wâ sa'dâh !</i> ; oh ! Que c'est admirable !: <i>mâ 'arwa'a hâthâ</i> , p. 49
7. DÉFINITION SANS ÉQUIVALENT	Berk !, beurk !: <i>Sawt yadullu 'alâ at-taqazzuz</i> [son qui exprime le dégoût], p. 143 Haro !: <i>Surâkh, SiyâH</i> [cri, hurlement], p. 6

Telles sont donc les sept principales procédures que nous avons identifiées. Souheil Idris a donc pris soin, dans la majorité des cas, de recourir aux paraphrases écrites entre parenthèses, de donner des précisions qui concernent les différents emplois et contextes puisque le sens de l'interjection est après tout un *sens pragmatique* (« *pragmatic meaning* », Cuenca 2000 : 32) que l'on ne peut interpréter que selon le contexte et les conditions dans lesquelles cette interjection est utilisée. Dans ce sens, Richet écrit : *l'équivalence figée disparaît sous le poids des multiples paramètres gouvernant le sens. Le contexte, le style, le ton, le personnage, l'auteur sont autant d'éléments qui viennent naturellement nuancer et complexifier la simple translation* (Richet 2001 : 6 ; voir aussi Amaka 2006 : 745b ; Thawabteh : 505). Ce souci transparait aussi dans les exemples de *Ah !* et *Olé !*, où Idris a choisi des phrases qui contiennent des interjections et les a traduites pour montrer les différents sens et les différentes façons de traduire.

Mais a-t-il toujours fait des choix judicieux ? Comment a-t-il géré le phénomène de la polysémie ? Avait-il raison de recourir aux emprunts internes aux dialectes arabes et externes au français ? Comment évaluer son choix de point de vue de l'appartenance catégorielle de l'équivalent proposé ? Cet équivalent est-il toujours un équivalent interjectif ? Et enfin, quels sont les moyens utilisés pour traduire l'intonation qui accompagne l'interjection ?

3.3. Problème de polysémie

Prenons l'exemple de *Ah !* (p. 49) qui montre le problème de polysémie aussi bien dans l'interjection française que dans l'équivalent

arabe. Dans le *Petit Robert* (1982 : 39) nous trouvons au moins sept sens : (1) sentiment vif de plaisir, (2) douleur, (3) admiration, (4) impatience ; (5) insistance et renforcement, (6) surprise et perplexité, (7) transcription de rire. Quant aux équivalents arabes : *'âh !, wâhan!, 'ay !, way !*, ils ne semblent pas suffisants pour rendre tous les sens cités dans le *Petit Robert* qu'Idris a commentés par la paraphrase : (*Sawt yadullu 'alâ al-faraH 'aw al-'alam 'aw al-'ijâb 'aw nafadh aS-Sabr* [son qui exprime la joie ou la douleur ou l'admiration ou l'impatience]). Ses équivalents arabes nous semblent loin de rendre le sens de 'réaction d'impatience et d'insistance'.

D'autre part, des intonations variées peuvent accompagner *'âh !* (dans quelques dialectes arabes au moins), et dénoter différentes interprétations non rencontrées français. Pour *aïe !* qui exprime la douleur en français, Idris donne deux équivalents : *'âh !* et *'ahâ !* ; le premier peut signifier, selon l'intonation, l'admiration, l'impatience, etc. ; et le second peut signifier 'je comprends maintenant !', etc.

Le choix d'équivalent ne nous paraît pas une tâche facile, mais il est possible de se servir des paraphrases pour expliquer les différents emplois et aider le traducteur (et le lecteur en général) à interpréter l'interjection utilisée dans un contexte bien déterminé.

3.4. Problème de néologie

En cas de cases vides, le lexicographe arabe se trouve obligé de recourir aux néologismes, soit par l'emprunt à la langue source, soit par l'emprunt aux dialectes. Le premier choix implique que la

communauté utilisant la langue cible soit habituée aux interjections empruntées par un contact culturel permanent. Mais ce n'est pas le cas partout. Les pays du Maghreb et le Liban sont plus habitués au français que les autres pays arabes. Quant au deuxième choix, l'emprunt aux dialectes, il pose encore une fois le problème de l'universalité des interjections entre les dialectes apparentés. C'est que les peuples arabes diffèrent parfois dans leurs façons d'exprimer leurs sentiments et réactions affectives vives. Quoique, de nos jours, grâce aux mass médias, il semblerait que ce problème en soit de moins en moins un.

Cependant, la nature polysémique et l'information intonative qui accompagne les interjections restent toujours des obstacles qui peuvent maintenir l'opacité au cours de l'interprétation. La traduction de *hum !* par *'iHim* n'est pas toujours satisfaisante parce que *'iHim* exprime, selon le contexte, la situation et l'intonation, un signe d'attention que l'homme utilise avant d'entrer une maison dans laquelle se trouve des femmes étrangères ou *Harims* ; mais ce son exprime aussi une position de refus pleine de moquerie envers la parole de quelqu'un d'autre.

3.5. Problème d'appartenance catégorielle

Une traduction parfaite et idéale serait peut-être une traduction proposant, pour chaque unité langagière dans une langue source, un équivalent de même appartenance catégorielle dans la langue cible. Mais, cela reste toujours difficile à concrétiser. Chaque langue a sa façon d'exprimer le monde selon un découpage spécifique et un système

de catégorisation. Il n'en reste pas moins vrai que la transmission du sens reste toujours le but « suprême » de la traduction, quelle que soit la procédure utilisée.

Idris a essayé, dans la plupart des cas, de trouver une particule pour rendre les interjections françaises : *'ism Sawt comme hâ* pour rendre *ha !* (2006 : 600) ; *'ism fi'l comme 'uffin !* pour rendre *foin !* (*op. cit.* : 539) ; et *'ah !* pour rendre *chut !* (*op. cit.* : 248).

Il a parfois recours aux équivalents non interjectifs comme les phrases nominales et verbales, les phrases exclamatives et affirmatives, et les « compléments absolus » (*maf'ûl mutlaq*) qui expriment de façon directe la réaction du sujet parlant : *'aHsanta* pour rendre *bravo !* (*op. cit.* : 174) ; *kafâ ! 'innaka mudhjr* pour rendre *basta !* (*op. cit.* : 135) ; *'ajaban !* pour rendre *bigre !* (*op. cit.* : 148) ; *tabban lahu !* pour rendre *bof !* (*op. cit.* : 157).

Nous pouvons constater aussi que Souheil Idris a essayé, dans la plupart des cas, d'ajouter un point d'exclamation après l'équivalent arabe pour dénoter la charge d'intonation et l'expression de réaction vive qui accompagnent les interjections.

4. Conclusion

Notre étude, qui a pour objet le traitement des interjections dans un dictionnaire bilingue français-arabe, montre les difficultés

de nature théorique et pratique que le lexicographe peut affronter. Nous avons évoqué parmi les problèmes théoriques, le problème de définition, identification et estimation de l'interjection dans de différentes langues, son aspect oral et écrit, et son universalité et sa traductibilité. Une étude descriptive du traitement des interjections françaises dans *Al-Manhal* a montré que Souheil Idris a essayé de résoudre le problème du choix d'un équivalent adéquat en recourant à plusieurs procédures comme la mention de plusieurs équivalents pour chaque interjection française, le recours à l'emprunt au français et aux dialectes arabes, la traduction de quelques exemples français, etc. Cependant, il nous semble que l'auteur aurait pu mieux utiliser les paraphrases pour évoquer les différents emplois des interjections, et il aurait pu profiter des particules dans l'arabe classique et surtout dans les dialectes arabes contemporains qui restent un champ d'étude linguistique non découvert.

Bibliographie

- AMEKA (F. K.) 1992, « Interjections : The universal yet neglected part of speech », in *Journal of Pragmatics*, Amsterdam, n° 18, pp. 101-118.
- AMEKA (F. K.) 2006, « Interjections », in *Encyclopedia of Language and Linguistics*, Brown (K.) ed., London, Elsevier, 2 vol. 6, pp. 743-746.
- BEN MRAD (Ibrahim) 2010, *Min al mu'jam 'ila al-qâmus* (Du lexique au dictionnaire), Tunis, Dâr al-gharb al-islâmi.
- CUENCA (M. J.) 2000, « Defining the indefinable? Interjections », in *Santaxis*, Madrid, n° 3, pp. 29-44.
- CUENCA (M. J.) 2006, « Interjections and Pragmatics Errors in Dubbing », in *Meta*, Montréal, 51 (1), pp. 20-35.
- DUBOIS (J.) et al. 2007, *Dictionnaire de linguistique & sciences du langage*, Paris, Larousse.
- HASSAN (T.) 2001, *al-Lughah al-'arabiyya : ma'nâhâ wa mabnâhâ* (La langue arabe : son sens et son structure), Casablanca, Dâr al-Thaqâfa.
- IDRIS (S.) 2006, *Al-Manhal* (Dictionnaire français-arabe), 35^e éd., Beirouth, Dâr al-âdâb.
- RICHET (B.) 2001, « Quelques données et réflexions sur la traduction des interjections », in *Oralité et traduction*, M. Ballard éd., Arras, Artois Presses Université, pp. 79-128.
- ROBERT (P.) 1982, *Le Petit Robert*, Paris, Le Robert.
- THAWABTEH (M. A.) 2010, « The translatability of interjections: a case study of Arabic-English subtitling », in *Meta*, Montréal, vol. 55, no 3pp. 499-515.

Dire la pensée complexe

Analyse de la microstructure comme prolégomènes à l'informatisation d'un dictionnaire ancien : le cas particulier des *Curiositez françoises* (1640) d'Antoine Oudin

Claire DUCARME
Département de langues et littératures romanes
Université de Liège

Résumé

Dans cet article, nous mènerons quelques réflexions préalables à la réalisation d'une édition informatisée d'un dictionnaire particulier du 17^e siècle, comportant à la fois des unités simples et des unités phraséologiques : *Curiositez françoises pour supplement aux dictionnaires ou Recueil de plusieurs belles proprietes, avec une infinité de Proverbes & Quolibets, pour l'explication de toutes sortes de Livres* (1640) d'Antoine Oudin. Nous nous attacherons, plus particulièrement, à l'analyse de la microstructure qui constitue, à nos yeux, une étape essentielle dans le balisage. Nous montrerons ensuite que toute édition informatisée d'un dictionnaire ancien doit être sous-tendue par une analyse métalexicographique fine. Nous aborderons, plus rapidement, quelques aspects techniques de l'informatisation.

Mots-clés :

informatisation, balisage, dictionnaire, lexicographie historique, métalexicographie, *Curiositez françoises*, microstructure.

1. Introduction

De nombreux dictionnaires actuels sont informatisés, mis en ligne et facilement consultables. Mais qu'en est-il des dictionnaires anciens ? Plusieurs d'entre eux ont trouvé une seconde vie grâce à l'outil informatique. Le projet d'informatisation a suscité notre intérêt dans le cadre de notre thèse de doctorat, qui consiste en l'analyse métalexigraphique d'un dictionnaire du 17^e siècle : *Curiositez françoises pour supplement aux dictionnaires ou Recueil de plusieurs belles proprietes, avec une infinité de Proverbes & Quolibets, pour l'explication de toutes sortes de Livres* (1640) d'Antoine Oudin. Nous exposerons, dans cet article, les principales réflexions préalables à la réalisation d'une édition informatisée de ce dictionnaire. Nous nous attacherons, tout particulièrement, à l'analyse de la microstructure qui constitue l'étape essentielle en vue du balisage d'un dictionnaire ancien.

Tout d'abord, nous mettrons en évidence, de manière générale, les spécificités microstructurelles des dictionnaires anciens (2). Après avoir présenté l'auteur et l'œuvre (3), nous nous attacherons ensuite à identifier les différents champs informationnels du programme lexicographique des *Curiositez* (4). Une fois ceux-ci établis, nous présenterons notre projet d'édition informatisée de ce dictionnaire, intrinsèquement lié au travail métalexigraphique (5). Nous ferons une critique d'une édition existante (5.1) et nous présenterons nos principes éditoriaux (5.2). Enfin, nous aborderons, d'un point de vue plus technique, des pistes pour le balisage (5.3).

2. Microstructure des dictionnaires anciens : problème général

La microstructure d'un dictionnaire, selon J. Rey-Debove (1971 : 151), est *l'ensemble des informations ordonnées suivant l'entrée*. La microstructure requiert donc une lecture horizontale : le consultant doit entrer dans le texte linéaire de l'article du dictionnaire afin de prendre connaissance des diverses informations fournies par le lexicographe. Ces informations sont structurées formellement : *cet ensemble possède une structure constante répondant à un programme et à un code d'information applicable à n'importe quelle entrée* (1971 : 151). Une fois le programme lexicographique adopté, le lexicographe *n'est plus libre d'ajouter un type d'information dans un article (et non dans tous) ni d'en omettre un sans déstructurer l'article* (1971 : 170). Le métalexigraphe qui tente une description et une analyse du microtexte se doit de mettre en évidence cette structure en dégagant des invariants formels et discursifs. D'éventuelles variations peuvent perturber la lisibilité du discours lexicographique. Celles-ci sont peu fréquentes dans les dictionnaires modernes, qui présentent une structure rigide et homogène, contrairement aux dictionnaires anciens.

Les principales études menées sur le microtexte des dictionnaires anciens (16^e-18^e s.) sont dues à I. Leroy-Turcan (1999) et T. R. Wooldridge (1993, 1997, 1999). Leur réflexion sur l'information microstructurelle a été principalement menée dans le but de baliser le texte de dictionnaires anciens en vue de l'informatisation de ceux-ci. Le balisage est défini par A. Tutin & Ch. Wionet (2001 : 47) *comme une technique qui consiste à isoler les parties logiques d'un document (par*

exemple, l'entrée, la partie grammaticale, la définition, etc.) à l'aide de jalons (appelés balises) sans modifier le contenu et l'ordonnement des parties textuelles. Comme le souligne G. Petrequin (2009 : 132), les difficultés rencontrées par le métalexigraphe qui tente une description et une analyse du contenu informationnel des articles d'un dictionnaire ancien sont semblables à celles que rencontre son collègue qui veut baliser informatiquement un tel texte : il doit éviter de dénaturer inconsciemment le sens de l'information d'un discours qui ne présente pas le même formalisme que celui des dictionnaires modernes. En effet, la principale caractéristique du microtexte des dictionnaires anciens tient à une *systématicité très relative du codage de l'information* (Petrequin 2009 : 132). Les lexicographes anciens s'accordent une plus grande liberté dans le traitement informationnel.

Même si les dictionnaires anciens ont une moins grande rigidité structurelle que les dictionnaires modernes, il est possible de mettre en évidence, au sein du discours lexicographique, des champs informationnels. On entend par « champs informationnels » *tous les éléments du discours lexicographique qui présentent à la fois une même thématique et un certain degré de récurrence d'un bout à l'autre de l'ouvrage étudié* (Petrequin 2009 : 133). La mise en évidence de ces champs se fait souvent à partir de critères formels typographiques, discursifs et de positions relatives dans le microtexte. Cependant, le métalexigraphe doit tenir compte du fait qu'une partie de l'information échappe à ce formalisme. En effet, d'une part, les informations relevant d'un même champ informationnel se présentent fréquemment sous des formes discursives différentes : il y a une diversité dans les formulations ; d'autre part, les champs informationnels ne sont pas tous présents systématiquement : il y a souvent un manque de

récursivité. Cela ne veut pas dire qu'il est impossible de dégager certaines régularités mais qu'il faut savoir adapter les outils développés dans les études métalexigographiques de dictionnaires modernes afin d'épouser aux mieux les caractéristiques du discours lexicographique ancien, comme nous le montrerons dans le cas des *Curiositez*.

3. Auteur et œuvre

Antoine Oudin (1595-1653) est secrétaire-interprète du roi Louis XIII et professeur d'italien du jeune Louis XIV. À côté de ses charges à la cour, il poursuit une carrière de grammairien et de lexicographe. En 1640, il publie les *Curiositez françaises pour supplément aux dictionnaires ou Recueil de plusieurs belles propriétés, avec une infinité de Proverbes et Quolibets, pour l'explication de toutes sortes de livres*.

Ce dictionnaire possède plusieurs spécificités. Destiné aux étrangers, il peut être qualifié de spécialisé, en ce sens qu'il recueille un vocabulaire vulgaire. Grâce à cet ouvrage, Oudin veut faciliter la lecture de « toutes sortes de livres » appartenant au genre burlesque florissant à cette époque. Il montre qu'à côté de la langue normée, il existe un vaste matériel linguistique qu'il traite avec un certain souci d'exhaustivité¹. Son recueil est très complet. Ainsi, les *Curiositez françaises* offrent de nombreuses premières attestations. Par exemple, *avorton*, dans le sens de « petit homme mal bâti », y est attesté pour

1. Les *Curiositez françaises* comportent quelque 8 000 entrées.

la première fois. Dadais, dans le sens de « niais », est enregistré dans ce dictionnaire mais non dans les autres ouvrages lexicographiques du 17^e siècle. Le dictionnaire de l'Académie ne recueille cette unité lexicale que dans sa quatrième édition de 1762. Une autre grande spécificité des *Curiositez* consiste dans la diversité des unités formant la nomenclature. En effet, celle-ci comporte aussi bien des unités simples que des unités polylexicales. Si le lexicographe a jugé bon d'enregistrer ces dernières dans son dictionnaire, c'est que leur compréhension n'est pas toujours évidente.

Les *Curiositez françaises* constituent un véritable trésor lexicographique. À la suite de M. Pfister (1989 : 97), nous pouvons affirmer qu'il s'agit d'une source précieuse pour l'étude du bas langage de la première partie du 17^e siècle, une mine lexicographique de la langue populaire.

4. Analyse de champs informationnels

La diversité de formes et de contenus demande de bien comprendre les modalités propres de fonctionnement de l'ouvrage étudié. Nous présenterons ci-après, en étudiant les principaux champs informationnels d'un article, quelques spécificités propres aux *Curiositez*.

4.1. Les différents champs d'un article

Proposer une structure type pour tous les articles des *Curiositez* peut paraître assez illusoire étant donné la grande diversité

des éléments qui les constituent. Toutefois, nous pouvons établir une description assez précise du programme lexicographique en déterminant des champs informationnels généraux. Ainsi, sous une apparente diversité, se dessine une certaine régularité. Tout article comporte trois champs, comme le montre l'exemple suivant : une marque diaphasique codée (*), une forme d'entrée et un corps².

- * *Soufflez moy dans l'Œil, cela se dit à un valet qui retourne de la cave, pour sçavoir s'il n'a point beu de vin.*

Le corps de l'article peut comporter quatre catégories principales d'informations : sens et valeur d'emploi (1), exemple (2), marque d'usage (3) et motivation du signe (4).

- * *Chaud, nostre vulgaire se sert de ce mot pour tesmoigner un mespris : par exemple, un chaud personnage ; un chaud Medecin, &c. i. mauvais, & de peu de doctrine, vulg.*
- * *Ange de Greve. i. crocheteur, gaignedenier. C'est à cause de leurs crochets qui forment comme des ailes.vulg.*

Dans le premier exemple, le corps est constitué des catégories 2 et 3, alors que, dans le second, il comporte les catégories 2 et 4. Le champ informationnel « sens et valeur d'emploi » est le plus fréquent mais, dans de rares cas, ne figure pas.

Examinons chaque champ informationnel.

2. Nous reprenons cette notion de *corps* à Ch. Rey (1999-2000). Lors du balisage de dictionnaires anciens, Ch. Rey (1999-2000 : 50, 51) accorde une place majeure au découpage de l'article en deux grands blocs, à savoir le bloc 'entrée' et le bloc 'corps'.

4.2. Marque diaphasique codée (*)

Tout article des *Curiositez* comporte toujours un champ transmettant une information de type diaphasique. Il s'agit du premier élément de l'article. Il se matérialise sous la forme d'un astérisque, appelé étoile par Oudin, ou de l'absence de celui-ci. Cette marque joue un rôle discriminatoire car elle permet au lecteur de distinguer les formes qui la possèdent de celles qui ne l'ont pas. Elle a une valeur d'annonce et de mise en évidence de l'unité. Elle fonctionne avec la séquence textuelle *vulg.* qui se situe soit après la forme d'entrée (cas rare), soit en fin d'article (cas plus fréquent) et qui a une valeur de renforcement. Il s'agit en quelque sorte d'une marque redondante. En effet, *vulg.* n'apparaît jamais sans l'étoile, sauf dans des cas très rares. L'étoile, par contre, peut figurer seule.

Canarder un homme. i. *le tirer de loing avec une harquebuse ou mousquet.*

* Chier des yeux. i. *pleurer.*

* Catholique à gros grain. i. *mauvais Catholique qui penche à l'hérésie.* *vulg.*

Ces deux marques ont la même valeur pour Oudin, comme il le montre dans son avertissement : *Pour ce qui est des estoiles, & du mot vulg. il faut entendre que ce ne sont pas (sic) des phrases dont on se doive servir qu'en raillant* (Oudin 1640 : « Avertissement »). De toute évidence, selon les propos tenus par le lexicographe, ces deux marques sont d'ordre diaphasique et indiquent que, pour les unités marquées, l'usage doit se limiter à un discours de raillerie.

Le rôle que joue ce champ informationnel répond à l'objectif que veulent remplir les *Curiositez*. Dans son avis aux étrangers, Oudin précise qu'il veut remédier à un manque de précision dans l'information lexicographique : *Le seul but où je vise, & que j'estime assez raisonnable, est de purger les erreurs qui se sont glissées dans la plupart des pieces que l'on a mises en lumiere pour l'instruction des Estrangers et encore : Touchant les Dictionnaires, ils sont si mal ordonnez que l'on n'a pas seulement eu le soin de marquer le bon d'avec le mauvais* (Oudin 1640 : « Aux étrangers »).

4.3. Forme

Le premier champ informationnel que nous venons de présenter est généralement suivi par la forme d'entrée. Il est néanmoins possible qu'une copule du type *vulgairement nous disons* l'introduise. La forme se présente en caractères romains. Elle se termine par un point ou parfois une virgule.

un Allant. i. *Un finet, un rusé.*

* *Vulgairement nous disons*, c'est ce que nos chiens ont tant Abbayé cette nuit, pour nous mocquer de ce que l'on nous vient proposer ou persuader.

Il est essentiel de distinguer, au sein de la forme, l'unité de traitement, c'est-à-dire la partie qui fait l'objet d'un traitement lexicographique. En effet, la forme, lorsqu'elle est polylexicale, n'est pas toujours traitée dans son entièreté.

* Qui bon l'Achepte bon le boit ; *vulg. c'est pour dire qu'il est mieux d'achepter une bonne marchandise chèrement, qu'une mauvaise à bon marché. Le reste du proverbe est, ou le respand en chemin.*

- * notre Pain est tendre ou notre pain ne se gaste pas. i. *retirez vous, vous n'avez que faire de nous venir escorniffler. Le commencement dit, Dieu vous soit en aide. vulg.*

De la forme globale qui est scindée en deux parties, le lexicographe n'en traite qu'une seule. *Qui bon l'Achepte bon le boit et notre Pain est tendre ou notre pain ne se gaste pas* constituent les unités de traitement, qui ne sont qu'une partie de la forme typographiée en caractères romains.

Au sein de l'unité de traitement, le mot vedette, qui est une unité simple, est marqué typographiquement par la majuscule et permet le classement des unités selon l'ordre alphabétique. Dans les exemples suivants, Allant, Lever, et Couleur, sont trois vedettes.

Un Allant. i. *un finet, un rusé.*

Lever boutique, *par Metaph. i. se marier.*

- * la Couleur nous en demeurera dans le ventre, *cela se dit, l'ors qu'on boit du vin rouge, plus volontiers que du blanc.*

Comme l'indique Oudin dans son avertissement, pour les unités polylexématiques, plusieurs mots peuvent potentiellement être des vedettes : *si cherchant un quolibet ou proverbe on ne le trouve en un lieu, il le faudra chercher autre-part ; par exemple, il semble à mon ventre que le Diable ait emporté mes dents, celui-cy se peut mettre à dents, Diable & ventre : observez la même chose pour tous les autres, & principalement où il y a un substantif & un verbe* (Oudin 1640, « Avertissement »). Le lexicographe utilise dans ce cas un système de renvois. Notons que les unités polylexématiques ne sont jamais

traitées comme des sous-entrées d'unités simples. Il n'y aura donc jamais une hiérarchisation du type entrée/sous-entrée. La vedette relève du procédé de classement macrostructurel alors que l'unité de traitement relève proprement de la nomenclature. D'autres éléments, comme les variantes graphiques et lexicales, que nous n'étudierons pas ici, doivent également être prises en compte dans l'analyse de ce champ informationnel.

4.4. Corps

4.4.1. Sens et valeur d'emploi

Le champ informationnel « sens et valeur d'emploi » est, comme nous l'avons déjà dit, le plus fréquent dans le corps des articles. Au sein de celui-ci, nous distinguons deux sous-classes principales : la définition proprement dite, que nous appelons « définition sémantique », et la définition d'emploi, qui précise la valeur d'emploi d'une unité. Il peut y avoir plusieurs sens et valeurs d'emploi.

4.4.1.1 Définition sémantique

La définition sémantique est en italique et suit la forme, mais pas toujours directement. En effet, elle est souvent introduite par la copule abrégée *i.*, mise pour *id* est, ou par une copule développée du type *c'est pour dire, c'est une sorte de...*

Donner Air. i. *publier, déclarer, mettre au jour.*

Jouer à l'Abbe, *c'est une sorte de jeu où il faut imiter celui qui passe devant les autres en tout ce qu'il fait.*

j'aimerois mieux le Licol que la beste, *pour dire que l'on aimeroit mieux la chaisne d'or, ou les perles que porte une femme, que sa personne mesme.*

Il existe différents types de définitions sémantiques, illustrés par les exemples ci-dessus : les définitions synonymiques et les définitions périphrastiques. Elles peuvent être accompagnées de marques rhétorico-sémantiques. Le plus répandu est le lien métaphorique qui est noté soit en pleines lettres par *metaphore*, soit en abrégé *metaph.* Sa position est variable, soit avant soit après la définition sémantique proprement dite.

Les tenans & Aboutissans. i. *Metaph.* *Tout ce qui depend d'un affaire.*
*tailler de la Besogne. i. *donner beaucoup à faire, donner du travail ou de la peine. Metaph.*

4.4.1.2. Définition d'emploi

Il est essentiel de distinguer, dans le dictionnaire des *Curiositez*, la définition sémantique de la définition d'emploi. Celle-ci, tout comme la définition sémantique, est en italique. Elle est toujours introduite par une copule du type *se dit lorsque*, *se dit pour se moquer*, etc. Elle ne sera, par contre, jamais introduite par la copule abrégée *i.*, ce qui prouve qu'Oudin distingue clairement les deux types de définition, comme le montre l'exemple suivant.

* il mourroit plustost un bon Chien de berger. i. *un honneste homme mourroit plustost, qu'un coquin ou meschant, cela se dit, lors que l'on a peur qu'une personne meurt dont il ne faut pas faire grand estat.* vulg.

Il existe deux types de définition d'emploi : l'une décrit la situation dans laquelle s'emploie l'unité définie, c'est-à-dire le contexte extralinguistique, l'autre décrit la valeur pragmatique de cette unité.

- * mouchez la Chandelle, *cela se dit à un qui se met devant nostre jour, & nous empesche de voir.* vulg.
- * n'oubliez pas la Confrairie des pourceaux, *d'autres disent*, le luminaire : *on se sert de ces mots lors que quelqu'un rotte.* vulg.
- * il faudroit amener le Chien & le chat, *cela se dit pour s'excuser de mener plusieurs personnes avec soy en quelque lieu.* vulg.
- * il donne l'avoine aux Chiens, *c'est pour se moquer d'un qui se dit maistre d'hostel d'une grande maison, & n'est qu'un coquin.* vulg.

Dans les deux premiers exemples, le lexicographe décrit le contexte dans lequel l'unité de traitement doit être employée, alors que les deux suivants mettent en évidence sa force illocutoire.

Définition sémantique et définition d'emploi participent toutes deux au programme lexicographique établi par Oudin pour ses *Curiositez*. Il s'agit de deux manières différentes d'aborder le contenu de l'unité, qu'elle soit simple ou polylexicale. Il est remarquable que la définition d'emploi soit majoritairement utilisée pour des unités de type phrastique. Cette remarque d'ordre métaléxicographique nous conduit à mener une analyse lexicologique plus générale invitant à considérer les unités polylexicales phrastiques de manière particulière.

4.4.2. Exemple

Le corps peut comporter un ou plusieurs exemples qui montrent l'emploi de l'unité de traitement dans un ou plusieurs contextes linguistiques particuliers. Ceux-ci, contrairement aux éléments appartenant au champ informationnel « sens et valeur d'emploi », ne sont pas métalinguistiques. L'exemple est un discours en mention qui montre le mot en usage. C'est cette différence de statut qui nous a conduite à en faire une catégorie différente.

Aller, *pour, estre bien seant*, v.g. ce rabat va bien ; &c.

* à Jean Guerin. i. *cecy se dit de toutes sortes de choses mal faites ou de mauvaise grace* ; v.g. La fille à Jean Guerin, & ainsi des autres. vulg.

L'exemple est parfois introduit par l'abréviation latine *v.g.* Il se trouve souvent à la fin de l'article, et toujours en caractères romains. Notons que les exemples sont exclusivement des exemples forgés par le lexicographe.

4.4.3. Marque d'usage

Outre la marque diaphasique systématique étoile portant sur l'ensemble d'un article (cf. 4.2), le corps de celui-ci peut comporter des marques d'usage de diverses natures. Nous nous limiterons ici à donner un rapide aperçu de la manière dont celles-ci se présentent.

Dans l'exemple suivant, la marque diachronique, en caractères romains, apparaît dans une position bien particulière, distincte de la

périphrase définitoire : elle se situe en fin d'article et est séparée par un point.

jouer à Boute-hors. i. *se chasser l'un l'autre*. Antique.

La marque diaphasique qualifiant le langage enfantin³ est placée soit directement après la forme, soit à la fin de l'article.

* faire Caca. i. *descharger le ventre ; mot d'enfant*.
un Dada, *mot d'enfant*. i. un cheval.

Bien qu'elle soit en italique, elle se distingue clairement du champ informationnel « sens et valeur d'emploi », dont elle est séparée par un point ou un point-virgule. Cependant, il faut noter que certaines marques sont parfois mal dégagées des autres champs informationnels du corps. Elles peuvent intervenir, entre autres, dans la rubrique « sens et valeur d'emploi ». Dans l'exemple suivant, la marque diatopique, à *Paris*, apparaît comme une incise au sein de la périphrase définitoire.

* Il est de l'Abbaye de Lonchamp, il tient des Dames, *cela se dit à Paris d'un homme qui aime les femmes*. vulg.

Le champ informationnel « marque d'usage », comme le montrent les divers exemples ci-dessus, ne se matérialise pas toujours

3. Nous qualifions ces marques de « diaphasiques », car, à la suite de J. Rey-Debove (1971 : 100), nous considérons le langage enfantin comme une sous-catégorie du langage familier.

de la même façon : soit il se présente de manière séparée des autres champs informationnels, soit il se fond dans l'un d'entre eux.

4.4.4. Motivation du signe

Lorsque l'unité de traitement est polylexématique, le lexicographe explicite souvent la motivation du signe. Ce champ informationnel est toujours en italique et possède une position variable, généralement avant ou après la partie « sens et valeur d'emploi ». Cette explication peut consister, pour les unités qui ont un sens figuré, en la description de la réalité visuelle extralinguistique à l'origine de l'image véhiculée dans la séquence.

- * Ange de Greve. i. *crocheteur, gaignedenier.*
C'est à cause de leurs crochets qui forment comme des ailes. vulg.

L'explication du lien entre forme et sens peut également se faire grâce à la mise en évidence du jeu de mots, plus ou moins explicite, présent dans la séquence, comme c'est le cas dans les deux exemples suivants.

- * les Cordonniers font des souliers, & les Tailleurs des robes, *c'est une sottise allusion du mot desrobent à des robes, pour dire que les tailleurs sont larrons.*
- * nous ne manquons pas de Raiponces, *c'est une allusion à responses, pour dire que nos valets nous respondent insolemment & mal à propos ; on y adjouste, il ne faut point aller aux Hales.*

Il y a, dans ces cas, une déconstruction de l'unité complexe.

L'explication de la motivation du signe peut aussi être présente dans le cas des unités non polylexicales. Elle se fait soit par la décomposition du signe linguistique, si celui-ci est morphologiquement décomposable, soit par un recours à l'étymologie-origine. Dans ce cas, le lexicographe ne cherche pas à faire de l'étymologie pour elle-même mais à expliquer le lien entre la forme et le sens, c'est-à-dire la motivation du signe, en ayant recours à la langue source.

Habler. i. *parler beaucoup. Le mot vient de hablar Espagnol.*

L'analyse métalexographique amène à considérer la description de l'image figurée dans la séquence, l'explicitation du jeu de mot et le recours à l'étymologie comme appartenant à un même champ informationnel dans le programme lexicographique des *Curiositez*.

Ces quelques analyses nous ont permis de montrer, d'une part, des spécificités des champs informationnels des *Curiositez* et, d'autre part, des particularités organisationnelles. Les champs informationnels sont, pour la plupart, assez spécifiques, comme par exemple « * », « sens et valeur d'emploi », « motivation du signe ». Ceci tient avant tout au fait qu'il s'agit d'un dictionnaire spécialisé destiné aux étrangers. Le programme microstructurel est donc nécessairement différent de celui d'un dictionnaire général. De plus, la structuration de ces champs informationnels est, elle aussi, particulière, comme le montrent les matérialisations variables de certains champs informationnels, tel que celui des marques d'usage.

5. De l'analyse de la microstructure à l'édition informatisée

C'est à partir d'une analyse rigoureuse du dictionnaire que nous envisageons l'informatisation des *Curiositez*. L'utilisation de versions informatiques de recueils offre incontestablement de multiples avantages. Outre l'accès facilité à l'ouvrage, le principal tient aux perspectives d'exploitation. Ainsi, T. R. Wooldridge (1993) souligne la linéarité de la lecture qu'impose l'imprimé au chercheur et montre que *l'ordinateur, s'il ne peut lui donner l'omniscience, lui offre néanmoins l'ubiquité, en lui permettant de convertir son corpus en base de données textuelles et ainsi de faire des lectures verticales qui traversent tout le texte*. La conversion du corpus en une base de données informatisée et augmentée des interprétations et analyses que nous proposons nous paraît constituer un avantage certain, non seulement pour notre recherche personnelle, mais aussi pour tout linguiste, chercheur ou curieux qui serait amené à consulter les *Curiositez*.

5.1. Une édition électronique des *Curiositez* (éd. Redon)

Il existe une version électronique des *Curiositez* parue aux éditions Redon en 1999. Il s'agit d'une version sur cédérom de l'*Atelier historique de la langue française* (AHLF 1999). L'AHLF offre l'accès à sept dictionnaires anciens : les *Curiositez françaises* d'Antoine Oudin (1640), le *Dictionnaire universel* d'Antoine Furetière (1690), le *Dictionnaire de l'Académie française* (éd. 1762), le *Dictionnaire philosophique* de Voltaire et compléments (1765), le *Dictionnaire universel des synonymes* de Guizot (1822), le *Dictionnaire de la*

langue française d'Emile Littré (1872 et supplément de 1877) et le *Dictionnaire de La Curne de Sainte-Palaye* (1876). Cette version électronique des *Curiositez* offre plusieurs avantages qui sont essentiellement au nombre de trois (cf. Delcourt 2002⁴). Le premier tient à la possibilité d'effectuer des recherches en plein texte : celle-ci peut se faire sur base de mots (ou de début de mots) ou sur des cooccurrences. Le deuxième est plutôt d'ordre pratique. En effet, l'utilisateur peut opérer des copier-coller afin de récupérer les données. Le troisième relève de la possibilité de passer rapidement d'un dictionnaire à l'autre grâce à des liens hypertextes. Mais cette édition souffre de plusieurs défauts. Le premier défaut vient du fait que l'éditeur ne mentionne pas clairement quelle version des *Curiositez* il a choisi d'éditer : celle de 1640, première édition chez Antoine de Sommaville à Paris, ou bien celle de 1656 imprimée à Rouen, ou bien celle qui se trouve à la suite du *Dictionnaire de La Curne de Sainte-Palaye* (1876). En consultant la liste des titres des dictionnaires présents dans l'AHLF, le lecteur pense avoir accès à l'édition originale de 1640, étant donné qu'il s'agit de la date qui figure à côté du titre du recueil et du nom de l'auteur (*Curiositez françaises* d'Antoine Oudin 1640). Il n'est pas fait mention des autres éditions. Cependant, la préface des *Curiositez* à laquelle donne accès la version sur cédérom est en réalité la notice sur Antoine Oudin rédigée par Léopold Favre⁵ dans le dictionnaire de La Curne. Le lecteur comprend qu'il a affaire à la version des *Curiositez* telle qu'elle se trouve à la suite de ce dictionnaire. Le choix de cette version

4. Bien que le compte rendu de Ch. Delcourt porte principalement sur la version électronique du Littré, L'AHLF permet plus de possibilités de recherche pour ce dictionnaire que pour les *Curiositez* d'A. Oudin.

5. Élaboré par Jean-Baptiste de La Curne de Sainte-Palaye, né en 1697, le *Dictionnaire historique de l'ancien langage français* ou *Glossaire de la langue française depuis son origine jusqu'au siècle de Louis XIV* n'a été publié qu'au 19^e siècle par Léopold Favre, érudit polygraphe.

n'est nullement expliqué et ne nous semble pas le meilleur. En effet, dans cette introduction, Favre indique que certaines expressions, peu nombreuses certes, ont été censurées :

Antoine Oudin s'est attaché à réunir les proverbes, les adages qui faisaient la joie et les délices de nos aïeux. Beaucoup sont gaulois, et même très gaulois ; ils ont une grande parenté avec les causeries de Rabelais ; mais enfin c'était le langage de nos pères, et si les paroles étaient libres, leurs actes certes valaient mieux que nos mœurs du jour. Cependant nous avons cru devoir supprimer quelques quolibets qui nous ont paru par trop salés ; nous avons bien peu élagué, et si nous n'avons pas opéré de plus grand ravage dans les pages de ce petit livre, c'est que nous savons qu'un Glossaire s'adresse à des personnes dont l'esprit mur et sérieux ne peut recevoir aucune atteinte d'expressions libres, qui ne retracent point d'actions obscènes. (L. Fav.) »

Cette version des *Curiositez* n'est donc pas fidèle à l'œuvre originale telle qu'elle a été pensée par Antoine Oudin. Le deuxième défaut est lui aussi lié à un manque de fidélité par rapport à l'original. En effet, l'*AHLF* supprime le symbole * présent devant certaines entrées du dictionnaire. Le renseignement n'est pourtant pas sans importance. Il a une valeur démarcative essentielle, comme nous l'avons montré au point 4.2. La suppression de cette marque implique une perte de sens dans le système microstructurel prévu par le lexicographe⁶. Le troisième défaut consiste en la présence de plusieurs erreurs. Celles, pourtant évidentes, commises dans la version « La Curne » ne sont pas corrigées. Ainsi, par exemple, *Qui bon l'Achepte bon le boit et par manière d'Aquit* sont présentés comme ne formant qu'un seul article. De nouvelles erreurs graphiques sont introduites.

6. La répartition des marques est quelque peu différente dans la version « La Curne » par rapport à l'originale.

Le non-respect des caractères romains et italiques originaux, ce qui enlève une partie du sens, constitue le quatrième défaut.

Si l'édition des *Curiositez* dans l'*AHLF* est accompagnée d'une première série d'outils et offre déjà une première base sur laquelle le chercheur peut mener des recherches, elle ne peut être qualifiée d'édition scientifique. Cette édition est à réaliser. Nous proposons d'en jeter ici les premiers jalons.

5.2. Nos principes éditoriaux

L'édition informatisée que nous voulons proposer est celle de la version de 1640 imprimée chez Antoine de Sommaville à Paris et qui a fait l'objet d'une réimpression chez Slatkine en 1993. D'un point de vue formel, il nous semble important de respecter le principe de fidélité absolue au texte de référence. Le texte est donc saisi manuellement par nos soins le plus fidèlement possible. Les graphies originales et les marques de ponctuation sont maintenues. Nous n'intervenons qu'en cas de coquilles manifestes (ex. confusion entre *le n* et *le u* dans *convent* mis pour *couvent*) ou bien lorsque l'original est peu lisible.

Le principe de fidélité doit lui aussi être appliqué aux contenus. Nous désirons que cette édition informatisée mette en évidence la structure spécifique de chaque article. Ceci rendra possible des recherches ciblées au sein des différents champs informationnels (recherche par types de formes, par types définitionnels, par marques d'usage...) intéressantes tant pour notre étude linguistique que pour tout chercheur qui aurait à consulter les *Curiositez*. Le repérage d'un

certain nombre d'éléments dans le texte lexicographique permettrait la comparaison et l'analyse exhaustive des données.

5.3. Pistes pour le balisage

Le balisage du texte du dictionnaire constitue la technique de base en vue de son informatisation. I. Leroy-Turcan (1999) a opposé le balisage formel, qui ne prend en compte que les aspects de la présentation matérielle du texte, au balisage analytique qui fait intervenir l'appréciation du métalexigraphe. Ce type de balisage, qui correspond à notre volonté éditoriale, requiert une expertise tant sur le plan métalexigraphe que sur le plan informatique.

Utilisé par Isabelle Leroy-Turcan mais surtout par A. Tutin et Ch. Wionet (2001), le balisage analytique⁷ se caractérise par une étude approfondie de la microstructure afin de fournir une interprétation précise du fonctionnement du dictionnaire informatisé. Dans ce cas, toutes les informations sont balisées : des jalons sont placés sur l'ensemble des champs informationnels de l'article. Les informations typographiques sont elles aussi balisées. Ce balisage analytique traduit la volonté d'offrir une grille de lecture quasi exhaustive au lecteur, en procédant à un découpage systématique des articles du dictionnaire. Le risque est de fausser le sens du texte en donnant une interprétation subjective des données. Afin d'éviter cet écueil, une

analyse métalexigraphe rigoureuse s'avère indispensable. Dans le cas des *Curiositez*, les différents champs informationnels que nous avons identifiés doivent être précisément balisés. Il est également intéressant d'avoir recours à des attributs types. Ceux-ci permettraient, par exemple, de préciser la nature de la forme donnée en entrée, simple ou polylexicale, ou encore le type de définition choisi pour le traitement lexicographique d'une unité.

Le chercheur-éditeur doit également s'interroger sur le langage de balisage à adopter. Il nous semble essentiel que notre édition soit diffusée sur Internet et qu'elle soit donc réalisée à l'aide de normes en vigueur dans la communauté universitaire. Un balisage normatif, c'est-à-dire propre à un langage ayant fait l'objet de recommandations et d'études en vue de devenir un standard, permet une portabilité plus grande qu'un balisage dit propriétaire, c'est-à-dire propre à un logiciel de traitement de corpus donné. Dès lors, afin de favoriser l'échange, nous privilégions l'emploi du langage XML (*Extensible Markup Language*), langage de toutes les attentions scientifiques actuelles. Il a été conçu pour pallier les problèmes rencontrés par les normes déjà existantes et s'attache particulièrement à la description de la structure logique des documents. Il est souple et riche, ce qui le rend particulièrement adapté à la restitution de la structure hiérarchique d'un dictionnaire, tel que nous voulons le faire.

Dans ce souci de partage, il est intéressant de prendre en compte les recommandations de la TEI (*Text Encoding Initiative*), qui ont pour but de promouvoir des formats communs d'encodage. Ces recommandations proposent un ensemble de conventions de codage

7. Le balisage analytique a été employé pour l'informatisation du *Dictionnaire étymologique ou Origine de la langue française* de G. Ménage (1694) par I. Leroy-Turcan et pour le projet d'informatisation du *Dictionnaire universel de Furetière revu par Basnage de Beauval* (1702) par Ch. Wionet et A. Tutin.

utilisables dans une grande variété d'applications : publication électronique, analyse littéraire et historique, lexicographie, traitement automatique des langues, recherche documentaire, hypertexte, etc. (Ide et Veronis 1996 : 5). Nous nous sommes exclusivement penchée sur celles proposées pour les dictionnaires papier. Les principaux constituants de la TEI sont assez généraux et semblent donc, de prime abord, pouvoir convenir pour le balisage de dictionnaires anciens à la structure parfois moins systématique que celles des contemporains. Cependant, les standards, comme la TEI, ne permettent pas de faire ressortir l'originalité du programme lexicographique tel qu'il a été conçu par Oudin. Ainsi, par exemple, il n'existe pas de balise TEI permettant de distinguer la forme de l'unité de traitement, ni de la vedette. Il en est de même pour la définition d'emploi qui est à distinguer de la définition sémantique stricte, tout en montrant l'appartenance des deux au champ informationnel « sens et motivation du signe ». La rubrique « marque d'usage » appartenant au corps de l'article est, elle aussi, à distinguer de la marque diaphasique codifiée, étoile, ce que ne permettent pas les balises proposées par la TEI. Le risque serait dès lors de forcer les articles des *Curiositez* à entrer dans un moule prédéfini et de proposer une grille de lecture qui dénature le texte original. De plus, la TEI ne permet pas de rendre compte des particularités organisationnelles. Pour ne prendre qu'un exemple, selon ce standard, la balise <def> est un élément terminal, ne pouvant pas faire l'objet de décomposition, alors que, comme nous l'avons montré, dans les *Curiositez*, le champ informationnel « marque d'usage » peut être contenu dans la définition. Ces quelques réflexions nous invitent donc à privilégier des balises établies par nos soins et prévues pour épouser parfaitement l'organisation microstructurelle. Nous pouvons,

néanmoins, envisager de garder certains noms de balises et d'attributs assez généraux de la TEI.

6. Conclusion

Par cet article, nous avons voulu présenter les analyses préalables à l'édition informatisée du dictionnaire des *Curiositez françaises* (1640) d'Antoine Oudin, véritable trésor lexicographique pour la connaissance de la langue vulgaire de la première moitié du 17^e siècle.

Après avoir mis en évidence les caractéristiques du microtexte des dictionnaires anciens, nous nous sommes attachée à identifier différents champs informationnels des *Curiositez*. Par cette étude, nous avons voulu mettre au jour les principales spécificités microstructurelles de ce dictionnaire. Nous avons montré que, sous une apparente irrégularité, se dessinait une certaine structure systématique. Nous avons ainsi pu dégager de grands champs informationnels et établir le schéma tripartite valable pour tout article des *Curiositez*. De nombreux champs informationnels sont apparus assez spécifiques. Les particularités du programme lexicographique sont dues avant tout à la nature même du dictionnaire étudié : il s'agit d'un dictionnaire spécialisé destiné aux étrangers. Cette étude a montré l'importance de distinguer le programme lexicographique d'Oudin, qui est cohérent et assez systématique, de la réalisation concrète de celui-ci, qui est, quant à elle, beaucoup plus souple et moins régulière.

Nous avons ensuite présenté notre projet d'édition informatisée des *Curiositez*. Après avoir montré l'absence de véritable édition scientifique, nous avons exposé les grandes réflexions qui doivent sous-tendre une telle édition. Selon nous, avant toute chose, elle se doit de mettre en évidence la structure de chaque article. Pour ce faire, nous avons choisi un balisage analytique fin qui consiste à placer des jalons autour des différents champs informationnels. Ceci implique au préalable la réalisation d'une analyse métalexigraphique rigoureuse permettant de les identifier. Le balisage à mettre en place pour les *Curiositez* doit respecter les spécificités microstructurelles et présenter dès lors une certaine souplesse.

D'un point de vue plus technique, nous avons montré l'intérêt d'utiliser un langage de balisage standard, par rapport à un langage propriétaire, afin de favoriser la portabilité. Le XML est apparu comme la norme en vigueur pour l'informatisation des dictionnaires anciens. Nous nous sommes penchée sur les recommandations pour le balisage des dictionnaires, telles que proposées par la TEI. Cependant, nous avons montré qu'elles ne permettent pas de rendre compte des spécificités de notre dictionnaire. Nous privilégions donc l'établissement de balises spécifiques.

Bibliographie

Sources primaires

AHLF 1999, *L'Atelier historique de la langue française : histoire des mots, du haut Moyen Âge au XIX^e siècle*, Marsanne, Redon.

LOUDIN (A.) 1640, *Curiositez françaises, pour supplément aux dictionnaires ou recueil de plusieurs belles propriétés, avec une infinité de proverbes et quolibets, pour l'explication de toutes sortes de livres*, Paris, Antoine de Sommaville. [id. Genève, Slatkine reprints 1993.]

Sources secondaires

BUCHI (E.) 1996, *Les structures du <Französisches Etymologisches Wörterbuch>*, Tübingen, Max Niemeyer Verlag.

DEL COURT (Ch.) 2002, « L'atelier historique de la langue française : histoire des mots, du haut Moyen Âge au XIX^e siècle », in *Revue belge de philologie et d'histoire*, Bruxelles, 80/3, pp. 1057-1061.

IDE (N.) et VERONIS (J.) 1996, « Présentation de la TEI : Text Encoding Initiative », in *Cahiers GUTenberg*, Orsay, 24, pp. 4-10.

LE ROY-TURCAN (I.) 1999, « Balisage formel ou balisage fin pour les dictionnaires anciens informatisés : objectifs et implications méthodologiques », in *L'informatisation des dictionnaires anciens* (Actes du colloque-atelier international DictA1998), T. R. Wooldridge éd., Toronto, Siehleda. (Texte publié en ligne http://homes.chass.utoronto.ca/~wulfric/siehlda/dicta1998/ilt2_tab.htm) (consultation : janvier 2013)

PETREQUIN (G.) 2009, *Le « Dictionnaire françois » de P. Richelet (Genève, 1679/1680). Étude de métalexigraphie historique*, Leuven, Peeters.

PFISTER (M.) 1989, « L'importance d'Antoine Oudin pour la lexicographie française et italienne », in *La lingua francese nel seicento*, M. Gross éd., Bari-Paris, Adriatica-Nizet, pp. 81-103.

REY (Ch.) 1999-2000, « Informatisation des dictionnaires anciens : l'exemple du métalangage grammatical dans le *Dictionnaire françois* de César-Pierre Richelet », mémoire, Université de Provence.

REY (Ch.) 2003, « Une expérience d'informatisation des dictionnaires anciens : les articles portant sur les sons dans l'*Encyclopédie méthodique* (1782-1832) », présentation-poster, Université de Provence. (Publication en ligne http://www.u-picardie.fr/LESCLaP/rey/presentation_poster_aix.pdf) (consultation : janvier 2013)

REY (Ch.) 2004, « Le balisage souple ou flottant : une piste pour l'informatisation des données encyclopédiques anciennes », Site du centre de recherches METAlexicographiques et DIctionnaires Francophones (Metadif). (publication en ligne : http://www.u-picardie.fr/LESCLaP/rey/Reyc_balisage-souple.pdf) (consultation : janvier 2013)

REY (Ch.) et ZAOUÏ (C.) 2004, « Balisage XML "ciblé" : Une nouvelle approche dans l'informatisation des corpus », in *Actes de la conférence internationale sur la fouille de texte*, M.-H. Antoni et Fr. Yvon éd., Paris, École nationale supérieure des télécommunications (ENST), pp. 121–133.

REY-DEBOVE (J.) 1971, *Étude linguistique et sémiotique des dictionnaires français contemporains*, La Haye/ Paris, Mouton.

TEI Consortium, TEI P5 : *Guidelines for electronic text encoding and interchange* (2.3.0.), <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html>. (consultation : janvier 2013)

TUTIN (A.) et WIONET (Ch.) 2001, *Pour informatiser le Dictionnaire Universel de Basnage (1702) et Trévoux (1704) ; approche théorique et pratique*, Paris, Honoré Champion.

WOOLDRIDGE (T. R.) 1993, « Le flou en informatique textuelle », in *Texte*, Toronto, 13/14, pp. 275-289.

WOOLDRIDGE (T. R.) 1997, « Baliser un texte, c'est le penser : le cas du Dictionnaire de l'Académie française ». [Version révisée d'une communication faite à Paris, le 23 mai 1997, dans le cadre d'une journée d'études organisée par le Groupe d'Études en Histoire de la Langue] (Texte publié en ligne <http://homes.chass.utoronto.ca/~wulftric/articles/gehlf597/>) (consultation : janvier 2013)

WOOLDRIDGE (T. R.) 1999, « L'informatisation du Dictionnaire de l'Académie française », in *L'informatisation des dictionnaires anciens* (Actes du colloque-atelier international DictA1998), T. R. Wooldridge (éd.), Toronto, Siehlda. (Texte publié en ligne http://homes.chass.utoronto.ca/~wulftric/siehlda/dicta1998/trw_acad.htm) (consultation : janvier 2013)

Le corpus Babeliris : d'une méthodologie de constitution aux principales caractéristiques linguistiques

Ward VAN de VELDE
Centre de recherche Termisti
Institut supérieur de traducteurs et interprètes
Haute École de Bruxelles

Résumé

Cet article décrit le processus d'encodage du corpus *Babeliris*, réunissant des textes médico-administratifs du réseau hospitalier *IRIS*. Concrètement et sans revenir sur les modalités de collecte des textes, il présente les balises *TEI-P5* qui ont servi à encadrer les données textuelles variables, à partir desquelles peuvent naître des sous-corpus. Dans ce cadre, il explique le fonctionnement de l'interface créée à cet effet. Par ailleurs, il soulève la problématique du choix d'un aligneur multilingue *ad hoc* et, partant, d'un gestionnaire de mémoire de traduction, en exposant les avantages et les inconvénients de trois logiciels testés. Enfin, il décrit les principales caractéristiques linguistiques du pan français du corpus, dans le but de pouvoir juger de la représentativité de celui-ci dans le vaste monde des corpus de langue spécialisée et de langue générale.

Mots-clés :

linguistique de corpus, médecine, *TEI-P5*, alignement, concordance, mémoire de traduction.

1. Introduction : de la recherche *Babeliris*¹ à la naissance d'un corpus original

S'agissant de l'accès aux soins de santé, la législation belge veut que *la communication avec le patient se déroule dans une langue claire* (Ministère 2002). Ce principe présuppose notamment que le médecin s'exprime à sa patientèle dans une langue simplifiée. Le respect de cette disposition légale, inspirée de l'*Evidence based medicine* (Sackett *et al.* 1996), devrait particulièrement s'imposer au contact d'une patientèle allophone et/ou étrangère. Or, dans les faits, on constate que la loi ne prévaut pas toujours. Ainsi, à Bruxelles – terrain de nos recherches –, tous les patients ne sont pas égaux devant la démarche d'information médicale, et ce, en dépit des nombreuses actions menées dans le but d'aplanir les inégalités sociales en matière de santé (Boïteké 2011 : 9-10).

Les interactions orales jouent certes un rôle crucial dans la relation patient-médecin, mais une mauvaise compréhension de l'écrit médical peut aussi affecter la santé du patient, même si *l'écrit* ne peut détrôner *l'oral* (HAS 2008 : 6-7). Le centre de recherche TERMISTI, auquel nous ressortons, et le *Centrum voor Vaktaal en Communicatie* ont clairement compris ce double enjeu et se sont associés pour mener le projet Babeliris. Celui-ci a pour objectif d'améliorer la communication interculturelle en milieu hospitalier, appliquée plus spécifiquement aux hôpitaux publics bruxellois du réseau IRIS. Il se ramifie en deux volets complémentaires dans le but d'embrasser les facettes orale

et écrite de la communication, respectivement traitées par les deux institutions susmentionnées.

Très concrètement, s'agissant du volet écrit, nous avons l'ambition de produire trois outils à valeurs formative et éducative destinés aux membres du réseau IRIS. Ce matériel didactique et pédagogique d'accompagnement du personnel soignant comprendra un *Guide de bonnes pratiques de rédaction* en français et en néerlandais destiné aux rédacteurs et aux traducteurs des hôpitaux, une banque de textes français-néerlandais alignés et une base de données terminologique multilingue, qui clarifiera le vocabulaire médico-administratif susceptible de poser problème à une patientèle ne maîtrisant ni les langues de la Région, ni l'actuelle *lingua franca* : l'anglais.

Afin d'appréhender la langue écrite en usage dans les hôpitaux IRIS au départ de laquelle nous concevrons ces instruments, nous avons appliqué les principes défendus par la linguistique de corpus, souscrivant à une approche résolument *corpus driven* (Tognini-Bonelli 2001 : 84-99). Dans ce cadre, il nous paraît essentiel d'exposer la méthodologie qui a présidé à l'élaboration de notre collection de textes, dans le but notamment d'assurer une certaine transparence pour les futurs utilisateurs de notre corpus (Tymoczko 1998 : 655). Par ailleurs, même si nous ne prétendons pas à l'exemplarité, il n'en demeure pas moins que nous espérons que le *modus operandi* exposé ici pourra inspirer d'autres chercheurs conduisant une recherche similaire.

1. Plus d'informations sur <http://babeliris.be>.

2. Vers une méthodologie d'encodage des documents

2.1. Du traitement primaire des textes à leur XMLisation : bref aperçu

Dans cette section, nous ne reviendrons ni sur les modalités de collecte documentaire ni sur les contraintes juridiques et/ou déontologiques que nous avons observées dans ce cadre (Van de Velde *à paraître*). Rappelons toutefois que les textes colligés obéissent à des critères bien définis au préalable : tous traitent d'une thématique médicale et/ou administrative, ont été rédigés par des membres du réseau IRIS, sont adressés au patient et sont encore en usage. Concrètement, nous avons réuni un vaste ensemble de documents médico-administratifs diffusés en toutes langues sur divers supports, auxquels risque d'être confrontée la patientèle IRIS : brochures, dépliants, affiches, panneaux, guides, formulaires, pages *Web*, etc. Enfin, nos textes étant relativement courts, nous avons retenu les écrits dans leur totalité : ce faisant, nous nous sommes inscrit dans la tradition de Sinclair (1991 : 19) et écarté de celle de Biber, théoricien du *stratified sampling* et du *proportional sampling* (1993 : 243-248).

Confronté à des écrits hétéroclites codés en divers formats, nous avons rapidement été amené à soulever la question de l'encodage textuel et à opter pour un langage de balisage adopté par une large communauté de spécialistes et de chercheurs. Celui qui a retenu notre intérêt, en raison du caractère pérenne et interopérable conféré aux écrits qu'il code, est le langage *XML* et, plus particulièrement, une application de celui-ci : la spécification *TEI*, dans sa version *P5*

(Bauman et Burnard : 2012), plus adaptée à des exploitations linguistiques et, selon Jacquemart (*s.d.* : 8), très souvent mise en œuvre dans la sphère médicale. Néanmoins, il y a loin d'un écrit papier ou d'un texte sauvegardé en *.pdf* à sa version *TEI*, sans compter le coût en temps et en moyens de cette entreprise de conversion (Popescu-Belis 2002 : 60-61). C'est de ce constat que rend compte ce schéma, inspiré des travaux de Hareide et Hofland (2012 : 89-90) et retraçant le traitement primaire des documents jusqu'à leur *XMLisation* :

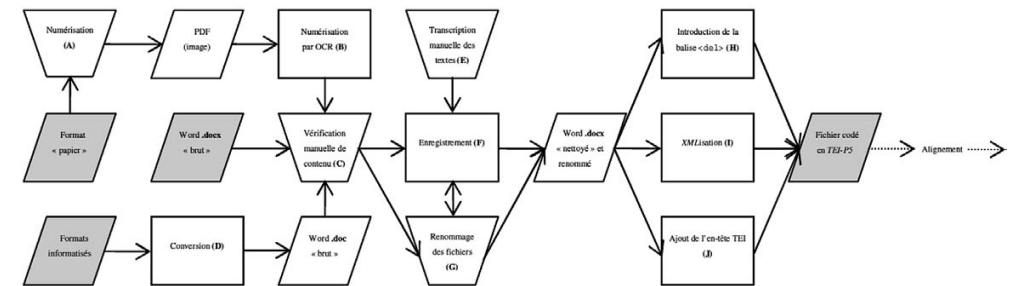


Figure 1 – Processus d'encodage documentaire :
d'un texte brut à sa version structurée

Sur fond grisé apparaissent les trois formats textuels d'entrée – c'est-à-dire ceux des documents bruts obtenus tels quels par les hôpitaux IRIS – et celui de sortie – autrement dit celui des versions textuelles codées en *TEI-P5* –, et entre ces extrêmes figure le processus analytique de traitement documentaire. Celui-ci diffère selon le support textuel, mais converge vers un même but : soumettre le corpus à de puissants algorithmes de recherche.

S'agissant des écrits papier disponibles uniquement dans ce format, nous les avons convertis en des documents électroniques exploitables. Pour ce faire, nous les avons scannés (A) – afin d'obtenir des fac-similés en format .pdf (image) – puis numérisés à l'aide d'un logiciel de reconnaissance optique de caractères (OCR) : *Readiris*TM (B). À ce travail de numérisation, long de sept mois, s'est ajoutée une minutieuse tâche de révision (C), qui avait pour objectifs de transformer les signes mal interprétés par l'OCR en des chaînes de caractères lisibles, d'une part, et d'assurer une adéquation de mise en forme entre les versions originales et numérisées, d'autre part. On sait qu'une mauvaise transcription ou un découpage textuel inexact peut biaiser les résultats d'une quelconque analyse linguistique par logiciel (Véronis 2000 : 163). Quant aux documents *Word* sous format .docx, nous en avons examiné le contenu, veillant à les dépouiller de leurs images, mais à conserver le texte qui accompagnait celles-ci (C). Les autres écrits informatisés ont tous été convertis automatiquement en fichiers *Word* .doc (D) avant de subir un traitement similaire (C). Enfin, quelques publications ont été retranscrites manuellement (E). Il convient de préciser que, pour tous les textes, nous avons maintenu à dessein les erreurs syntaxiques et orthographiques, car celles-ci participent de la difficulté à comprendre le message.

Les corps des textes une fois validés, nous les avons enregistrés au format *Word* .docx (F) afin de permettre une XMLisation automatisée de nos fichiers. À ce stade, nous les avons aussi renommés sur le modèle *Titre du document_Hôpital concerné_Langue_Public cible_Support textuel_Visée du document_Date de diffusion* (G). Faute de règles bien définies régissant l'appellation d'un document .xml

(Dufournaud et Gratsac-Legendre 2012 : 7), nous avons nommé nos fichiers de façon intuitive, veillant à définir au mieux les paramètres variables qui les caractérisent et tenant compte de l'exploitation informatique que nous envisageons de notre corpus.

La dernière étape du traitement liminaire des textes IRIS, dite d'XMLisation (I)², a consisté à convertir par lots notre banque de textes désormais codés en .docx en des documents .xml conformes à la TEI-P5, et ce, à l'aide de l'éditeur *Tei Vesta*. Dans le même temps, et toujours de manière automatisée, on a fait précéder chaque texte d'un en-tête *TEI-P5* (J), dans lequel on a injecté entre les balises *ad hoc* les sept métadonnées variables contenues dans le titre de nos documents préalablement enregistrés en .docx. Enfin, on a balisé les corps mêmes des messages – contenus entre les marques <body> – au niveau du paragraphe <p> et, à l'instar des travaux de Vihla (1998 : 75), de Sansonetti (2003 : 73-74) ou de Hareide et Hofland (2012 : 90), on y a neutralisé, à l'aide de la balise <del resp="Ward Van de Velde"> (Bauman et Burnard 2012 : 884-886 et Van de Velde à paraître), les données superflues susceptibles de biaiser toute analyse linguistique : adresses et courriels, contacts et numéros de téléphone, index et tables des matières, bibliographies et références bibliographiques... Manquant de précision, cette dernière tâche d'annotation intratextuelle mériterait toutefois une vérification manuelle, coûteuse en temps et en moyens.

2. Cette tâche a été confiée à M. Cédric Libert, jeune linguiste de l'Université catholique de Louvain-la-Neuve (Belgique) qui se spécialise en informatique.

2.2. De la sélection des balises variables à la constitution de l'en-tête *TEI*

La section précédente suppose que toute publication codée en *TEI* s'articule autour de deux éléments : un `<body>` ou corps du message, d'une part, et un `<teiHeader>` ou en-tête *TEI*, d'autre part. Cet en-tête comporte à son tour quatre éléments majeurs (`<fileDesc>`, `<encodingDesc>`, `<profileDesc>` et `<revisionDesc>`), dont une exploitation rigoureuse permettra de travailler sur le corpus de manière sélective en fonction d'un (ou de plusieurs) paramètre(s) de recherche et, partant, de générer des sous-corpus d'étude.

Ainsi, à côté des balises fixes standards utilisées pour tous les documents du corpus, nous avons dû sélectionner des balises variables, qui définissent au mieux les paramètres que nous souhaitons étudier plus spécifiquement, et que Biber (1993) comme Sinclair (1996) nomment *dimensions*³. Concrètement, il s'agit des sept métadonnées dégagées en 2.1. au moment de renommer les fichiers avant leur XMLisation automatisée, auxquelles a été ajoutée la spécification du domaine. Avant de nous engager plus avant dans leur description, il importe de noter que les informations insérées entre balises ont été rédigées en français, tandis qu'à chaque attribut a été associée une valeur en langue anglaise, tel que recommandé par la *TEI* (2012 : xliii) :

1. `<title>xxx</title>` (dans `<fileDesc>`)⁴ : titre du document francisé et, lorsque manquant, composé de toutes pièces au départ des données contenues dans le texte.

3. Zweigenbaum *et al.* (2001 : 293) et Habert *et al.* (1997 : 152-153) nous donnent plusieurs exemples de ces dimensions.

4. Cette indication entre parenthèses désigne la balise-mère à laquelle est subordonnée la balise traitée.

2. `<publisher>xxx</publisher>` (dans `<fileDesc>`) : hôpital d'où émane l'écrit.
3. `<date when-iso="xxxx"/>` (dans `<fileDesc>`) et `<change who="iris hospitals" when-iso="xxxx"/>` (dans `<revisionDesc>`) : année de diffusion du document sous la forme d'une date à quatre chiffres (de 1991 à 2012), codée 0000 par défaut.
4. `<language ident="xx">xxx</language>` (dans `<profileDesc>`) : langue de rédaction du document, spécifiée pour l'attribut à l'aide du code ISO 639-1 (2002).
5. `<channel>xxx</channel>` (dans `<profileDesc>`) : support textuel⁵, déterminé en fonction d'une catégorisation *sui generis* composée des valeurs brochure, dépliant, affiche, page Web, formulaire, ordonnance et autre (p. ex. lettre). Parfois, la frontière entre brochure et dépliant est ténue (Van Ommen et Van Kuppenveld 1995 : 59) et les critères de distinction varient selon les auteurs. Ainsi, si pour les différencier certains prennent en considération le nombre de pages (Huigen 2004 : 7), d'autres s'inspirent – tout comme nous – de leur étymon. Ainsi la brochure désignerait une *publication brochée*, autrement dit un *petit ouvrage imprimé et broché*⁶ ou *relié sommairement, après pliage des feuilles et assemblage des cahiers, avec une couverture de papier*, le dépliant faisant référence, lui, à un *prospectus plié plusieurs fois*.
6. `<interaction passive?="group">xxx</interaction>` (dans `<profileDesc>`) : lecteur modèle du document. Il peut s'agir du **patient** ou d'un **responsable** de celui-ci, d'un acteur **non défini** ou **autre** (p. ex. lettre d'un médecin à son confrère à signer par le patient).
7. `<purpose type="xxx"/>` (dans `<profileDesc>`) : visée du document. Les valeurs accordées à l'attribut sont soit "inform" en présence d'un document de type descriptif, soit "persuade" pour les écrits de type injonctif, p. ex. ceux précisant un traitement à suivre ou des dispositions à respecter avant un examen invasif.

5. Nous présumons ici une variation linguistique liée au support. La création de cette rubrique nous a par ailleurs été demandée par les responsables des hôpitaux IRIS, désireux de connaître les paramètres de rédaction de leurs pages Web.

6. Ces définitions sont tirées du *Nouveau Petit Robert* sur cédérom (2009).

7. L'attribut "passive" peut surprendre au premier aspect, alors qu'en réalité il précise tout simplement « le nombre de participants passifs (ou récepteurs) à qui un texte est destiné » : "group" dans notre cas, c'est-à-dire « un ensemble indéfini mais limité de personnes, comme une conférence par exemple » (Bauman et Burnard 2012 : 1048-1049).

8. `<term>xxx</term>` (dans `<keywords>`, inclus dans `<profileDesc>`) : service et unité auxquels se rattache l'écrit. Admettant le principe selon lequel toute classification s'avère délicate à établir en raison de son caractère intuitif (McEnery *et al.* 2006 : 21) et celui selon lequel « *les catégories ne s'excluent pas forcément [...] ou semblent être en intersection* » (Habert 2000), nous avons opté pour une typologie de *genre* reposant sur le domaine, en raison de sa légitimité *per se* (Lee 2001 : 37-39, Biber 1993 : 245 et Sinclair 1996 : 7). Concrètement, la notion de *genre* s'appuie sur des critères extralinguistiques, externes à la production textuelle, autrement dit sur des paramètres situationnels et fonctionnels de l'écrit : contexte ou mode de production, situation d'énonciation, lecteur modèle, mode de publication, fonction ou visée et, dans le cas présent, thème ou domaine. Ayant pour ambition de nous rapprocher de la réalité du réseau IRIS et d'inclure des aspects tant administratifs, (para)médicaux, infirmiers que sociaux dans une seule taxinomie, nous avons considéré les domaines suivants (Van de Velde à paraître) :

TABLE 1. DOMAINES RETENUS POUR LE CORPUS *BABELIRIS*

SERVICES	UNITÉS
ADMINISTRATIF ET LOGISTIQUE	Accueil – admission – consultation ; archivage médical – bibliothèque ; droits, devoirs et protection du patient (service juridique et contentieux) ; hygiène hospitalière ; informatique – pages Web ; médiation – communication – relations publiques ; médiation interculturelle ; service social ; trésorerie – facturation ; non catégorisé
INFIRMIER	Soins infirmiers
MÉDICAL	Anatomie pathologique et cytogénétique ; anesthésiologie – réanimation ; biologie clinique et prélèvements ; cardiologie – chirurgie cardiaque ; chirurgie cervico-faciale et thoracique ; chirurgie mammaire et pelvienne ; chirurgie plastique, réparatrice et esthétique ; chirurgie vasculaire ; dermatologie ; endocrinologie ; gériatrie ; gynécologie – obstétrique – périnologie – prénatale ; hémato-oncologie – dépistage et prévention du cancer – cancérologie mammaire ; hépato-gastro-entérologie – chirurgie digestive et abdominale ; hôpital de jour ; immuno-allergologie ; infectiologie – maladies infectieuses ; maladies tropicales ; médecine interne ; néonatalogie ; néphrologie – dialyse ; neurologie – neurochirurgie ; ophtalmologie ; ORL ; orthopédie – traumatologie – chirurgie orthopédique ; pédiatrie ; pneumologie – tabacologie – médecine du sommeil ; psychiatrie ; radio-isotopes – médecine nucléaire ; radiologie – imagerie médicale ; recherche hospitalière – éthique ; revalidation locomotrice, neurologique et cardio-pulmonaire ; rhumatologie ; sénologie ; sexologie – urologie ; soins supportatifs, continus et palliatifs ; soins intensifs – urgences ; stomatologie – dentisterie – orthodontie – chirurgie maxillo-faciale ; non catégorisé
PARAMÉDICAL	Diététique – nutrition ; ergothérapie ; kinésithérapie – ostéopathie ; logopédie ; pharmacie ; psychologie

Très concrètement, tout en-tête TEI-P5 de notre corpus se structure comme suit. Pour des raisons liées à l'impossibilité de développer ici la totalité de notre `<teiHeader>`, nous avons décidé de présenter uniquement le jeu de balises contenues dans la description du profil `<profileDesc>` :

```
<teiHeader>
  <fileDesc> [...]
</fileDesc>
  <encodingDesc> [...]
</encodingDesc>
  <profileDesc> [...]
  <langUsage>
    <language ident="FR">français</language>
  </langUsage>
  <textClass>
    <keywords scheme="www.babeliris.org">
      <term>Service médical</term>
      <term>Infectiologie – maladies infectueuses</term>
    </keywords>
  </textClass>
  <textDesc>
    <channel>dépliant</channel>
    <constitution type="single"/>
    <derivation type="original"/>
    <domain type="medicine"/>
    <factuality type="fact"/>
    <interaction passive="group">non défini</interaction>
    <preparedness/>
    <purpose type="inform"/>
  </textDesc>
</profileDesc>
<revisionDesc status="approved">
</revisionDesc>
</teiHeader>
```

3. La genèse d'une interface génératrice de sous-corpus

Le choix des balises exposées *supra* laisse entrevoir que le langage *TEI* ne constitue pas un format *stricto sensu*, mais plutôt un vaste ensemble de recommandations destinées à coder un document

XML. En réalité, la *TEI* permet d'encoder une même information de diverses manières en fonction de l'interprétation donnée au balisage et de l'objectif de celui-ci en lien avec une étude donnée. Faute d'unicité des métadonnées textuelles, aucun logiciel développé à ce jour ne permet donc d'interpréter invariablement le jeu de (sous-)balises *XML/TEI*. Cette réalité a pour corollaire de devoir construire ou réaménager – à la lumière du travail d'analyse à effectuer – des outils intégrant la logique sémanticienne *TEI*.

En outre, malgré l'universalité supposée du langage TEI-P5, il n'en reste pas moins que certains concordanciers et outils d'analyse du discours posent des problèmes d'exploitation du balisage *.xml* et supportent uniquement des formats propriétaires (Daoust et Marcoux 2006 : 330-331). D'autres encore imposent l'usage de *DTD* particulières. Il serait toutefois ambitieux d'adapter constamment le marquage *.xml* à ces outils, sachant qu'ils s'accommodent du format de texte brut *.txt*. Comme l'atteste le tableau 2, qui met en correspondance certains des « linguisticiels » utilisés à TERMISTI avec leurs propres formats de représentation de données, le plus petit commun dénominateur des outils testés demeure le format *.txt* :

**TABLE 2. FORMATS D'IMPORTATION DOCUMENTAIRE
SUPPORTÉS PAR LES DIVERS LOGICIELS TESTÉS**

	OUTIL LINGUISTIQUE	FORMAT D'IMPORTATION TEXTUELLE REQUIS
1	AntConc	.htm, .html, .txt et .xml
2	Cordial Analyseur	.cnr, .htm, .rtf, .sam, .txt et .wp
3	mkAlign	.htm, .html, .tmx, .txt et .xml
4	NooJ	.doc, .docx, .htm, .html, .pdf, .rtf, .txt, .xml...
5	TXM	.tmx, .txt, .xml(-tei et dérivés)

Fort de ces constatations, nous avons fait développer⁸ un outil capable de générer des sous-corpus codés en *.txt* au départ de nos documents XMLisés. Donner naissance à ce genre d'outil est intéressant à plusieurs égards : pour comparer certains paramètres de rédaction de divers sites hospitaliers, pour étudier des écrits d'un seul service quel que soit l'hôpital dont ils émanent, etc. La constitution de ces sous-corpus repose bien entendu sur l'interrogation des *dimensions* variables de l'en-tête *TEI*, isolées en 2.1. (à l'exception du titre de l'écrit, unique par principe) : service, unité, site hospitalier, langue, lecteur modèle, support, visée, année de publication ou encore association de plusieurs de ces éléments. Ainsi, la *figure 2* indique que l'on recherche, dans notre corpus, tous les documents du service médical et, en particulier, de l'unité d'hémato-oncologie – dépistage et prévention du cancer – cancérologie mammaire de l'Institut Jules Bordet, rédigés en français, à visée injonctive, tirés de pages Web, ayant le patient pour lecteur modèle et composés à une date quelconque.

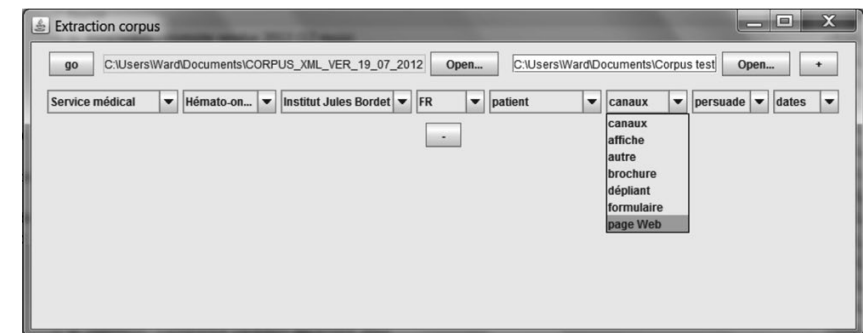


Figure 2 – Interface génératrice de sous-corpus

8. Toujours avec le soutien de C. Libert.

Enfin, suivant les conclusions de notre étude des concordanciers disponibles, nous avons veillé à ce que notre interface de sélection textuelle produise comme résultat de sortie un seul fichier regroupant tous les documents concernés par notre requête, codé au format .txt, et facilement exploitable par les outils d'analyse du discours présentés dans le tableau 2.

4. L'alignement du corpus *Babeliris*

Dans le contexte de notre projet *Babeliris*, nous avons aussi composé un corpus parallèle français-néerlandais au départ de l'ensemble de nos textes médico-administratifs bilingues⁹. En réalité, nous avons conçu un corpus spécialisé de traduction unidirectionnel, fondé sur un appariement de segments français-néerlandais, étant donné que la plupart des textes *IRIS* sont traduits du français vers le néerlandais comme le laisse supposer le tableau 3 (40,9 % de documents disponibles uniquement en français, contre 7,6 % pour le néerlandais) :

TABLEAU 3 – TEXTES UNILINGUES ET TRADUITS DU CORPUS *BABELIRIS*

	LANGUE	NOMBRE TOTAL D'ÉCRITS	NOMBRE D'ÉCRITS UNILINGUES	POURCENTAGE D'ÉCRITS UNILINGUES	NOMBRE D'ÉCRITS ALIGNABLES	POURCENTAGE D'ÉCRITS ALIGNABLES
1	Français	1 126	461	40,9 %	665	59,1 %
2	Néerlandais	720	55	7,6 %	665	92,4 %
	TOTAL	846	516	28,0 %	1 330	72,0 %

9. Obéissant au souhait de la structure faitière de l'IRIS de penser l'aménagement linguistique prioritaire au sein des hôpitaux bruxellois, nous avons décidé d'aligner uniquement les documents disponibles en français et en néerlandais. La banque de textes ainsi alignés, de laquelle naîtront les outils mentionnés en 1, pourra également servir à alimenter les cours de langues sur objectifs spécifiques destinés au personnel (<http://www.iris-hopitaal.eu/fr/>).

L'alignement phrastique de segments et de sous-segments prend sens à la lumière des objectifs visés par le projet *Babeliris*, puisqu'il permet une extraction aisée de certains termes et de leurs équivalents vulgarisés, dégagés par la comparaison structurelle des langues des bi-textes¹⁰. Certains de ces termes vulgarisés viendront ensuite alimenter notre base de données terminologique. Par ailleurs, les textes parallèles alignés constitueront une banque de données bilingue, que nous convertirons en une vaste mémoire de traduction à faire valider par les traducteurs de la structure *IRIS* faitière.

Afin de rencontrer au plus vite ces objectifs, nous avons dû choisir, pour nos écrits, un aligneur multilingue un gestionnaire de mémoires de traduction. Les trois outils testés dans ce cadre, dont les principales caractéristiques sont résumées dans le tableau 4, sont *WinAlign* de *SDL Trados*, *MultiTrans Prism* de *MultiCorpora* et *mkAlign*.

TABLEAU 4 – COMPARAISON DES ALIGNEURS TESTÉS

	LANGUE	NOMBRE TOTAL D'ÉCRITS	NOMBRE D'ÉCRITS UNILINGUES	POURCENTAGE D'ÉCRITS ALIGNABLES
1	Mise à jour du logiciel	2007	2011	2012
2	Nombre de langues alignables simultanément	2	2	2
3	Coût	Payant	Payant	Gratuit
4	Niveau d'appariement automatique le plus fin	Phrase	Phrase et sous-segment	Phrase
5	Niveau d'intervention manuelle autorisée	Mot	Sous-segment (encadré d'une ponctuation forte)	Mot
6	Méthode d'alignement	Flèches	Surlignage jaune	Tableau à double entrée
7	Format d'exportation échangeable	.tmw	.tmx	.tmx

10. Ainsi, pour un profane du langage médical, le terme *carence martiale* sera moins transparent que son équivalent néerlandais *ijzertekort*, qui littéralement se traduit par *manque de fer* et que nous retiendrons désormais sous cette forme. Par ailleurs, la confrontation des segments source et cible permettra aussi de relever certains cas de vulgarisation traductionnelle : ainsi le terme *vitamine liposoluble* renfermera davantage de renseignements que son équivalent néerlandais *oplosbare vitamine* en raison de la présence du préfixe d'origine grecque *λιπος* (graisse, corps gras), mais son sens sera plus complexe d'interprétation. C'est pourquoi nous préférons finalement retenir la version *vitamine soluble* du terme.

Nous avons finalement retenu *Multi Trans Prism* dans sa version Expert. Même si ce dernier a son coût et ne permet pas une intervention manuelle au niveau du mot, il présente toutefois de nombreux avantages :

- Il produit censément des alignements automatisés de textes parallèles
- aussi 1 : N et N : 1¹¹ – avec une exactitude confirmée avoisinant les 95 %. Son algorithme sophistiqué est basé sur la longueur des phrases et des paragraphes, la mise en page et les marques de ponctuation. Dès lors, cet outil n'exploite ni lexiques bilingues ni cognats¹², si bien qu'en l'absence d'un ancrage lexical, il s'avère indépendant des langues traitées ;
- Il est doté d'une interface aussi ergonomique qu'intuitive et nous a semblé, en ce sens, plus convivial que les deux autres outils testés ;
- Il permet d'exporter tout appariement de segments vers le format universel d'échange de mémoires de traduction .tmx.

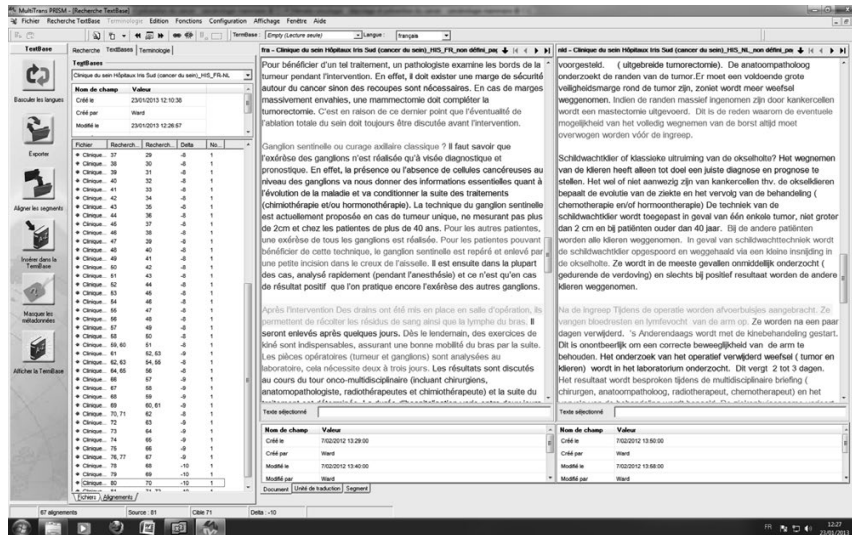


FIGURE 3 – INTERFACE DE L'ALIGNEUR *MULTI TRANS PRISM*

11. Pour sa part, Véronis évoque l'idée de correspondances complexes m:n où m, n > 1 (2000 : 159).

12. La notion de *cognat* est traitée entre autres par Véronis (2000 : 159-160).

5. Les principales caractéristiques linguistiques du corpus français Babeliris

Dans cette section, nous exposerons quelques traits linguistiques du pan français de notre corpus (abrégé BAB., 519 393 mots), créé à l'aide de notre interface génératrice de sous-corpus, et les confronterons aux spécificités langagières de trois autres corpus, constitués à des fins de comparaison linguistique¹³ :

- Corpus spécialisé de pédiatrie (abrégé PÉD., 154 864 mots) : élaboré au départ du cours polycopié du professeur P. Tounian (2002) de l'Université Pierre et Marie Curie ;
- Corpus spécialisé de notices pharmaceutiques (abrégé NOT., 335 431 mots) : conçu à partir des 141 notices¹⁴ mises à jour entre le 15 et le 31 décembre 2012, glanées sur le site de l'Agence nationale de sécurité du médicament et des produits de santé (ANSM) ;
- Corpus de langue de presse généraliste et vulgarisée *L'Essentiel* (abrégé ess., 156 082 mots) : constitué au départ de 271 articles de journal rédigés en français simplifié et publiés en ligne¹⁵ en Belgique entre le 1^{er} mars 2009 et le 31 décembre 2012.

Très concrètement, nous analyserons notre corpus français *Babeliris* et ces trois autres corpus au regard de cinq caractéristiques

13. Nous sommes parti de l'hypothèse selon laquelle notre corpus *Babeliris* devait sans doute renfermer des caractéristiques linguistiques à la croisée du langage spécialisée et de la langue générale vulgarisée.

14. Sur son site *Internet*, cette agence française définit une notice comme étant une « *annexe de la décision d'autorisation comportant des informations notamment sur les indications thérapeutiques, contre-indications, modalités d'utilisation et les effets indésirables d'un médicament* » et précise que « *ces informations sont insérées dans le conditionnement (emballage) contenant le médicament [et sont plus] particulièrement destinées au patient* » (<http://agence-prd.ansm.sante.fr/php/ecodex/glossair/glossair.php#Notice>).

15. Ces articles sont consultables sur <http://www.journal-essentiel.be/>.

syntaxiques propres au langage spécialisé, dégagés par Van Campenhoudt (2011).

Dans son cours, celui-ci s'intéresse aux parties du discours. Opposant des statistiques sur l'emploi de classes grammaticales dans des corpus de textes spécialisés (SPÉ.) et de textes littéraires (LITT.) (*op. cit.* : 23), il observe que les écrits de spécialité renferment sensiblement plus de substantifs et d'adjectifs pour moins de verbes et de pronoms. C'est donc en toute logique que nous avons alimenté, à l'aide du logiciel *Cordial Analyseur* (fonction *Étiquetage de textes*), le tableau 5 qui reprend, pour les quatre classes grammaticales concernées, le pourcentage d'attestations par rapport au nombre total de formes du corpus.

TABLEAU 5 – ÉVALUATION DES QUATRE CLASSES GRAMMATICALES VARIABLES DANS NOS SIX CORPUS

	CLASSE GRAMM.	LITT.	SPÉ.	BAB.	PÉD.	NOT.	ESS.
1	Substantifs	18,20 % - 24,30 %	18,20 % - 24,30 %	18,20 % - 24,30 %	18,20 % - 24,30 %	27,30 % - 30,80 %	29,60 %
2	Adjectifs	4,40 % - 5,30 %	5,30 % - 11,10 %	7,30 %	11,30 %	7,40 %	6,00 %
3	Verbes	13,10 % - 14,90 %	9,00 % - 11,10 %	10,40 %	9,00 %	10,70 %	12,20 %
4	Pronoms	9,00 % - 12,70 %	2,80 % - 5,30 %	4,20 %	2,70 %	3,50 %	5,60 %

Notre analyse laisse apparaître, pour notre corpus BAB., un taux élevé de substantifs et d'adjectifs pour un faible pourcentage de verbes et de pronoms. Sur le plan des classes grammaticales, le corpus bab. tend ainsi à se rapprocher du corpus spécialisé SPÉ., de PÉD. et de NOT., et à s'écarter des valeurs conférées au corpus littéraire de référence LITT.

Un degré élevé de figements de langue constitue une autre caractéristique du langage spécialisé (*op. cit.* : 14-16). Ces figements peuvent être évalués notamment grâce à l'information mutuelle (IM), étant entendu qu'un score supérieur à 3 bits permet de rendre compte d'une cooccurrence pertinente ou d'une collocation (Church et Hanks 1990 : 23-24). Le tableau 6 donne un aperçu, pour nos six corpus, du nombre de cooccurrents ayant une im égale ou supérieure à 3 bits, du total en bits de ces IM égales ou supérieures à 3 bits et, enfin, de la moyenne en bits par cooccurrent pertinent retenu en 1 (mesures calculées à l'aide d'*Oxford WordSmith Tools 4.0*¹⁶).

TABLEAU 6 – CALCUL DE L'INFORMATION MUTUELLE

	DONNÉE D'IM	LITT.	SPÉ.	BAB.	PÉD.	NOT.	ESS.
1	Nombre de cooccurrents im ≥ 3	14 638	19 157	27 190	6 434	21 690	6 229
2	Total en bits im ≥ 3	83 129,35	127 788,50	184 955,86	41 764,95	162 559,86	38 125,76
3	Moyenne en bits / cooccurrent	5,68	6,67	6,80	6,49	7,49	6,12

Nos corpus étant de tailles diverses¹⁷, nous nous intéresserons uniquement aux moyennes en bits des cooccurrents égaux ou supérieurs à 3 bits. On notera à ce propos que les valeurs obtenues par les cooccurrents de LITT. dépassent à peine les 5,50 bits, alors

16. Pour ce faire, nous avons dégagé des informations de « base 3 », c'est-à-dire que nous avons retenu des traces de figement apparaissant au moins 3 fois dans le corpus, égales ou supérieures à 3 bits, et pour lesquelles le mot pivot est séparé du mot qui le renforce par 3 mots tout au plus (*span 3*) ou une fin de phrase (*sentence break*).

17. Pour information, les corpus LITT. et SPÉ. comptent respectivement 434 816 et 394 272 mots.

que celles des textes de spécialité excèdent sans problème les 6,50 bits. Cet écart notable permet dès lors de distinguer un texte littéraire (LITT.) ou de langue générale (ESS.) d'un texte spécialisé (SPÉ.) Ainsi, avec ses 6,80 bits, notre corpus bab. se range du côté du discours spécialisé, marqué par des segments de phrase répétitifs (Van Campenhoudt 2011 : 15).

L'emploi des temps constitue une autre variable pertinente pour reconnaître les écrits spécialisés (*op. cit.* : 21-23). Kocourek nous apprend que dans le texte technoscientifique, *le présent de l'indicatif, à la voix active ou passive, est la forme verbale la plus fréquente, avec environ 85 %. À peu près dix pour cent sont réservés au passé composé et au futur, et le reste est réparti entre les autres temps ou modes* (1991 : 71). Le tableau 7 rend compte de la répartition de ces temps, ainsi que de l'imparfait et du passé simple, parmi l'ensemble des verbes (données obtenues à l'aide de la fonction *Étiquetage de textes de Cordial Analyseur*).

TABLEAU 7 – RÉPARTITION DES TEMPS VERBAUX DANS NOS SIX CORPUS

	TEMPS VERBAL	LITT.	SPÉ.	BAB.	PÉD.	NOT.	ESS.
1	Présent	18,60 % - 57,50 %	63,60 % - 77,00 %	60,80 %	78,40 %	51,60 %	63,10 %
2	Futur simple	1,70 % - 5,10 %	0,80 % - 9,80 %	6,70 %	2,50 %	2,90 %	3,70 %
3	Passé composé	3,20 % - 12,60 %	2,80 % - 12,80 %	4,60 %	3,00 %	6,70 %	11,50 %
4	Imparfait	4,80 % - 30,20 %	0,20 % - 3,10 %	0,70 %	0,60 %	0,40 %	4,20 %
5	Passé simple	0,30 % - 22,50 %	0,00 % - 1,70 %	0,20 %	0,10 %	0,00 %	0,10 %

Le temps qui domine dans BAB., comme dans PÉD., est le présent de l'indicatif. Quant aux emplois du futur simple et du passé composé (11,30 %), ils atteignent largement les dix pour cent évoqués par Kocourek (1991 : 71). Le futur simple est d'ailleurs très répandu dans bab., peut-être parce que « *dans les textes médicaux, [il] semble souvent utilisé pour donner des consignes en évitant la fonction conative* » (Van Campenhoudt 2011 : 22). En réalité, cette stratégie s'avère fréquente lorsqu'il s'agit de donner des instructions sous une forme modalisée. Enfin, l'imparfait et le passé simple semblent davantage réservés aux textes littéraires et sont presque absents de BAB., de PÉD., de NOT. et de ESS. rédigé en français simplifié.

Quatrième caractéristique d'un texte spécialisé : un faible embrayage pronominal. Van Campenhoudt évoque *une très faible présence des pronoms personnels des premières et deuxième personnes dans les textes spécialisés [...] L'exception la plus notable est celle des modes d'emploi, instructions et autres notices vulgarisatrices* (*op. cit.* : 18-19). Pour sa part, Kocourek nous apprend que *le premier trait saillant de la syntaxe technoscientifique est la prédominance de la 3^e personne* (1991 : 71). Ces traits langagiers ne semblent toutefois pas respectés dans BAB., comme l'indique le tableau 8 résumant le pourcentage de pronoms personnels embrayés par rapport à l'ensemble des pronoms personnels (mesures recueillies à l'aide de la fonction *Étiquetage de textes de Cordial Analyseur*).

TABLEAU 8 – EMBRAYAGE PRONOMINAL DANS NOS SIX CORPUS

	PRON. EMBRAYÉ	LITT.	SPÉ.	BAB.	PÉD.	NOT.	ESS.
1	1 ^{re} pers. sg	24,30 % - 33,30 %	0,00 % - 6,00 %	5,40 %	0,00 %	0,00 %	3,50 %
2	2 ^e pers. sg.	0,10 % - 1,60 %	0,00 % - 0,30 %	0,60 %	0,30 %	0,20 %	0,40 %
3	1 ^{re} pers. pl.	7,20 % - 9,60 %	0,00 % - 6,50 %	7,90 %	0,60 %	0,00 %	4,00 %
4	2 ^e pers. pl.	2,00 % - 6,20 %	0,00 % - 1,90 %	32,10 %	0,00 %	63,10 %	3,90 %

L'utilisation de la première personne, surtout au pluriel (1 358 occ. de *nous*), est très fréquente dans BAB. et désigne souvent le personnel *IRIS*. Quant à l'emploi du pronom de deuxième personne, il semble surreprésenté au pluriel de politesse (5 982 occ. de *vous*) et évoque la patientèle du réseau hospitalier. Cette distribution pronominale atypique nous empêche tout logiquement de rapprocher BAB. d'un autre corpus à notre disposition, même si celui-ci se démarque, tout comme dans NOT., par une forte présence de la 2^e p. pl.

Enfin, même s'il convient de les interpréter avec prudence en l'absence d'une analyse sémantique rigoureuse, les longueurs moyennes des mots et des phrases ainsi que l'indice de Flesch-De Landsheere permettent aussi de caractériser un texte spécialisé. Kocourek écrit ainsi que « *les phrases technoscientifiques se caractérisent par leur longueur [...] moyenne [...] entre 28 et 29 mots* » (1991 : 73). Les mots du discours scientifique semblent aussi suivre cette tendance à être longs (Van Campenhout 2011 : 10). Enfin, l'indice de Flesch-De Landsheere permet d'évaluer grossièrement la lisibilité d'un texte, qui sera généralement inférieure pour un écrit spécialisé (*ibid.* : 13). Le tableau 9 précise,

pour nos six corpus, les trois mesures lexicométriques susdécrites, définies avec *Cordial Analyseur* et *TextStat*.

TABLEAU 9 – QUELQUES RELEVÉS LEXICOMÉTRIQUES DE NOS SIX CORPUS

	MESURE LEXICOMÉTRIQUE	LITT.	SPÉ.	BAB.	PÉD.	NOT.	ESS.
1	Nombre moyen de lettres par mot	4,32 - 4,61	4,60 - 5,29	5,19	5,39	5,42	4,71
2	Nombre moyen de mots par phrase	16,29 - 23,16	27,09 - 46,45	10,95	15,22	10,53	11,77
3	Indice de Flesch-De Landsheere	41,86 - 48,32	19,00 - 42,66	17,63	4,24	14,55	32,35

La longueur moyenne des phrases de BAB. (mais aussi de NOT. et de PÉD.) étant biaisée par la présence de nombreux tableaux et de listes à puces, nous nous focaliserons uniquement sur les variables *1* et *3*. S'agissant du nombre moyen de lettres par mot, on notera que les termes de BAB., de PÉD. et de NOT. sont relativement longs et rejoignent en ce sens les modalités d'écriture de SPÉ., à la différence de ceux de ESS. qui sont censés être accessibles à un large public. Enfin, des indices de Flesch-De Landsheere bas pour BAB., PÉD. et NOT. attestent d'une véritable complexité, d'autant plus marquée qu'ils sont largement inférieurs aux scores de SPÉ. Les textes de BAB., de PÉD. et de NOT. pourraient donc être qualifiés de « *très difficiles* » (De Landsheere 1982).

Nous serions tenté d'en conclure pour l'instant que nous sommes en présence d'un corpus original, qui laisse apparaître des caractéristiques globales du langage spécialisé (parties du discours, emploi des temps, information mutuelle et mesures lexicométriques) et qui se démarque par une sur-représentativité de la fonction conative, typique des notices pharmaceutiques et des modes d'emploi.

6. Conclusion et perspectives

Aujourd'hui, de nombreuses études empiriques du langage appliquent les méthodes de la linguistique de corpus. L'exploitation d'une vaste collection de textes s'est ainsi imposée à nous comme une évidence.

Dans cette logique, il nous a fallu procéder à la *phase de préparation du corpus* (Sansonetti 2003 : 73) ou, pour le dire avec les mots d'Habert, à la *phase de dépouillement du corpus*, à savoir à cette « exigence de standardisation des textes contenus dans un corpus [...] destinée avant tout à les rendre comparables, à les stabiliser le temps d'une expérience » (1997 : 194). C'est dans ce contexte que se sont inscrites nos réflexions sur les balises *TEI-P5* et sur un aligneur *ad hoc* pour nos textes bilingues. La genèse de notre interface génératrice de sous-corpus s'est, elle aussi, posée comme un préalable à l'étude linguistique, en l'espèce du pan français de notre corpus *Babeliris*. À l'aide de cet outil, nous avons facilement pu en conclure que le français des hôpitaux *IRIS* rejoignait celui des textes spécialisés marqués par une prédominance de la 2^e p. pl.

Enfin, les perspectives d'exploitation pour notre corpus sont nombreuses. On pourrait imaginer générer des sous-corpus dans le but d'analyser, par exemple, les spécificités langagières propres à chaque hôpital *IRIS* ou à chaque unité. Il serait aussi intéressant de développer un outil qui croiserait interface de sous-corpus et aligneur, en vue d'étudier les procédés traductionnels spécifiques à un pan de la littérature *IRIS* et d'en envisager la vulgarisation. C'est que notre objectif in fine est, avant tout, d'encourager la lisibilité des écrits destinés au patient.

Bibliographie

- BAUMAN (S.) et BURNARD (L.) 2012, *TEI-P5: Guidelines for Electronic Text Encoding*, Charlottesville, Text Encoding Initiative Consortium, version révisée 2.0.2. (www.tei-c.org/guidelines/p5)
- BIBER (D.) 1993, « Representativeness in Corpus Design », in *Literacy and Linguistic Computing*, 8/4, Oxford, pp. 243-257.
- BOÏTEKÉ (P.) 2011, *État de la question. Emploi des langues à Bruxelles : mieux tenir compte de la réalité sociologique bruxelloise*, rapport commandité par la Fédération Wallonie-Bruxelles et réalisé par l'Institut Emile Vandervelde, Bruxelles.
- CHURCH (K.W.) et HANKS (P.) 1990, « Word Association Norms, Mutual Information, and Lexicography », in *Computational Linguistics*, Cambridge, MA, 16/1, pp. 22-29.
- DAOUST (F.) et MARCOUX (Y.) 2006, « Logiciels d'analyse textuelle : vers un format XML-TEI pour l'échange de corpus annotés », in *Actes des 8^{es} journées internationales d'analyse statistique des données textuelles (JADT) 2006*, Besançon, Presses universitaires de Franche-Comté, 1, pp. 327-340.
- DE LANDSHEERE (G.) 19825, *Introduction à la recherche en éducation*, Liège, Thone.
- DUFOURNAUD (N.) et GRATZAC-LEGENDRE (V.) 2012, *Manuel d'encodage XML-TEI – édition numérique de manuscrits baroques. Recommandations pour une application TEI ENBaCH, École des hautes études en sciences sociales (EHESS)*, Paris, 60 pp.
- HABERT (B.) 2000, « Des corpus représentatifs : de quoi, pour quoi, comment ? », in *Linguistique sur corpus. Études et réflexions*, M. Bilger, Perpignan, Presses Universitaires de Perpignan, pp. 11-58.
- HABERT (B.), NAZARENKO (A.) et SALEM (A.) 1997, *Les linguistiques de corpus*, Paris, Masson et Armand Colin, 240 pp., Coll. U Linguistique.
- HAREIDE (L.) et HOFLAND (K.) 2012, « Compiling a Norwegian-Spanish parallel corpus », in OAKES (M.P.) et JI (M.), *Quantitative Methods in Corpus-Based Translation Studies*, Studies in Corpus Linguistics, 51, Amsterdam, John Benjamins Publishing Company, pp. 75-113.
- HAUTE AUTORITÉ DE LA SANTÉ 2008, *Élaboration d'un document écrit d'information à l'intention des patients et des usagers du système de santé*, Saint-Denis, Haute Autorité de Santé.

HUIGEN (M.) 2004, *Zelf brochures schrijven*, coll. TaalAnker: *Hoe formuleer ik het? Cahierreeks voor taalgebruikers*, 48, Alphen aan den Rijn, Kluwer.

ISO 639-1 2002, *Codes pour la représentation des noms de langue. Partie 1 : Code alpha-2*, Genève, Organisation internationale de normalisation.

JACQUEMART (P.) s.d., *Conception d'un corpus de textes médicaux*, rapport de stage effectué dans le laboratoire DIAM-SIM/DSI /AP-HP sous la direction de P. Zweigenbaum, Paris (www.limsi.fr/npz/ftpapiers/jacquemartdea99.doc).

KOCOUREK (R.) 1991, *La langue française de la technique et de la science. Vers une linguistique de la langue savante*, Wiesbaden, Oscar Brandstetter Verlag & co.

LEE (D.) 2001, « Genres, Registers, Text Types, Domains, and Styles: Clarifying the Concepts and Navigating a Path through the BNC Jungle », in *Language Learning & Technology*, s.l., 5/3, pp. 37-72.

McENERY (T.), TONO (Y.) et XIAO (R.) 2006, *Corpus-Based Language Studies: An Advanced Resource Book*, Routledge Applied Linguistics, Londres, Routledge.

MINISTÈRE BELGE DES AFFAIRES SOCIALES, DE LA SANTÉ PUBLIQUE ET DE L'ENVIRONNEMENT 2002, *Loi relative aux droits du patient*, publiée au *Moniteur belge* le 26 septembre 2002, pp. 43719-43724.

POPESCU-BELIS (A.) 2002, « Constitution de banques de textes multilingues : un mécanisme fondé sur le standard XML », in *Cahiers du Rifal*, Paris, 23, pp. 56-60.

SACKETT (D.L.) et al. 1996, « Evidence based medicine: what it is and what it isn't », in *BMJ*, 312, pp. 71-72.

SANSONETTI (L.) 2003, « Approche lexicométrique de corpus d'interactions verbales entre un adulte et un enfant en cours d'acquisition du langage. Résultats d'expérience », in WILLIAMS (G.), *Textes et Corpus : Actes des Troisièmes Journées de la Linguistique de Corpus*, pp. 71-84 (web.univ-ubs.fr/corpus/jlc3/jlc2003.pdf).

SINCLAIR (J.) 1991, *Corpus, Concordance, Collocation. Describing English Language*, Oxford, University Press.

SINCLAIR (J.) 1996, *Preliminary Recommendations on Corpus Typology*, rapport technique, EAGLES (Expert Advisory Group on Language Engineering Standards), Birmingham University (www.ilc.cnr.it/eagle596/browse.html).

TOGNINI-BONELLI (E.) 2001, *Corpus Linguistics at Work*, Amsterdam, John Benjamins Publishing.

TOUNIAN (P.) 2002, *Pédiatrie, DCEM 3*, Paris, Université Pierre et Marie Curie, cours photocopié.

TYMOCZKO (M.) 1998, « Computerized Corpora and the Future of Translation Studies », in CLAS (A.), *Meta : journal des traducteurs*, Montréal, 43/4, pp. 652-660.

VAN CAMPENHOUDT (M.) 2011, *Linguistique française et exercices III. Le français spécialisé*, Bruxelles, Institut supérieur de traducteurs et interprètes, cours photocopié de 3e bachelier, document interne.

VAN DE VELDE (W.) à paraître, « La construction du corpus d'écrits médico-administratifs BABELIRIS : vers une méthodologie de compilation et d'encodage des textes », in *Actes de la conférence Traduction et Innovation*, Université Paris Diderot (Paris VII), 13-15 décembre 2012.

VAN OMMEN (H.) et VAN KUPPENVELD (E.) 1995, *Professionele bedrijfscommunicatie. Het handboek voor tekstschrijvers*, Groningen, Wolters-Noordhoff.

VÉRONIS (J.) 2000, « Alignement de corpus multilingues », in PIERREL (J.-M.), *Ingénierie des langues. Informatique et systèmes d'information*, chapitre 6, Paris, Hermès Sciences.

VIHLA (M.) 1998, « MEDICOR: a corpus of contemporary American medical texts », in *ICAME Journal*, Bergen, 22, pp. 73-80.

ZWEIGENBAUM (P.) et al. 2001, « Building a text corpus for representing the variety of medical language », in V. Patel et al., *Medinfo 2001*, Proceedings Corpus Linguistics, Lancaster, pp. 290-294.

L'enseignement/apprentissage du français juridique : une proposition de démarche terminologique et didactique

Christina DECHAMPS
Centro de Linguística
Universidade Nova de Lisboa

Résumé

De nos jours, l'enseignement/apprentissage des langues de spécialité et, en particulier, de la langue juridique, représente un domaine d'activité important où, de façon contradictoire, manquent des ressources fiables susceptibles d'aider, à la fois, l'enseignant et l'apprenant. Et ce manque est d'autant plus évident lorsque l'on confronte langue française et langue portugaise. De cette façon et dans le but d'apporter notre contribution au manque signalé ci-dessus, nous avons consacré nos recherches 1) à l'élaboration d'un corpus comparable bilingue français-portugais de textes juridiques qui nous a permis, ultérieurement, de relever les principales collocations verbales dans les deux langues et 2) au développement, après la sélection et l'analyse des collocations relevées dans le corpus, d'une plateforme numérique rassemblant divers exercices capables d'aider les apprenants lusophones à acquérir les collocations verbales de la langue juridique française. L'objectif de cet article est donc de présenter la plateforme numérique réalisée ainsi que les diverses méthodologies adoptées pour son élaboration.

Mots-clés :

collocation, langue juridique, corpus, enseignement/apprentissage, plateforme numérique d'apprentissage.

1. Introduction

Pour bien des raisons, l'enseignement/apprentissage (E/A) de la langue juridique représente un domaine d'activité important. Au-delà des obstacles terminologiques, s'impose à l'enseignant et à l'apprenant l'ardue tâche d'enseigner/apprendre des structures linguistiques semi-figées que nous appellerons ici « collocations ». En effet, si l'acquisition de termes juridiques est loin d'être une tâche simple, celle des collocations de cette même langue de spécialité l'est encore moins. Paradoxalement, les ressources qui pourraient aider l'apprenant manquent, particulièrement dans le contexte français-portugais. De fait, il règne encore un certain vide dans la recherche à ce propos ainsi qu'un manque de communication entre linguistique et didactique, deux situations qui ne favorisent pas la mise en place de nouvelles pratiques pédagogiques pour l'amélioration de l'E/A des collocations.

2. Présentation et objectifs de la recherche

De cette façon et dans le but de combler le manque signalé ci-dessus, nous avons consacré notre recherche à l'étude de la collocation verbale (V + N) dans la langue juridique française. Nous entendons par « collocation », une combinaison non libre constituée d'une base (N = terme), choisie librement et d'un collocatif (verbe) qui permet d'attribuer un sens spécifique à l'expression et/ou qui l'encadre dans un discours plus ou moins spécialisé.

Vu l'étendue de la question, nous nous sommes limitée aux collocations contenant un verbe transitif et un terme COD ; ce qui nous a permis d'analyser les collocations à la voix active (*le Président de la République ratifie un traité*), à la voix passive (*un traité est ratifié par le Président de la République*) et avec l'emploi seul du participe passé (*un traité ratifié le 31 janvier ...*). Cette analyse a été réalisée sur la base d'un corpus comparable bilingue de textes juridiques et, à partir des résultats obtenus, un produit numérique à caractère pédagogique – une plateforme comprenant des activités diverses – a été élaboré, dans le but de renforcer l'E/A des collocations juridiques.

3. Constitution du corpus - méthodologie

Afin de faciliter le relevé des principales collocations verbales dans les deux langues, nous avons ainsi procédé à l'élaboration d'un corpus comparable bilingue français-portugais de textes juridiques. D'emblée, nous devons mentionner que nous avons opté pour un corpus comparable, et non pour un corpus parallèle, pour deux raisons. La première est que nous ne disposons pas de traduction des ouvrages sélectionnés ; et la deuxième est que nous estimons que les textes authentiques sont plus fiables en termes terminologiques et discursifs, contrairement aux traductions, plus susceptibles de comporter des erreurs.

Notre corpus a la particularité de rassembler des textes juridiques à caractère didactique, c.-à-d. des ouvrages destinés au public

estudiantin qui s'initie à la science du droit et marqués par les caractéristiques du discours de semi-vulgarisation scientifique ou discours scientifique pédagogique (Loffler-Laurian 1983, Eurin-Balmet 1992). Ce choix repose sur le fait que ces textes présentent une certaine sélection terminologique en ne conservant que les concepts jugés essentiels pour le public auquel ils s'adressent – nous donnant ainsi accès aux structures collocationnelles considérées comme pertinentes dans le contexte de l'E/A de la langue juridique. Cela dit, il serait erroné de penser que ce corpus est moins riche qu'un corpus de textes spécialisés. En fait, nous y retrouvons des citations de textes doctrinaux, de textes législatifs et de jurisprudence, pour reprendre la classification de textes juridiques proposées par Bocquet (2008), ce qui fait de notre un corpus un recueil de textes riches, même s'ils sont très marqués par le discours scientifique pédagogique. Par ailleurs, le type de discours que nous retrouvons dans ce corpus se rapproche beaucoup du type de communication que nous avons dans le cadre d'un échange entre avocat et client, où nous allons retrouver de fortes marques de vulgarisation mais aussi de didacticité¹. Ce type de communication correspond souvent aux besoins communicatifs des apprenants en langue juridique.

Le corpus français – JURIFRAN – comprend cinq ouvrages et 447 275 mots ; et le corpus portugais – JURIPOR – , deux ouvrages et 185 072 mots. Il s'agit d'un corpus relativement restreint mais assez représentatif. En fait, les écrits juridiques de ce type sont relativement peu nombreux, surtout en langue portugaise ; toutefois, nous avons veillé à sélectionner des textes qui sont considérés comme des ouvrages de référence.

1. Terme que nous utilisons suivant l'acception de Moirand (1993).

Pour l'extraction des candidats à collocations, nous avons principalement utilisé le logiciel Hyperbase 9.0 d'Étienne Brunet (Université de Nice). Cependant, ponctuellement, nous avons recouru à d'autres programmes de traitement de données terminologiques : TermoStat 3.0 de Patrick Drouin (Université de Montréal)² et SCP 4.0.9 ou Simple Concordance Program 4.0.9 d'Alan Reed³.

4. Extraction et sélection des collocations : méthodologie

Les différents logiciels que nous avons eu l'occasion d'expérimenter sont malheureusement peu performants dans l'extraction automatique des collocations verbales. La fonction Bigrammes de TermoStat en identifie un nombre très réduit et la fonction Cooccurrence d'Hyperbase reconnaît tous les mots qui se retrouvent dans un environnement proche, sans qu'il y ait pour autant une quelconque relation syntaxique entre les deux éléments.

C'est pourquoi l'extraction a dû se faire en grande partie manuellement, en recourant principalement aux fonctions Concordance et Contexte d'Hyperbase. Nous sommes ainsi partie de la base de la collocation, c.-à-d. le terme (substantif). À l'analyse des listes de fréquences obtenues grâce aux logiciels Hyperbase et SCP et de la liste des candidats termes fournie par TermoStat, nous nous sommes rendu compte que le sous-domaine conceptuel le plus représentatif dans notre corpus est celui des sources du droit, de l'écrit juridique en général.

2. Voir http://olst.ling.umontreal.ca/~drouinp/termostat_web/index.php (consulté le 21-2-2013).

3. Voir <http://www.textworld.com/scp> (consulté le 21-2-2013).

Cette constatation est loin d'être une surprise, eu égard à l'importance des textes écrits dans la culture juridique romano-germanique. Ainsi, nous avons pris le parti d'étudier les phénomènes combinatoires qui se manifestent avec les termes qui appartiennent à ce sous-domaine conceptuel et qui se retrouvent dans notre corpus.

En ce qui concerne la sélection des collocations, nous avons appliqué à notre premier relevé plusieurs critères : critères de fréquence, de figement, critères terminologiques et didactiques.

Dans un premier temps, nous avons retenu les combinaisons présentant, en termes de fréquence, un minimum de quatre occurrences ; à noter que ce seuil de 4 est celui qui est adopté par Patrick Drouin dans *TermoStat* et qu'il nous a paru adéquat à la taille de notre corpus.

Dans un deuxième temps, nous avons appliqué les critères de figement ainsi que les critères terminologiques et didactiques.

Pour ce qui est du figement, nous avons analysé chaque candidat à collocation en observant les possibilités de combinaison du verbe avec :

- d'autres termes de la même classe conceptuelle (*ratifier un traité, une ordonnance, une convention*) ;
- d'autres termes juridiques (pour *ratifier*, il n'existe pas d'occurrences) ;
- des unités lexicales de la langue courante (*idem*).

Cette analyse a mis en évidence un degré de figement très

variable parmi les structures retenues. C'est le cas notamment de *modifier une loi* (moins figé) et *entériner une jurisprudence* (très figé).

Quant aux critères terminologiques, nous nous sommes reportée à ceux qui sont proposés par L'Homme (1997 : 16) :

- La collocation terminologique doit comporter des mots utilisés dans la langue de spécialité ;
- Une relation grammaticale doit exister entre les unités de la collocation terminologique ;
- L'utilisation de la collocation terminologique doit être le reflet de conventions ;
- La collocation terminologique doit être composée d'une unité terminologique et d'une autre unité lexicale qui est souvent polysémique ;
- Les deux unités de la collocation terminologique doivent posséder d'autres emplois en langue de spécialité :
 - > le verbe se combine avec plusieurs termes
 - > le terme se combine avec plusieurs verbes
 - > le verbe avec un sens déterminé se combine avec plusieurs termes de la même classe conceptuelle.

Ces critères ont facilité l'identification des collocations terminologiques parmi les structures candidates à ce statut. *Ratifier un traité* est un bon exemple de combinaison qui répond à toutes les conditions énoncées ci-dessus.

Finalement, les critères didactiques nous ont aidée à sélectionner les collocations les plus pertinentes dans le cadre de l'E/A de la langue juridique. Dans cette optique, nous avons considéré :

- Les besoins et les objectifs d'apprentissage des apprenants. Le fait de travailler sur un corpus de textes à caractère pédagogique est déjà une façon de répondre au problème. Par ailleurs, focaliser notre attention sur les collocations qui appartiennent au domaine conceptuel des sources du droit est une manière de nous concentrer sur un thème crucial pour l'apprenant dans ses différents besoins communicatifs.
- La problématique de la synonymie ou, mieux, de la parasynonymie en langue juridique que nous retrouvons au niveau collocationnel. Pour un apprenant, des collocations telles que *ratifier un traité* et *approuver un traité* peuvent facilement être perçues comme synonymes et leurs éléments, comme aisément interchangeable dans le discours. Or, il faut savoir que *ratifier un traité* ne correspond pas à *approuver un traité*. En effet dans le premier exemple, on se réfère à l'acte officiel où l'on approuve expressément et selon les normes requises.
- La polysémie interne. Il s'agit certainement de l'une des principales caractéristiques de la langue juridique (Cornu 2000). Au niveau collocationnel, il ne faudra pas, par exemple, confondre *édicter une règle* et *édicter une peine*.
- La charge culturelle présente dans les structures collocationnelles. *Casser un arrêt*, *une décision* renvoie évidemment à la Cour de Cassation, juridiction suprême équivalente au *Supremo Tribunal de Justiça* portugais.
- La question des faux-amis. *Celebrar um contrato* ne se traduit pas par *célébrer un contrat* mais par *conclure un contrat* qui, à son tour, ne pourra jamais se traduire par *conclure um contrato*.

Tous ces critères ont été appliqués sur l'ensemble du relevé de candidats à collocations obtenu après l'application du critère de fréquence ; ce qui a favorisé, dans certains cas, le maintien dans notre sélection définitive de combinaisons qui auraient été éliminées avec l'un ou l'autre critère.

Ex. : *Modifier une loi, un décret / alterar uma lei*

Ici la collocation française présente un degré de figement très faible qui pourrait mettre en doute son statut de collocation. Néanmoins l'expression devient pertinente dans une perspective bilingue, étant donné que le verbe *altérer* se combine peu avec le terme *loi*.

Ex. : *Observer une règle / respecter une règle*

Dans ce cas, les deux combinaisons sont apparemment synonymes mais elles appartiennent chacune à des types de discours spécifiques. *Observer une règle* s'utilisera plutôt dans un discours juridique plus spécialisé.

Ex. : *Refondre un code / entériner une jurisprudence*

Les deux combinaisons présentées ci-dessus sont très figées, caractéristique qui, de nouveau, peut remettre en cause leur statut de collocation mais elles sont, toutefois, intéressantes dans une perspective bilingue, vu qu'elles présentent des entraves au niveau de la réception et de la production et qu'elles sont, par ailleurs, difficilement traduisibles en portugais.

Suite à tout ce processus de sélection, nous avons retenu 77 verbes qui se combinent avec l'un – ou plusieurs – des 48 termes appartenant au domaine conceptuel des sources du droit. Sur l'ensemble de ces verbes, nous observons la coexistence de différents types de verbes : (i) des verbes supports (Gross 1996) ou passe-partout (Lerat 2009 : 219) ; (ii) des verbes très spécialisés (Lerat 2009 : 219) ou pleins (Giraldez 2007) et (iii) des verbes polysémiques avec au moins un sens

spécialisé (Lerat 2009 : 219) qui sont, d'ailleurs, les plus nombreux dans notre relevé.

5. Traitement didactique : méthodologie

En plus des analyses que nous avons développées dans le cadre de la sélection, la théorie, d'une part, des fonctions lexicales de Mel'čuk et, d'autre part, celle des structures actanciennes (L'Homme 2011 et 2012 ; Pimentel 2011 et 2012) ou schémas d'arguments (Gross 1994 et 1998 ; Lerat 2002a, 2002b, 2004, 2005, 2006, 2008, 2009) ont été primordiales pour le traitement didactique des collocations. Elles ont permis de mieux décrire les collocations en mettant en évidence une série de régularités intéressantes dans le cadre de l'E/A. Concrètement, la première a été essentielle pour souligner la présence de grandes classes sémantiques parmi les combinaisons retenues ; la deuxième, pour révéler les récurrences par rapport aux classes d'objets sélectionnées par les verbes attestés dans ces mêmes collocations. Par ailleurs, nous avons étudié les phénomènes de nominalisation et d'équivalences en portugais.

5.1. Les fonctions lexicales

Pour traiter les collocations sélectionnées, il nous a paru judicieux d'adapter, à notre objet d'étude, les fonctions lexicales de Mel'čuk, comme le propose Verlinde (2003 : 19 et 2006 : 91). Un bon exemple de cette adaptation se retrouve dans la description des collocations dans le dictionnaire de Cohen (1986) dédié au domaine de la Bourse⁴.

Ainsi nous avons conservé les fonctions suivantes : début, suite, déclin, fin. Ces catégories s'appliquent assez facilement au cycle de vie d'une source du droit.

Ensuite, nous avons élaboré le classement sémantique des collocations verbales extraites du corpus JURIFRAN. Il faut être conscient que l'application pure et simple d'une classification proposée pour un autre domaine de spécialité est assez délicate malgré la volonté d'universalisme exprimée par tous les auteurs consultés. Certaines adaptations seront toujours à faire, surtout au vu du statut particulier du domaine du droit.

Dans la section « Début », nous avons repris toutes les expressions qui se réfèrent à la création et au début de l'existence d'une source du droit.

Ex. : *Ratifier un traité*
Voter, adopter, promulguer une loi.

Dans la section « Suite », nous avons réuni tous les procès et actions qui surviennent pendant la « vie » d'une source du droit. Nous avons décrit en particulier ceux qui sont en rapport avec la transformation et le déclin⁵.

Ex. : *Adapter une loi, une règle, codifier un texte, une loi,*
un décret, refondre un code, transposer une directive (transformation)
Attaquer un jugement, un arrêt, contester une décision (déclin)

4. Voir L'Homme (2004 : 113-114).

5. Ces derniers items ont été appliqués au domaine de l'économie (Cohen 2011) mais peuvent s'adapter, *mutatis mutandis*, à celui du droit.

Pour terminer le cycle, nous avons la section « Fin » qui correspond à la disparition d'une source du droit.

Ex. : *Abroger une loi, annuler un acte, rompre un contrat, casser un jugement*

Avec les informations que nous pouvons dégager de ce classement, il est possible d'élaborer des exercices onomasiologiques. Pour l'apprenant, il est extrêmement utile de pouvoir établir le lien entre le sens et la forme, étant donné qu'il s'agit de la situation à laquelle il sera confronté en cas de production orale ou écrite. En guise d'exemple, il est intéressant de savoir que l'on dit *abroger une loi, une directive, un décret, un texte mais annuler un acte, casser un arrêt et rompre un contrat* lorsque nous désirons exprimer la fin d'une source du droit. Des expressions telles que *annuler une loi* ou *abroger un contrat* seraient donc jugées incorrectes.

5.2. La structure actancielle du prédicat et actants

Au-delà de tout ce qui vient d'être exposé, il nous a également paru essentiel de mettre en évidence les différents actants de la structure prédicative, c'est-à-dire, dans le cas présent, ne pas nous limiter à la présentation de la seule collocation mais de la considérer comme un élément d'un ensemble plus vaste. Notre objectif est évidemment d'enrichir notre description déjà amorcée des collocations, en tentant de pallier un manque existant à ce niveau.

C'est ainsi que nous nous sommes concentrée plus spécifiquement sur les verbes, c.-à-d. sur les collocatifs des collocations relevées.

Souvent les noms ont servi de point de départ à partir duquel les verbes sont découverts en fonction des liens qu'ils partagent entre eux. Mais il n'est pas impossible de déplacer le point focal de l'analyse et d'y placer le verbe. Cela amènera le terminologue à découvrir d'autres unités plus directement reliées au verbe. (L'Homme 2012 : 100)

Comme l'affirment Le Pesant et Mathieu-Colas (1998 : 14), *les prédicats demeurent à l'arrière-plan et ne font pas l'objet d'un recensement systématique*. Et dans l'optique des langues de spécialité, ils ajoutent (1998 : 27) : *l'introduction des classes d'objets dans l'analyse des langues spécialisées n'en est encore qu'à ses débuts*, et ils défendent l'application de cette approche à l'étude des langues de spécialité – approche qui s'inscrit dans la perspective de la sémantique lexicale. De plus, ce type de description est essentiel pour mieux cerner le poids terminologique des verbes, plus ou moins lourd selon les cas.

Generalmente los lemas verbales que aparecen en textos especializados suelen ser de uso genérico y adquieren valor especializado en contexto cuando su contorno inmediato (sujeto y/o complementos regidos) adquiere también valor especializado. (Lorente 2002 : 4)

Dans le même ordre d'idées, en plus du repérage des COD déjà effectué – rappelons que nous avons relevé les structures V + Terme où ce dernier occupe la position de COD –, nous nous sommes attachée sur l'identification des sujets dans le but d'en établir une éventuelle classification sémantique. D'ailleurs, remarquons d'ores et déjà, que dans *voter une loi / adopter une loi / promulguer une loi*, exemples cités précédemment, les différents verbes pourraient être

éventuellement considérés comme synonymes ou, mieux, parasynonymes, mais ils renvoient à des phases spécifiques de la création d'une loi réalisées par des acteurs bien déterminés.

Ex. : *Le parlement adopte ou vote une loi - Le Président de la République promulgue une loi.*

Pour notre analyse, nous nous sommes donc inspirée des travaux de Gaston Gross (1994, 1998, entre autres) et Le Pesant/Mathieu-Colas (1998), repris, entre autres, par Lerat (2002a, 2002b, 2004, 2005, 2006, 2008, 2009) et Dolota-Zarod (2011) pour la langue juridique.

Ces auteurs parlent de « schéma d'arguments » où nous retrouvons *prédicats* et *arguments*. Ces derniers sont décrits sémantiquement et sont regroupés en classes d'objets.

En 1994, Gaston Gross insiste sur l'importance de distinguer *traits* et *classes d'objets*, la première catégorie étant trop générale et la deuxième, en tant que sous-catégorie, permettant une description qui présente plus précisément l'emploi des différents *prédicats* et *arguments* (1994 : 18). *C'est à l'intérieur des classes et seulement dans ce cadre que l'on peut prédire les restrictions qui permettent de rendre compte de la synonymie. [...] Les classes d'objets permettent [aussi] d'éviter une grande partie des erreurs de traduction* (Gross 1994 : 29).

Nous nous sommes aussi basée sur les recherches de L'Homme

et de Pimentel (2011, 2012, entre autres). Ces linguistes – qui partent du cadre théorique de Fillmore – parlent de *structure actancielle* où nous avons un *terme prédicatif* et des *actants*⁶. L'Homme, déjà en 1997, défend une classification des différents actants en classes conceptuelles.

Les principales classes d'objets que nous avons déterminées pour les actants sujets sont les suivantes : <source du droit>, <institution> et <citoyen>. Dans certains cas, il a été nécessaire, pour des raisons de clarté, de subdiviser ces classes ou de les préciser. Ainsi :

- <source du droit> a été divisé, dans certains cas, en <source directe>, <source indirecte> et <source spécifique>.
- <institution>, pour certaines collocations, est devenu <tribunal>. Il faut comprendre ici <tribunal> comme une étiquette qui renvoie à tout type de tribunal, y compris les différentes cours et autres juridictions spécialisées comme les Prud'hommes.
- Le trait <humain> réunit diverses étiquettes comme <personne morale>, opposée à <personne physique>, <citoyen>, à comprendre ici comme le simple citoyen, ou encore <magistrat> ou <autorité> qui renvoient à des citoyens jouissant d'une certaine fonction ou autorité, au niveau du pouvoir judiciaire, législatif ou exécutif.

L'attribution de l'une ou l'autre étiquette dépend beaucoup des contextes dans lesquels la collocation a été relevée. Dans tous les cas, nous avons recherché la pertinence de cet étiquetage dans le cadre de l'utilisation concrète de la combinaison collocationnelle dans le discours.

6. Dénominations que nous adoptons.

Nous pourrions reprocher à ces classes d'objets leur manque de précision. En effet, le découpage aurait pu être encore plus fin. Toutefois, vu la dimension restreinte du corpus, nous n'avons pas désiré affiner plus l'étiquetage au risque de créer une description qui ne corresponde pas vraiment à l'usage réel de la langue juridique en contexte pédagogique. L'idée ici est surtout d'identifier des tendances plutôt que de présenter des règles strictes de construction d'énoncés, surtout dans le cadre de notre corpus portugais qui est tout de même assez réduit et qui ne permet pas, à coup sûr, d'identifier les comportements collocationnels réels de la langue juridique.

Après avoir déterminé la structure actancielle de chaque verbe sélectionné, il est intéressant d'en faire autant dans l'autre langue afin de mettre en évidence les éventuelles divergences existant entre le français et le portugais. À ce propos, il est assez intéressant de remarquer que, dans certaines collocations, en portugais, les actants sujets sont seulement des institutions alors qu'en français, ces derniers sont des institutions mais aussi, par exemple, des magistrats.

Ex. : <institution> <tribunal>, <magistrat> *rendre une ordonnance, un arrêt, une décision*
↔ <institution> <tribunal> *proferir uma sentença, uma decisão*

Signalons qu'une analyse fine de la structure actancielle des termes prédicatifs dans les deux langues permet de mieux souligner les différents phénomènes d'équivalence totale ou partielle (Pimentel 2012).

Cette approche et ce portrait des verbes collocatifs nous semblent extrêmement opportuns dans l'optique de l'E/A⁷, car ils nous permettent, entre autres, de présenter les différents emplois de certains verbes comme de désambiguïser certains d'entre eux. Les exemples ci-dessous en sont une excellente illustration.

Ex. : <institution> *prend une décision* □ *décision = source du droit*
↔ <citoyen> *prend une décision* □ *décision ≠ source du droit*
<source du droit> *établit une règle, un principe* □ 'donner une existence officielle'
↔ <citoyen> *établit un acte, un écrit* □ 'rédiger selon un modèle officiel'

De plus, pour l'apprenant, ces informations représentent une véritable aide à la production.

En ce qui concerne les exercices que nous proposons pour l'acquisition des collocations verbales, nous accorderons une attention particulière au classement sémantique des actants sujets. Dans la mesure du possible, nous tentons, par cette voie, de mettre en évidence les éventuelles divergences présentes dans la structure actancielle en langue portugaise, comme c'est le cas dans ce dernier exemple.

Ex. : <source du droit> ou <tribunal> *applique une règle*
↔ <tribunal> *aplica uma norma*

7. Comme dans l'optique de la traduction et de la rédaction spécialisée, vu que ce type d'informations est peu présente dans les dictionnaires.

5.3. Nominalisation du prédicat verbal

En plus de toutes les analyses présentées ci-dessus, il nous a semblé essentiel de nous étendre sur les possibilités de nominalisation (Verlinde 2003) ou sur les substantifs déverbaux (Lerat 2009) appelés aussi « noms déverbaux » (Dolata-Zarod 2011). Comme il est admis par de nombreux chercheurs, la nominalisation est l'une des principales caractéristiques de la langue de spécialité et, en particulier, de la langue juridique dont le discours prend ainsi un tour plus impersonnel et plus objectif. De fait, *la nominalisation est, dans les textes spécialisés, la façon la plus fréquente d'exprimer les opérations, les états, les processus et les propriétés* (Lerat 2009 : 221-222).

En observant de plus près la langue juridique, nous remarquons aisément que la plupart des nominalisations sont des noms d'actions provenant de verbes collocatifs⁸. Dans certains cas, à partir du moment où nous reconnaissons le statut terminologique de ces mêmes verbes et de leurs nominalisations respectives – vu qu'ils possèdent un contenu conceptuel identique –, nous pouvons parler de termes à nature prédicative (L'Homme & Pimentel 2011) dont il est possible d'analyser les différents actants, comme nous l'avons démontré dans le chapitre précédent.

Pour certains verbes, aucune nominalisation n'existe et, quand celle-ci existe, nous avons analysé sa fréquence dans JURIFRAN, ainsi que le nombre d'attestations lorsqu'elle est utilisée avec le terme base de la collocation (actant COD).

8. Voir Lerat (2009).

Ex. : *Ratifier un traité* (16)
Ratification (13)
Ratification d'un traité (5)

Cette recherche nous a permis de constater que, contrairement à l'idée défendue par plusieurs auteurs, la nominalisation n'est pas un phénomène aussi présent qu'on le pense, du moins en ce qui concerne les verbes sélectionnés dans le type de corpus choisi pour notre travail. À ce propos, L'Homme affirme que *certaines nominalisations sont beaucoup plus fréquentes que le verbe correspondant. Toutefois, ce phénomène n'est pas systématique* (2012 : 97). Lerat (2009 : 222), quant à lui, précise que *le choix entre formulation verbale et formulation nominale* dépend de plusieurs facteurs comme, par exemple, la langue ou l'auteur. La présence moins marquée de la nominalisation dans JURIFRAN nous incite à penser qu'il s'agit d'une caractéristique propre du discours scientifique pédagogique.

Concrètement, sur les 77 verbes retenus, seules 7 nominalisations sont plus fréquentes que la structure collocative correspondante.

Ex. : *Appliquer une règle* (13) – *application d'une règle* (31)
Appliquer une loi (12) – *application d'une loi* (41)

Notons tout de même que la nominalisation utilisée sans l'actant COD – dont le rôle sémantique est «patient» – est régulièrement plus fréquente que la forme verbale.

Ex. : *Codifier un texte* (9), *codifier une loi* (8), *codifier un décret* (5)
Codification (90)
Codification d'un texte (1), *codification d'une loi* (1),
codification d'un décret (1)

Cela dit, il faut être aussi conscient que le substantif résultant de cette nominalisation peut être polysémique et que ses emplois peuvent sortir du contexte des emplois de la collocation verbale.

Ex. : *Citer une loi* (4) – *citation* (13)

□ *citation d'une loi* (0) >< *citation* (à comparaitre)

Imposer une règle (6) – *imposition* (13)

□ *imposition d'une règle* (0) >< *imposition* (voir impôts)

Par ailleurs, pour certains verbes, ce même substantif issu de la nominalisation n'est jamais utilisé avec le sens du verbe collocatif.

Ex. : *Affecter un acte* (6) >< *affectation* (25)

□ *affectation d'un acte* (0)

Dans les contextes où *affectation* apparaît dans le corpus, ce terme prend le sens de 'décision par laquelle une autorité assigne quelqu'un ou quelque chose à un usage particulier' alors qu'*affecter* dans la collocation *affecter un acte* prend celui de 'produire une modification, souvent négative'.

Du point de vue didactique, nous allons réutiliser ces données pour travailler la nominalisation, souligner les cas où elle est possible ou inacceptable ou encore inexistante.

6. Présentation de la plateforme numérique

La plateforme que nous avons élaborée ne se présente pas comme un cours de français juridique complet mais plutôt comme un complément à un cours de ce type, les objectifs principaux de ce travail étant :

- Présenter, sur support informatique, un ensemble d'exercices avec les collocations juridiques que nous avons sélectionnées ;
- Observer jusqu'à quel point il était possible de faire converger sur un produit concret le résultat des analyses terminologiques effectuées et les exigences didactiques et pédagogiques en la matière.

Les deux applications que nous avons utilisées pour construire notre plateforme d'exercices – *Hot Potatoes* et *Moodle* – ne sont pas très récentes, mais elles ont déjà fait leurs preuves dans le monde de l'éducation.

Le programme *Hot Potatoes*, que nous pouvons télécharger sur le site <http://web.uvic.ca/hrd/hotpot/>, a été développé par l'Université de Victoria au Canada, plus concrètement, par le groupe de recherche et développement du Centre informatique et média. C'est probablement le programme le plus utilisé par les sites éducatifs. Trois raisons expliquent ce succès :

- sa simplicité ;
- la diversité des exercices qu'il est possible de créer ;
- son caractère gratuit⁹.

9. Le logiciel n'est pas réellement gratuit. Il s'agit en fait d'un programme *freeware* ou d'un gratuitiel. Il est libre d'accès pour tous ceux qui vont l'utiliser dans un contexte éducatif mais, en revanche, ces mêmes utilisateurs doivent s'enregistrer et postérieurement diffuser leurs exercices sur l'Internet.

Ce programme propose cinq outils pour la construction d'exercices autocorrectifs.

- JCloze pour élaborer des exercices avec des textes lacunaires ;
- JCross pour créer des mots croisés ;
- JMatch pour construire des exercices d'appariement ;
- JMix pour faire des puzzles, où l'élève devra ordonner des segments phrastiques ou des mots ;
- JQuiz pour construire des questionnaires à choix multiples (QCM) et des questionnaires à réponses courtes.

Quant à *Moodle*¹⁰, acronyme de *Modular Object-Oriented Dynamic Learning Environment*, il s'agit d'une plateforme éducative où interagissent enseignants, apprenants et contenus, en se présentant comme un environnement d'apprentissage en ligne. Créée par l'Australien Martin Dougiamas en 2002¹¹, cette application se fonde sur les propositions du constructivisme pédagogique. Concrètement, à partir d'une interface conviviale où se combinent des fonctions pédagogiques et communicatives avec un système de gestion de contenu, *Moodle* permet d'intégrer des ressources (fichiers de divers formats, pages web, pages de texte, etc.) et des activités (forums, clavardage, devoirs, glossaires, bases de données, wikis, tests, etc.). Il permet aussi d'obtenir des rapports d'activités et d'évaluer en ligne l'apprentissage des apprenants.

Vu les nombreux avantages présentés par *Moodle* et sa popularité auprès des communautés éducatives, nous l'avons choisi comme

interface où seront importés nos exercices sur les collocations verbales de la langue juridique.

La petite centaine d'exercices que nous avons créés ont donc été logés sur une plateforme *Moodle*¹². La page qui réunit les exercices porte le nom de « Terminologie juridique française ».

Pour y accéder, il est nécessaire de s'inscrire sur la plateforme *Moodle* en fournissant les différentes données demandées par le système¹³. De plus, pour entrer sur la page « Terminologie juridique », le mot de passe suivant est nécessaire : *collocation*.

Nous avons souhaité construire une plateforme qui se présente comme un outil où l'apprenant travaille en complète autonomie pour la simple raison que le profil typique du public qui participe à ce genre de formation dispose d'une disponibilité assez réduite en raison de ses occupations professionnelles.

Ensuite, nous avons rapidement défini que les activités à présenter sur cette même plateforme devraient développer particulièrement la compétence lexicale en langue juridique, en s'attardant sur les collocations, consciente du fait que l'acquisition de nouvelles structures terminologiques (c.-à-d. les collocations) en français juridique, malgré sa difficulté, améliore considérablement les compétences de réception et de production écrite et orale.

10. Voir <http://moodle.org/> (consulté le 26-9-2012).

11. La dernière version de l'application date de juin 2012.

12. Nous les retrouvons en cliquant sur le lien suivant : <http://www.fcsh-elearning.edu.pt/flv/>

13. <http://www.fcsh-elearning.edu.pt/flv/login/index.php>.

Enfin, nous avons décidé de cibler, par le biais de ces exercices, l'acquisition d'un niveau B2 en compétence lexicale spécialisée. En d'autres mots, les activités proposées sont réalisables par des apprenants qui ont déjà un niveau B1 acquis en langue française. En réalité, ce niveau B2 correspond au niveau validé par le diplôme de français professionnel-juridique de la CCIP ; il est donc pertinent que les différentes activités de la plateforme s'encadrent dans une préparation à l'obtention de ce même diplôme.

Pour ce qui est de la méthodologie, nous avons privilégié les exercices directs qui reprennent le contexte d'apparition de la collocation, ainsi que l'approche onomasiologique.

Ces contextes sont authentiques, retirés directement du corpus que nous avons constitué. Nous avons veillé à en varier les auteurs et à les introduire aussi bien dans les énoncés des exercices, que dans les commentaires, en guise d'exemples d'emploi. Dans certains cas, il a été nécessaire d'altérer légèrement le texte original mais toute

modification apportée a été signalée entre crochets dans le texte. Notons que tous les contextes ne se prêtent pas toujours à une didactisation : phrases trop longues, structures syntaxiques complexes, etc.

En ce qui concerne l'approche onomasiologique, rappelons que, dans les exercices présentés, l'apprenant est souvent conduit, sur la base des informations fournies dans l'énoncé, à trouver la forme linguistique correcte en langue étrangère. Toutefois les exercices d'ordre sémasiologique ne sont pas négligés et sont tout de même disponibles sur la plateforme.

En termes de programmation des contenus, nous avons sélectionné huit problématiques qui se présentent au moment de l'acquisition des collocations. Ces problématiques renvoient, d'une part, à la polysémie interne, à la paronymie et au degré de figement des collocations, thèmes développés au moment de la sélection des structures collocationnelles, et, d'autre part, aux fonctions lexicales, au schéma actanciel, à la nominalisation et à la traduction FR-PT, thèmes développés lors de l'analyse des collocations retenues pour la didactisation.

De cette manière, la page « Terminologie juridique française », après une brève introduction, comporte huit sections distinctes :

- Les exercices sur le degré de figement des collocations

– Les exercices de traduction PT>FR

– Les exercices sur l'usage discursif des collocations

Pour chaque section, nous avons proposé quelques exercices sur les structures qui nous semblaient les plus problématiques pour un apprenant lusophone. Nous avons aussi fourni une petite introduction qui permet à l'apprenant de savoir sur quelle problématique il va travailler. Pour ce faire, il a été nécessaire de banaliser certains concepts linguistiques afin de les rendre plus accessibles à un public d'étudiants, non nécessairement initiés à la linguistique.

L'application *Jquiz* et les types d'exercices qu'elle permet de produire (QCM et questionnaires à réponses courtes) ont été privilégiés pour la didactisation des collocations. Néanmoins, nous avons veillé à introduire des activités différentes (textes lacunaires, exercices d'appariement, mots croisés) afin d'éviter la monotonie.

Dans l'objectif de rendre la plateforme plus interactive, nous avons ajouté à chaque section un forum permettant une discussion asynchrone entre l'apprenant et l'enseignant sur d'éventuels doutes concernant les exercices. Par ailleurs, certains exercices comportent, comme aide, des liens renvoyant l'apprenant vers des sites où ce dernier pourra trouver des informations concernant les collocations étudiées. Il s'agit principalement de dictionnaires ou glossaires comme le *Lexique des termes juridiques* du Ministère de la Justice québécois¹⁴.

En résumé, nous avons cherché à concrétiser sous la forme d'un dispositif d'exercices, la synthèse que nous avons établie à partir de travaux récents en terminologie et en didactique des langues de spécialité, par rapport à une problématique particulière : l'acquisition des collocations de la langue juridique à des apprenants lusophones. Cependant, nous reconnaissons que cet ensemble d'exercices mériterait d'être enrichi, en offrant des activités qui portent sur les collocations d'autres domaines conceptuels du droit et/ou qui favorisent le développement d'autres compétences. Néanmoins, nous pensons avoir, dans ce projet, déblayé le terrain, permettant ainsi ultérieurement les modifications nécessaires à l'optimisation de cette plateforme.

14. Voir <http://www.justice.gouv.qc.ca/francais/publications/generale/termes/index.htm> (consulté le 21/11/2012).

7. Conclusion

Premièrement, cette recherche nous a montré qu'il serait erroné de se limiter à la seule collocation – phénomène linguistique d'ailleurs difficilement cernable – mais, au contraire, qu'il est préférable de considérer chacun de ses éléments comme les pièces d'un ensemble plus vaste. Nous avons notamment fait référence dans ce travail à la structure actancielle qui dérive de théories extrêmement pertinentes dans le cadre de l'E/A des collocations.

Deuxièmement, nous avons remarqué que, régulièrement, la technologie informatique ne résout pas tous les problèmes. Elle possède ses limites, que ce soit au niveau du traitement des données terminologiques ou au niveau de la didactisation de ces mêmes données sur un support numérique. Le regard critique d'un terminologue, d'un spécialiste, d'un enseignant et d'un didacticien continue d'être nécessaire.

Troisièmement et à la suite de ce qui vient d'être énoncé, nous affirmons que la collaboration entre ces spécialistes (terminologue, spécialiste, enseignant et didacticien) est indispensable pour développer de produits pédagogiques de qualité. Malheureusement, ce n'est pas toujours le cas et c'est dans le sens de cette coopération que nous avons travaillé, en associant, sur le même plan, recherches récentes en terminologie et en didactique des langues étrangères et expérience professionnelle et en tirant parti des meilleures contributions de ces différentes perspectives.

Quatrièmement, il convient d'admettre que la recherche présentée dans ces pages mériterait de plus amples développements. La collaboration d'autres chercheurs serait souhaitable pour améliorer qualitativement et quantitativement les résultats déjà obtenus. Toutefois, nous pensons avoir apporté notre modeste pierre à l'édifice du savoir, en comblant une lacune existant dans le cadre de l'enseignement/apprentissage de la langue juridique française, l'un de nos objectifs étant de construire un produit concret « prêt à l'emploi ».

Comme le dit bien Damette (2007 : 193), les études autour du français juridique ne sont pas encore à la hauteur des enjeux politiques et économiques de la diffusion du droit et du français dans le monde. Et nous en dirons de même pour le portugais juridique, vu l'importance croissante que prend la langue de Camões à l'échelle mondiale. Pour contrer ce que Renard (2000) appelle la *macdonalisation* linguistique, il est crucial d'affirmer la présence d'autres langues dans le monde des sciences et des techniques. Aider à développer les méthodologies d'E/A d'une langue de spécialité est, en quelque sorte, renforcer sa diffusion comme langue privilégiée de communication. Et c'est dans cet axe que nous avons inscrit notre recherche.

Bibliographie

- BOCQUET (C.) 2008, *La traduction juridique – Fondement et méthode*, Bruxelles, De Boeck Université, coll. Traducto.
- COHEN (B.) 2011², *Lexique des cooccurrents - Bourse et conjoncture économique*, Montréal, Linguattech.
- CORNU (G.) 20003, *Linguistique juridique*, Paris, Montchrestien, col. Domat. Droit privé.
- DAMETTE (É.) 2007, *Didactique du français juridique – français langue étrangère à visée professionnelle*, Paris, L'Harmattan.
- DOLATA-ZAROD (A.) 2011, « La prédication dans la langue du droit » in *Romanica Cracoviensia*, 11, pp. 99-106.
- EURIN-BALMET (S.) et al. 1992, *Pratiques du français scientifique*, Paris, AUPELF-UREF.
- GIRALDEZ (J. C.-E.) 2007, *Las colocaciones léxicas en el lenguaje jurídico del derecho civil francés*, thèse de doctorat – Universidad Complutense de Madrid, Facultad de Filología, Departamento de Filología Francesa.
- GROSS (G.) 1994, « Classes d'objets et description des verbes » in *Langages*, Paris, 115, pp. 15-30.
- GROSS (G.) 1996, *Les expressions figées en français*. Paris, Ophrys.
- GROSS (G.) 1998, « Pour une véritable fonction 'synonymie' dans un traitement de texte », in *Langages*, Paris, 131, pp. 103-114.
- LE PESANT (D.) 1998, « Introduction aux classes d'objets », in *Langages*, Paris, 131, pp. 6-33.
- LERAT (P.) 2002a, « Qu'est-ce qu'un verbe spécialisé ? Le cas du droit », in *Cahiers de Lexicologie*, Paris, 80, pp. 201-211.
- LERAT (P.) 2002b, « Vocabulaire juridique et schémas d'arguments juridiques » in *Meta*, Montréal, 47-2, pp. 155-162.
- LERAT (P.) 2004, « Web et terminologie philologique », in *Linguistica Antverpiensia*, Anvers, 2004-3, pp. 231-244.
- LERAT (P.) 2005, « Le prédicat sémantique droit sur le web », in *Linx*, Paris, 52, pp. 155-161.
- LERAT (P.) 2006, « Terme et microcontexte. Les prédications spécialisées », in D. Blampain et al., éd., *Mots, termes et contextes. Actes des septièmes Journées scientifiques du réseau de chercheurs Lexicologie, Terminologie, Traduction*, Bruxelles 8-10 septembre 2005, Paris, Éditions des Archives Contemporaines et Agence universitaire de la Francophonie, coll. Actualité scientifique, pp. 89-98.
- LERAT (P.) 2008, « Restrictions paradigmatiques et traduction des schémas d'arguments », in *Meta*, Montréal, 53-2, pp. 434-442.
- LERAT (P.) 2009, « La combinatoire des termes. Exemple : Nectar de fruit », in *Hermes – Journal of Language and Communication Studies*, Aarhus, 42, pp. 211-232.
- L'HOMME (M.-Cl.) 1997, « Méthode d'accès informatisé aux combinaisons lexicales en langue technique », in *Meta*, Montréal, 42-1, pp. 15-23.
- L'HOMME (M.-Cl.) 2004, *La terminologie : principes et techniques*, Montréal, Presses Universitaires de Montréal, coll. Paramètres.
- L'HOMME (M.-Cl.) et al. 2011, « Annotation syntaxico-sémantique de contextes spécialisés : application à la terminographie bilingue » in M. Van Campenhoudt et al., éd., *Passeurs de mots, passeurs d'espoir – Lexicologie, terminologie et traduction face au défi de la diversité, Actes des 8^e journées scientifiques du Réseau LTT*, Lisbonne, 15-17 octobre 2009, pp. 651-670.
- L'HOMME (M.-Cl.) 2012, « Le verbe terminologique : un portrait de travaux récents », in *Actes du 3^e Congrès mondial de Linguistique Française*, Lyon, 4-7 juillet 2012 http://www.shs-conferences.org/index.php?option=com_toc&url=%2Farticles%2Fshsconf%2Fabs%2F2012%2F01%2Fcontents%2Fcontents.html (consulté le 25-2-2013).
- LOFFLER-LAURIAN (A.-M.) 1983, « Typologie des discours scientifiques : deux approches », in *ÉLA*, 51, pp. 8-20.
- LORENTE (M.) 2002, « Verbos y discurso especializado », in *ELiEs* 16 <http://elies.rediris.es/elies16/Lorente.html> (consulté le 25 février 2013).
- MELČUK (I.) 2003, « Les collocations : définition, rôle et utilité », in *Les collocations lexicales : analyse et traitement*, F. Grossmann et A. Tutin, Amsterdam, De Werelt, pp. 23-31.
- MOIRAND (S.) 1993, « Autour de la notion de didacticité », in *Carnets du Cediscor*, 1 <http://cediscor.revues.org/600> (consulté le 11 avril 2013).
- PIMENTEL (J.) 2012, « Identifying Equivalents of Specialized Verbs in a Bilingual Comparable Corpus of Judgments: a Frame-based Methodology », in *Proceedings of LREC 2012*, pp. 1791-98.

PIMENTEL (J.) *et al.* 2011, « General and specialized lexical resources: a study on the potential of combining efforts to enrich formal lexicons », in *International Journal of Lexicography*, Oxford, 25 (2), pp. 152-190.

RENARD (R.) 2000, *Une éthique pour la francophonie. Questions de politique linguistique*, Paris, Didier Érudition.

VERLINDE (S.) *et al.* 2003, « Les collocations dans les dictionnaires d'apprentissage : repérage, présentation et accès », in *Les collocations : analyse et traitement. Travaux et Recherches en Linguistique Appliquée*, série E - n°1, F. Grossmann et A. Tutin, Amsterdam, De Werelt, pp. 105-115.

VERLINDE (S.) *et al.* 2006, « Corpus, collocations et dictionnaires d'apprentissage », in *Langue française*, Paris, 150, pp. 84-98.

La traduction des aspects juridiques de la loi Gacaca du Rwanda Regards croisés

Télesphore NGARAMBE
Faculté des lettres, médias et sciences sociales
Université nationale du Rwanda

Résumé

Dans le présent article, l'auteur applique l'approche socioculturelle et contextuelle à la traduction des aspects juridiques incorporés dans la loi Gacaca du kinyarwanda vers le français et vers l'anglais. Les aspects identifiés lors de l'analyse sont, entre autres, la nomination du comité des juges *inyangamugayo* dans un tribunal *Gacaca*, le quorum requis pour que le comité de la juridiction *Gacaca* au niveau de la Cellule puisse siéger, les critères de sélection des juges dans le *Gacaca* moderne et la compétence de l'assemblée générale de la juridiction *Gacaca* au niveau des Cellules. Les résultats de cette étude révèlent que ces aspects sont difficiles à traduire. Cette difficulté est due à la nature même de la loi du *Gacaca* moderne, qui englobe des aspects juridiques issus de la tradition aussi bien que de la modernité. Par conséquent, la traduction de tous ces aspects nécessite une analyse approfondie du contexte socioculturel et linguistique dans lequel cette loi a été produite.

Mots-clés :

traduction, droit, tradition, contexte, système.

1. Introduction

La traductibilité des textes juridiques a toujours été l'une des préoccupations majeures pour les traducteurs et les spécialistes de la traduction. Ce domaine devient beaucoup plus compliqué lorsqu'il traite des textes composés par un mélange d'aspects issus de plus d'un système juridique. La situation devient encore plus complexe quand il s'agit de traduire un système de droit coutumier comme le *Gacaca* du Rwanda qui était tombé en désuétude depuis très longtemps. Il a été rétabli en 2001 en vue de faire face aux conséquences néfastes du génocide qui s'est produit au Rwanda en 1994.

Gacaca est un mot en kinyarwanda qui désigne normalement une sorte d'herbe ou une pelouse traditionnelle où les gens s'asseyaient pour discuter de problèmes divers ou régler certains différends. Étymologiquement, le terme vient du mot *umucaca* que Jacob (1987) définit comme *une espèce d'herbe à gazon dont le nom scientifique est cynodon dactylon*. Cette notion est utilisée d'une façon métonymique pour faire référence à un système juridique que Mangquku (2004) décrit comme :

a traditional court where perpetrators and victims resolve their differences before a community and a panel of eminent persons [un tribunal traditionnel où les bourreaux et leurs victimes règlent leurs différends devant les membres de la communauté et un comité de personnalités éminentes]¹.

1. Toutes les traductions, sauf indication contraire, sont de l'auteur.

Plus précisément, selon Vanhove (1941), ce droit coutumier Rwandais se composait des règlements de la société en matière de succession, de mariage, d'obligations et contrats. Conformément à ce système, le roi était l'autorité suprême du pays, et à l'époque, la plupart des affaires pénales et civiles étaient réglées à l'amiable ou en famille au moyen du système dénommé *Gacaca*.

Après le génocide de 1994, les tribunaux classiques rwandais, limités en nombre et en capacité, ont été débordés. En guise de solution, le *Gacaca* a été restauré et juxtaposé au système juridique moderne en usage au pays. Dans le but de rendre le *Gacaca* accessible à la plus large majorité de la population en l'adaptant à la politique de multilinguisme (MINEDUC 2003 : 14) préconisée par le gouvernement rwandais, la loi a été traduite en français et en anglais – deux des trois langues officielles du pays. Néanmoins, la différence de famille de langues et de système juridique complique la traduction de la loi *Gacaca* dans ces langues étrangères. Le système juridique traditionnel rwandais diffère sensiblement des autres systèmes juridiques impliqués dans la traduction de cette loi en matière de procédure, de structure, de terminologie, d'administration, etc. Pour l'ajuster au contexte actuel, le *Gacaca* a été revisité et transformé en un système juridique qui combine tradition et modernité. C'est ce nouveau système que désignerons, dans cet article, *Gacaca* moderne, pour le distinguer du *Gacaca* traditionnel.

Cette hybridation s'est avérée inévitable car, dans les juridictions du *Gacaca* traditionnel, les procédures étaient principalement

orales — sans aucune codification des dispositions — et basées sur la jurisprudence. Les juridictions du *Gacaca* moderne sont en fait une adaptation du *Gacaca* traditionnel pour relever les défis exceptionnels du contexte post-génocide qui requiert la poursuite et le jugement d'un nombre particulièrement important de suspects du génocide.

À l'instar de l'ancien, le *Gacaca* moderne incarne la justice participative tout en visant son objectif primordial de réparer le tissu social. Néanmoins, le *Gacaca* moderne diffère considérablement du modèle original dans la mesure où il s'occupe de crimes graves, comme celui de génocide. Toutes ces différences rendent très complexe la traduction de ce système juridique vers les langues étrangères. Dans cet article nous nous proposons (1) d'explorer les textes de la loi Gacaca et de les analyser dans leur contexte socioculturel ; (2) d'identifier et d'analyser des problèmes rencontrés dans la traduction de la loi *Gacaca* ; (3) de fournir des solutions appropriées afin de faciliter une meilleure compréhension du système *Gacaca* par le public cible.

2. Présentation des systèmes juridiques majeurs impliqués

Les systèmes juridiques clés impliqués dans cette étude ont été identifiés sur la base des textes présentés en versions parallèles dans les trois langues de la loi *Gacaca* publiée dans le *Journal officiel de la République du Rwanda*, 43^e année, numéro spécial du 19 juin 2004. L'analyse de ces textes nous a permis d'identifier un nombre important d'aspects spécifiques du système juridique français, anglais et rwandais.

C'est pourquoi une description sommaire de ces systèmes en rapprochement avec le système *Gacaca* est indispensable en préliminaire.

2.1. Le système juridique français

Comme de nombreux autres systèmes juridiques, le système juridique français n'était initialement pas écrit. Plus tard sous l'influence romaine et anglaise, certaines collections de coutumes ont été consignées par écrit. En premier lieu, cela a été fait *under the authority of local seigneurs (lords), and later under royal ordinance [sous l'autorité des seigneurs et plus tard sous l'ordonnance royale]* (Dadomo et Farran 1996 : 5). Cette initiative a scindé la France en deux parties de droit : le pays du droit écrit, dans le sud, où le droit romain avait la plus grande influence, et le pays du droit coutumier dans le nord.

En raison de leur contact historique – pensons aux conquêtes et, actuellement, à l'adhésion à l'Union européenne – le système juridique français ressemble au système juridique anglais à quelques différences près. Le Rwanda, lui, a été en contact avec le système juridique français dès l'arrivée des colonisateurs belges au début du XIX^e siècle. Depuis cette époque jusqu'à l'indépendance, en 1962, le Rwanda a connu un double système de droit : d'une part, les lois, décrets et ordonnances rédigés en français et adoptés par le législateur colonial, et, d'autre part, les autres lois émanant du pouvoir législatif réel détenu par les autorités coutumières avant l'arrivée du colonisateur. Comme illustré dans le tableau ci-dessous, la première Constitution du Rwanda de 1962, en son article 99, a mis en place une structure judiciaire calquée sur celle de la France.

FRANÇAIS	Kinyarwanda
<i>Tribunal de canton</i>	<i>Urukiko rwa kanto</i>
<i>Tribunal de première instance</i>	<i>Urukiko rwa mbere rw'iremezo</i>
<i>Cour d'appel</i>	<i>Urukiko rw'ubujurire</i>
<i>Cour de cassation</i>	<i>Urukiko rusesa imanza</i>

De la plus basse à la plus haute juridiction, ces institutions portent des noms typiquement français pour lesquels la traduction en kinyarwanda a généralement eu recours à la technique du calque ou de l'emprunt. Dans la plupart des cas, la même approche a été appliquée à la rédaction et à la traduction de la loi régissant le *Gacaca* moderne.

2.2. Le système juridique anglais

Le système juridique anglais a été énormément influencé par d'autres systèmes juridiques, particulièrement par les systèmes romain, anglo-saxon et français. Selon Alcaraz et Hughes (2002 : 5), le droit anglais est né d'un système qui s'est développé au moyen-âge quand le latin, renforcé par la puissance et le prestige de l'Église romaine, était la *lingua franca* dans toute l'Europe. Le droit romain avait été si bien écrit que tous les juristes le prenaient pour modèle.

Contrairement à de nombreux pays dont les systèmes juridiques sont basés sur le code continental ou napoléonien divisé en quatre parties distinctes (le droit civil, le droit pénal, le droit administratif et le droit du travail), le système anglais fait une distinction entre seulement deux sortes de droit : civil et pénal. Le système juridique anglais opère également une distinction entre les termes

tribunal et *cour*. Selon Bailey et Gunn (1996 : 16), le terme *tribunal* peut être utilisé d'une façon générale pour signifier n'importe quelle assemblée judiciaire, y compris une cour. Dans ce contexte, le mot *tribunal* est utilisé pour faire allusion à des tribunaux spécialisés ayant des pouvoirs spéciaux.

Même si certains Rwandais connaissaient le droit anglais bien avant, l'influence de ce dernier sur le droit rwandais ne s'est accentuée qu'après 1994 avec le rapatriement massif des Rwandais en provenance des pays anglophones et l'officialisation de l'anglais³. Cette officialisation a fortement influé sur la rédaction et la traduction des lois du pays, y compris la loi régissant le *Gacaca* moderne.

2.3. Systèmes juridiques coutumiers et modernes rwandais

Le système de droit coutumier rwandais consiste en un système monolithique appelé *Gacaca* que Muberanziza (1998) décrit comme “[...] *urukiko rwo mu muryango cyangwa se rw'abaturanyi bigiragamo ibibazo bivutse hagati y'abantu aba n'aba maze bigakemurwa mu bwumvikane ntawe uhutajwe akaba ari na byo byatumaga ubwizerane n'ubusabane byarakomezaga kuranga imiryango*”. [Gacaca est un tribunal de famille ou de voisins dans lequel on règle pacifiquement des litiges. C'est un mécanisme de restauration de la confiance mutuelle et de la compréhension entre les familles].

3. Voir l'article 5 de la Constitution de la République du Rwanda (2003).

Comme mentionné plus haut, lorsque les colonisateurs belges sont arrivés dans le pays, leurs traditions juridiques eu progressivement une très grande influence sur le système judiciaire rwandais. À côté du droit coutumier, les colons ont mis en place d'autres lois pour régir les différents domaines de la vie nationale. *[D]epuis l'indépendance, outre les dispositions de la législation coloniale non abrogée, nous avons des dispositions ayant valeur de loi : loi, ordonnance-loi, décret-loi* (Ruhashyankiko 1977 : 57). Toutefois, comme Ruhashyankiko l'a déploré, le système que les colonisateurs ont imposé n'était pas bien adapté au contexte local : *[a]u système coutumier est venu se superposer sans s'y substituer un droit colonial mal adapté et inapproprié* (Ruhashyankiko 1977 : 57).

Contrairement au système traditionnel oral, le droit moderne rwandais est un système écrit hérité de l'autorité coloniale. Il émane des lois promulguées par le Parlement, des règlements, de la jurisprudence et de la coutume. Le système judiciaire moderne rwandais est donc le résultat de l'hybridation de deux espèces principales : le système juridique rwandais et les systèmes français et anglais. En plus de la langue nationale, le kinyarwanda, le Rwanda élabore et promulgue actuellement ses lois en français – la langue du colonisateur – et en anglais – la langue imposée par le contexte actuel de la mondialisation⁴.

Cette situation complique le travail du traducteur qui doit servir de médiateur entre ces trois systèmes juridiques différents. La

section suivante est consacrée à l'élaboration d'un cadre théorique et méthodologique adapté à la traduction de la loi régissant le Gacaca moderne, cette loi qui est, comme nous venons de le voir, un véritable mélange d'aspects juridiques traditionnels et modernes.

3. Démarche méthodologique et théorique

Cette étude adopte la théorie développée par Bassnett et Lefevere en 1990 et connue en anglais sous le nom de *cultural turn*. Dans cette théorie, Lefevere (1992 : 17) considère la langue comme *the expression and repository of a culture [l'expression et la mine de la culture]*. Dans le même ordre d'idées, selon Bassnett et Lefevere (1990 : 8) *neither the word, nor the text, but the culture becomes the operational 'unit' of translation [ce n'est pas le mot, ni le texte, mais la culture qui devient l'unité opérationnelle de la traduction]*.

Pour l'examen et l'analyse du texte de départ, cet article applique les approches d'analyse socioculturelles recommandées par Williams et Chesterman (2002), celles d'intertextualité et d'interdiscursivité préconisées par Fairclough (2003) et adoptées par Bryman (2008). Ces approches considèrent les *texts in terms of different discourses, genres and styles they draw upon and articulate together [textes en ce qui concerne les différents discours, genres et styles dont ils inspirent et articulent ensemble]* (Fairclough 2003 : 3). À cet égard, selon Bryman (2008 : 509) l'établissement des relations intertextuelles *enables a focus on the social and historical context in which discourse*

4. La prédominance de l'anglais sur les autres langues dans le monde est largement expliquée dans l'article « The Rise of English: The Language of Globalization in China and the European Union » par Anne Johnson (2009).

is embedded [permet de mettre l'accent sur le contexte social et historique dans lequel le discours est incorporé].

Sur le plan méthodologique, l'étude examine les textes de départ (en kinyarwanda) et d'arrivée (en français et en anglais) de la loi régissant le *Gacaca* moderne. L'analyse de cette loi a permis d'identifier les aspects juridiques qui causeraient des problèmes aux traducteurs. Pour arriver à la traduction de la présente loi, cet article exploite primordialement les données primaires qualitatives recueillies auprès de 30 enquêtés. Ces derniers ont été sélectionnés à choix raisonné sur la base de leur maîtrise des principaux systèmes juridiques impliqués dans cette étude. Plus précisément, la sélection de ces enquêtés était fondée sur les critères suivants : participation à l'élaboration de la loi et à sa mise en exécution, expertise culturelle, défense juridique, activisme pour les droits de la personne, compétence linguistique, connaissances de l'histoire, expérience de la vie, etc.

Pour collecter les données, un guide d'entretien a été administré auprès d'au moins une personne employée aux institutions nationales ou internationales échantillonnées : le Ministère de la Justice, le Ministère des Sports et de la Culture, le Service national des juridictions *Gacaca* (SNJG), la Commission nationale de lutte contre le génocide (CNLG), le Barreau national, Human Rights Watch et Amnesty International. En plus, le guide d'entretien a été utilisé pour recueillir les idées auprès de trois spécialistes des trois systèmes juridiques impliqués dans cette étude (rwandais, anglais et français). Ces spécialistes comprenaient des juristes, des juges du *Gacaca* moderne, des historiens, des linguistes, des sociologues et des anthropologues.

L'analyse a appliqué une approche socioculturelle pour tirer profit de toutes les données recueillies dans la documentation et sur le terrain. Cela a permis d'améliorer la traduction déjà faite et, dans la mesure du possible, de formuler un complément d'informations de base au début de chaque aspect problématique identifié pour éclairer davantage sa signification socioculturelle et contextuelle pour le public cible qui ne maîtrise pas le système *Gacaca*. Ce faisant, cette approche épouse l'avis de Newmark (1988 : 23) qui recommande de « to supplement the linguistic level, the text level with the referential level, the factual level with the necessary additional information (no more) from this level of reality, the fact of the matter » [compléter le niveau linguistique, le texte avec le niveau référentiel, le niveau factuel avec l'information supplémentaire nécessaire] en vue de traduire convenablement la réalité et les faits exprimés.

4. Analyse et traduction des aspects juridiques identifiés

Après analyse des textes de la loi régissant le *Gacaca* moderne, quatre aspects difficiles à traduire nous sont apparus : la nomination du comité d'une juridiction *Gacaca*, les critères de sélection d'un *inyangamugayo*, le quorum requis pour qu'une juridiction *Gacaca* de cellule siège, et la compétence de l'assemblée générale d'une juridiction *Gacaca*.

Chaque aspect identifié est mis en relief dans le texte de départ reproduit dans la première colonne d'un tableau, et accompagné, en

deux colonnes, de ses traductions en français et en anglais textuellement recopiées du *Journal officiel de la République du Rwanda*. Chaque tableau est suivi d'une description et d'une analyse, puis d'un tableau final présentant ma traduction, en français et en anglais, produite sur la base de l'approche socioculturelle et contextuelle.

KINYARWANDA	FRANÇAIS	ANGLAIS
<i>Abagize Inteko y'Urukiko Gacaca bitoramo, ku bwiganze burunduye bw'amajwi, Inama Mpuzabikorwa igizwe Perezida, Visi Perezida wa mbere, Visi Perezida wa kabiri, n'abanditsi babiri [...].</i>	<i>Les membres du Siège de la Juridiction Gacaca choisissent parmi eux, et à la majorité absolue, le Comité de Coordination composé d'un Président, d'un premier Vice-Président, d'un deuxième Vice-Président et de deux Secrétaires [...].</i>	<i>Members of the Seat for a Gacaca Court elect among themselves, with a simple majority, the Coordination Committee made up with a President, a first Vice-President, a second Vice-President and two secretaries [...].</i>

4.1. Nomination du Comité d'une juridiction Gacaca (Art. 11)

Le problème de cette disposition réside dans la façon dont le comité de coordination des *inyangamugayo*, juges d'une juridiction Gacaca, est formé. Cet article stipule que les *inyangamugayo* sélectionnent en leur sein, les membres du comité de coordination du siège d'une juridiction Gacaca. Cette procédure est tout à fait différente de celle du système de droit classique où, normalement, ces membres sont nommés par un organisme compétent. Pour le cas du Rwanda, les juges sont nommés par le Conseil supérieur de la magistrature⁵. En outre, cette procédure diffère du système du Gacaca traditionnel qui ne préconisait pas une telle structure de comité, composée d'un

président, d'un vice-président, etc. Cette différence a été clairement signalée dans différents avis de la majorité des enquêtés, en particulier dans les avis des juristes et des avocats :

- Dans le droit coutumier, le président et les autres n'étaient pas élus, on procédait tout simplement par le plus âgé [sic] ;
- Le plus vieux de la colline assistait les autres *inyangamugayo* pour trancher les différends sans qu'il soit nécessairement considéré comme président ;
- Ceci ne concorde pas avec la pratique dans les juridictions classiques parce que lorsque le litige est entendu par une collégialité des juges, celui qui se tient au milieu est appelé président⁶.

La majorité des répondants de la catégorie des anciens, des historiens et des sociologues s'accordent à dire qu'il y avait une sorte d'organisation : les chefs des deux familles impliquées en conflit coprésidaient les séances. Un employé de la Commission nationale de droit de la personne a révélé quelque chose de similaire : *dans le Rwanda traditionnel, les chefs de famille et de ménage étaient en même temps présidents des réunions et des structures du contentieux*. Un conseiller juridique dans les juridictions Gacaca a donné plus de précisions : *il n'y avait aucun ordre hiérarchique comme l'on observe aujourd'hui, mais l'inyangamugayo le plus expérimenté et respectable présidait au siège*.

Pour rendre cette disposition compréhensible par le lecteur de la langue d'arrivée, il est important de souligner la nécessité d'adapter la loi au système moderne de l'organisation des organes de prise de

5. Voir l'Article 157(2) de la Constitution de la République du Rwanda (2003).

6. Données recueillies sur le terrain (précision valable pour toutes les citations d'enquêtés reproduites dans cet article).

décision. La solution est fournie dans l'information placée au début de la disposition tel que présenté dans le tableau ci-après.

FRANÇAIS	ANGLAIS	
<i>[Pour adapter la loi au mode moderne d'organisation], les membres du Siège de la Juridiction Gacaca choisissent parmi eux, et à la majorité absolue, le Comité de Coordination composé d'un Président, d'un premier Vice-Président, d'un deuxième Vice-Président et de deux Secrétaires [...].</i>	<i>[To adapt the law to the modern mode of organisation], members of the Bench of the Gacaca Court shall select among themselves, by simple majority rule, the Coordination Committee made up of a President, a first Vice-President, a second Vice-President and two secretaries [...].</i>	
KINYARWANDA	FRANÇAIS	ANGLAIS
<i>Umuntu wese w'inyangamugayo ufite nibura imyaka maku-myabiri n'umwe (21) kandi utazitwira n'iri Tegeko Ngenga, ashobora gutorerwa kujya mu nzego z'Inkiko Gacaca [...]</i>	<i>Toute personne intègre, âgée de vingt et un (21) ans au moins et remplissant toutes les conditions exigées par la présente loi organique, peut être élue membre des organes des Juridictions Gacaca, [...]</i>	<i>Any person of integrity who is at least 21 years old and meeting all the conditions required by this organic law, can be elected a member of the organ for a Gacaca Court [...]</i>

4.2. Les critères de sélection d'un *inyangamugayo* (article 14§3)

Le problème de cette disposition est qu'elle fixe l'âge comme l'un des principaux critères de choix d'un *inyangamugayo*. Même si cela a en quelque sorte un rapport avec le *Gacaca* traditionnel où l'âge mûr était considéré comme un signe de sagesse et de droiture, une personne âgée de 21 ans semblera trop jeune aux traditionalistes, alors que le droit moderne la considérera comme adulte.

Comme exprimé par la majorité des personnes enquêtées, cette notion de maturité n'a pas été facile à restituer pour les traducteurs qui avaient la tâche de rendre cette idée en français et en anglais. Un certain nombre d'enquêtés parmi les anthropologues, les historiens, les anciens et même les *inyangamugayo* modernes ont critiqué le fait de considérer le critère d'âge (21 ans) comme l'une des conditions préalables pour recruter les juges au *Gacaca*. Pour eux, les *inyangamugayo* traditionnels n'étaient pas élus d'une manière officielle comme c'est le cas dans les juridictions du *Gacaca* moderne. Le statut d'*inyangamugayo* n'était pas le résultat d'une élection, il était partie intégrante de la personne, et ce statut se manifestait dans la façon dont la personne se comportait dans la société. Par conséquent, la procédure actuelle de sélectionner les *inyangamugayo* diffère de celle du passé. Comme un juriste du MINIJUST l'a noté, elle est même différente de celle préconisée par le droit classique. Concernant toujours la question d'âge, un autre enquêté a réagi en ces termes :

Sur le plan international, l'âge de la majorité est un sujet essentiel, mais la maturité aurait été un meilleur critère. Même dans les juridictions du *Gacaca* moderne, dans le processus d'élection de l'Assemblée générale, l'âge est un facteur déterminant. Un candidat âgé de quarante ans avait plus de chances de vaincre celui âgé de 24.

Dans leurs explications, tous les enquêtés ont reconnu que le critère d'âge a été fixé à 21 ans pour essayer de répondre à plusieurs conditions et défis socioculturels inhérents à la période d'après-génocide. Les raisons majeures sont que, pendant ces temps de crise grave, beaucoup de personnes étaient mortes et qu'une grande partie de la

population avait fui le pays, tandis qu'une autre avait trempé dans le génocide. Ainsi, il n'y avait d'autre choix que l'assouplissement de certaines conditions requises pour être sélectionné comme *inyangamugayo*. Pour traduire cet aspect d'âge et de sélection des *inyangamugayo* par la population, il est donc préférable de concilier à la fois la tradition, la modernité, et le contexte de la période post-génocide.

En outre, il est également important de signaler la contradiction et la redondance sous-entendues dans la disposition dont la première phrase commence par *umuntu wese w'inyangamugayo*. Ceci signifierait que cette personne est déjà *inyangamugayo* et donc, qu'elle est censée réunir toutes les conditions requises. Comme solution, il n'est pas nécessaire d'employer les termes *inyagamugayo*, *intègre* et *integrity* au début de la disposition pour qualifier le mot personne, comme on l'a fait dans les traductions qui font l'objet de notre analyse. Si une personne est *inyangamugayo*, elle l'est déjà et il ne devrait pas y avoir d'autres conditions supplémentaires à remplir. En principe, dans la signification de ce concept, il n'y a pas de demi-mesure. À cet égard, il est important de fournir des renseignements généraux afin d'expliquer le contexte et la pertinence de ce critère d'âge adopté par le *Gacaca* moderne. En outre, pour être plus neutre, le terme *inyangamugayo* doit être remplacé par toute personne. Les traductions proposées à cet effet sont présentées dans le tableau ci-après:

FRANÇAIS	ANGLAIS
<i>[Suite aux défis de la période qui a suivi le génocide], toute personne âgée de vingt et un (21) ans au moins et remplissant toutes les conditions exigées par la présente loi organique, peut être choisie par la communauté comme membre des organes des Juridictions Gacaca, sans discrimination aucune, notamment de sexe, d'origine, de religion, d'opinion ou de position sociale [...].</i>	<i>[Due to challenges of the period after the genocide], any person who is at least 21 years old and who fulfils all the conditions required by this organic law, can be selected by the community as a member of the organs for the Gacaca Courts without any discrimination, such as that based on sex, origin, religion, opinion or social position [...].</i>

KINYARWANDA	FRANÇAIS	ANGLAIS
<i>Inama Rusange y'Urukiko Gacaca rw'Akagari iterana ku buryo bwemewe n'amategeko iyo yitabirwe nibura n'abantu ijana (100) mu bayigize.</i>	<i>L'Assemblée Générale de la Juridiction Gacaca de la Cellule ne siège valablement que si au moins cent (100) de ses membres sont présents.</i>	<i>The General Assembly of the Gacaca Court of the Cell only meets legitimately if at least one hundred (100) members are present.</i>

4.3. Quorum requis pour que siège une juridiction Gacaca de cellule (Art. 18)

Dans son article 18, cette loi fixe le quorum requis pour que la juridiction *Gacaca* au niveau de cellule puisse siéger valablement. Cet aspect n'existait pas dans le *Gacaca* traditionnel et il n'est pas non plus reconnu comme tel par le système juridique classique. Dans les deux systèmes, un siège n'a pas nécessairement besoin d'un nombre déterminé de membres présents pour constituer une audience ou statuer sur un cas. Cette nouveauté du *Gacaca* moderne peut entraver la compréhension de cette loi, si le traducteur ne fournit pas les informations nécessaires pour éclairer davantage le contexte.

Le terme *quorum* est un concept très moderne qui dérive du droit positif. Selon un enquêté (avocat), il s'agit d'un nombre minimal requis pour qu'une assemblée puisse prendre une décision valide. Dans le cas du *Gacaca* moderne, l'assemblée est la réunion de tous les membres de la communauté au niveau de la cellule.

Plus précisément, selon les enquêtés, cette condition de quorum s'est avérée nécessaire pour répondre aux exigences de la période

post-génocide qui sont quasiment toutes différentes de celles qui prévalaient dans le Rwanda ancien. À cette époque, les audiences étaient improvisées chaque fois que surgissait un cas nécessitant une intervention des *inyangamugayo*, et seules les parties intéressées et leurs voisins avaient l'obligation morale d'y participer.

Toutefois, la majorité de enquêtés s'accordent sur le fait que, dans le cas du *Gacaca* moderne, il était très opportun de déterminer le quorum de manière à faire participer le plus grand nombre possible de membres de la population et ainsi recueillir suffisamment d'informations en rapport avec le génocide dans le but ultime de la réconciliation nationale.

Suivant la théorie d'interdiscursivité de Fairclough (2003), il ressort de notre analyse que le concept de quorum est un phénomène social construit par les acteurs sociaux (c.-à-d. les législateurs de la loi régissant le *Gacaca* moderne) pour répondre à la situation qui prévalait dans la période post-génocide.

FRANÇAIS	ANGLAIS	
<i>[Pour créer un environnement permettant de révéler la vérité sur le génocide et arriver à la réconciliation], l'Assemblée Générale de la Juridiction Gacaca de la Cellule ne siège valablement que si au moins cent (100) de ses membres sont présents.</i>	<i>[To create an environment that would allow to reveal the truth about the genocide and achieve re conciliation], the General Assembly of the Gacaca Court of the Cell meets legitimately only if at least one hundred (100) members are present.</i>	
KINYARWANDA	FRANÇAIS	ANGLAIS
<i>Inama Rusange y'Umurenge ifite inshingano zikurikira: 1° kugenzura imikorere y'Inkiko Gacaca zose ziri mu Murenge nta wivanga mu mikirize y'imanza ; [...]</i>	<i>L'Assemblée Générale du Secteur exerce les attributions suivantes: 1° superviser le fonctionnement de toutes les juridictions du Secteur sans toutefois s'immiscer dans leur manière de rendre les jugements ; [...]</i>	<i>The general Assembly for the Sector has the following attributions: 1° to supervise the functioning of all Gacaca Courts in the Sector, without meddling in their ways of conducting proceedings; [...]</i>

4.4. La compétence de l'Assemblée générale d'une juridiction *Gacaca* (article 35§1)

Cette disposition définit les attributions de l'Assemblée générale de la juridiction *Gacaca* au niveau de secteur⁷. Le problème majeur identifié est la synchronisation des trois systèmes juridiques (le *Gacaca* traditionnel, le *Gacaca* moderne et le droit positif) dont la structure et les responsabilités des cours et tribunaux diffèrent totalement.

La compétence du *Gacaca* traditionnel se limitait au niveau de la communauté et les cas les plus compliqués devaient être transférés aux chefs ou au roi. Traditionnellement, il n'y avait aucune structure formelle de supervision ni de suivi des décisions des juridictions *Gacaca*. Toutes les structures étaient indépendantes.

La spécificité de la structure et l'appellation des cours et tribunaux sont aussi problématiques. Le problème a été identifié dans la traduction en français qui manque de précision sur ce que le traducteur a appelé *toutes les juridictions de Secteur*. Dans le secteur, il pourrait y avoir plusieurs sortes de tribunaux qui ne rentrent pas dans la structure de *Gacaca*. Comme un avocat l'a fait remarquer, cette structure complique la traduction française de la disposition :

Faute très grave dans cette traduction qui parle de la supervision de toutes les juridictions du Secteur. Avec les lois, il faut éviter toute chose qui prête à confusion. On n'a pas précisé qu'il s'agit des juri-

7. Entité administrative au niveau de base qui vient en troisième lieu après l'*umudugudu* et la cellule. Voir MINALOC (2007 : 9).

dictions *Gacaca* alors qu'il y a beaucoup de secteurs qui comptent en leur sein des juridictions ordinaires et même spécialisées !

À ce sujet, les magistrats et les avocats ont révélé que cette pratique était tirée de la structure judiciaire moderne où les cours et tribunaux supérieurs supervisent ceux des instances inférieures, sans toutefois s'immiscer dans leurs décisions. L'enquête du MINIJUST trouve que cette stipulation a donné trop de pouvoirs au *Gacaca* moderne, et que ceci pouvait avoir un impact négatif sur le fonctionnement des juridictions qui sont sous sa supervision. Pour lever l'inquiétude de l'enquête précédent, le conseiller juridique du CNLG a réagi en soulignant que le concept d'*Inama Rusange, qui est l'Assemblée générale, est chargé des questions administratives et n'a pas la capacité de régler les affaires judiciaires.*

Le point de vue partagé par la plupart des enquêtés est que l'assemblée générale et la supervision des tribunaux sont des aspects spécifiques au système du *Gacaca* moderne. Ainsi l'analyse de ces aspects cadre bien avec la théorie adoptée pour cette étude qui prouve que tous ces phénomènes ont été construits pour faciliter l'organisation de différentes activités des tribunaux du *Gacaca* moderne. À cet effet, la traduction de cette disposition devrait adopter et mettre en exergue cette nouveauté qui a été introduite afin d'assurer l'indépendance des structures du système du *Gacaca* moderne.

5. Conclusion

Cet article applique l'approche socioculturelle et contextuelle à la traduction, du kinyarwanda vers le français et vers l'anglais, de quatre aspects juridiques identifiés dans la loi régissant le *Gacaca* moderne – les aspects analysés et traduits étant la nomination du comité de la juridiction *Gacaca*, le quorum, les critères de sélection des juges et la compétence de l'assemblée générale.

Les informations recueillies auprès des enquêtés ont révélé que certains de ces aspects sont spécifiques de la culture rwandaise tandis que d'autres sont dérivés du système juridique moderne. C'est la configuration de ces aspects qui reflète la nature hybride de la présente loi et qui requiert une attention soutenue lors de l'opération traduisante. Les résultats de cette étude montrent que la meilleure façon de traduire ces aspects qui ont retenu notre attention et d'autres, de moindre importance qui n'ont pas été évoqués ici, est de les conceptualiser et de les comprendre au regard de la société, de la culture et du contexte qui les ont produits. À cet égard, cet article suggère l'antéposition des informations de base nécessaires au début de chaque partie problématique pour éclairer davantage son sens, son contexte et sa pertinence.

FRANÇAIS	ANGLAIS
<i>[En considération de la structure spéciale des Juridictions du Gacaca moderne et en vue d'assurer leur indépendance], l'Assemblée Générale du Secteur exerce les attributions suivantes : 1° superviser le fonctionnement de toutes les juridictions Gacaca du Secteur sans toutefois s'immiscer dans leur manière de rendre les jugements ; [...]</i>	<i>[Considering the special structure of the modern Gacaca court and in order to ensure their independence], the General Assembly for the Sector has the following attributions: 1° to supervise the functioning of all Gacaca Courts in the Sector, without meddling in their ways of conducting proceedings; [...]</i>

Bibliographie

A. Source principale

GOUVERNEMENT RWANDAIS 2004, « Loi organique n° 16/2004 du 19 juin 2004 portant organisation, compétence et fonctionnement des juridictions *Gacaca* chargées des poursuites et du jugement des infractions constitutives du crime de génocide et d'autres crimes contre l'humanité commis entre le 1er octobre 1990 et le 31 décembre 1994 », in *Journal officiel*, 40/40 (15 mars 2001). [En trois versions: kinyarwanda, français et anglais].

B. Sources secondaires

- ALCARAZ (E.) et HUGHES (B.) 2002, *Legal Translation Explained*, Manchester & Northampton, St Jerome Publishing.
- BAILEY (S.H.) et GUNN (M.J.) 1996, *Smith and Bailey on the English Legal System*, London, Sweet & Maxwell.
- BASSNETT (S.) et LEFEVERE (A.) 1990, "Grandmother and the Thousand and One Nights: The 'Cultural Turn' in Translation Studies", in *Translation, History & Culture*, S. Bassnett et A. Lefevere, London, Pinter Publishers.
- BRYMAN (A.) 2008, *Social Research Methods*, Oxford, Oxford University Press.
- DADOMO (C.) et FARRAN (S.) 1996, *The French Legal System*, London, Sweet & Maxwell.
- FAIRCLOUGH (N.) 2003, *Analysing Discourse: Textual Analysis for Social Research*, London, Routledge.
- GOUVERNEMENT RWANDAIS 2003, *La Constitution de la République du Rwanda*, Kigali.
- JACOB (I.) 1987, *Dictionnaire rwandais-français*, I, II & III, Butare, INRS.
- JOHNSON (A.) 2009, "The Rise of English: The Language of Globalization in China and the European Union", in *Macalester International*, 2, pp. 131-168.
- LEFEVERE (A.) 1992, *Translating Literature: Practice and Theory in a Comparative Literature Context*, New York, The Modern Language Association of America.

MANGQUKU (L.) 2004, "A Seat in the Grass. In South African Institute of International Affairs", in *e-Africa: The Electronic Journal of Governance and Innovation*, Johannesburg, 2 (mars 2004), pp. 10-11.

MINALOC 2007, *Rwanda Decentralisation Strategic Framework: Towards a Sector-Wide Approach for Decentralisation Implementation*, Kigali.

MINEDUC 2003, « Education Sector Policy » (http://www.mineduc.gov.rw/IMG/pdf/EDUCATION_POLICY.pdf, dernière consultation 13/2 2012).

MUBERANZIZA (A.) 1998, *Ubucamanza na Gacaca mu Rwanda*, in *Revue scientifique du droit*, Butare, p. 42.

NEWMARK (P.) 1988, *A Text Book of Translation*, New York and London, Prentice Hall.

RUHASHYANKIKO (N.) 1977, *Le droit coutumier rwandais et africain*, Butare, Faculty of Law/ NUR.

VANHOVE (J.) 1941, *Essai de droit coutumier du Ruanda*, Bruxelles, Institut colonial belge.

WILLIAMS (J.) et CHESTERMAN (A.) 2002, *The Map: A Beginner's Guide to Doing Research in Translation Studies*, Cornwall, St Jerome Publishing.

Au cœur de la théorie Sens-Texte

Défigement dans les textes poétiques. Typologies et exemples en français, espagnol, catalan, polonais, russe et biélorusse

Yauheniya YAKUBOVICH
*Laboratoire fLexSem,
Universitat Autònoma de Barcelona*

Résumé

De nos jours, le défigement est une ressource stylistique d'un grand potentiel non seulement dans le langage de la presse et de la publicité, mais aussi dans le langage poétique. Notre recherche en cours aborde le sujet du défigement dans des corpus poétiques en six langues (français, espagnol, catalan ; polonais, russe, biélorusse). Son objet est de systématiser et d'analyser les phrasèmes défigés (collocations, locutions, noms propres, proverbes, pragmatèmes, clichés, citations...) extraits de textes poétiques contemporains, en nous basant sur l'appareil conceptuel de la Théorie Sens-Texte et en adoptant une nouvelle technique d'analyse poétique formelle, avancée par Blanco (2012 et à paraître). Dans cet article, nous présentons les objectifs, la méthode et les premiers résultats de ladite recherche (pour l'instant, seulement en trois langues du projet : français, biélorusse et russe).

Mots-clés :

défigement, phrasème, texte poétique, langues slaves.

1. Introduction à la recherche

Jean-Paul Sartre affirmait que les poètes sont des hommes refusant d'utiliser le langage. Saint-John Perse, à son tour, soutenait que le poète est celui qui rompt avec l'accoutumance... Il est bien évident que le langage poétique défie la norme. Les mots dans la poésie ne sont certainement pas associés de la même manière que nous les combinons dans le langage courant. Nous pouvons rappeler à ce propos la technique des « mots en liberté » d'Arthur Rimbaud, des expérimentations surréalistes, la pratique oulipienne... Parfois, dans les cas où le bon usage exige de nous une combinaison contrainte de lexèmes, les poètes contournent les impératifs de la normativité, en libérant ces lexèmes de leur combinatoire restrictive, comme on peut le voir dans les lignes suivantes :

j'avais le **pied marin** et l'estomac de même
Une traversée en 1922 (Queneau 2006 : 297)

Ce pays je le **connais à peine**
Sandy je le **connais** à plaisir
Histoires des chiens (Prévert 1980 : 206)

c'est l'heure où **la goutte d'eau**
qui **fait déborder le vase**
tombe dans le bec de l'hirondelle
qui ne fait pas le printemps
Les mystères de la chambre noire (Prévert 1980 : 67)

On était **du même bois**
Un peu rustique un peu brut
 Dont on fait n'importe quoi
Sauf naturellement **les flûtes**
Auprès de mon arbre (Brassens 1963 : 75)

Ces anamorphoses que les poètes pratiquent sur les énoncés figés de toute sorte, en les détournant, mettent en relief leur caractère contraint et révèlent le potentiel sémantique et stylistique de l'énoncé qui se dévoile en séquence défigée. Les locuteurs de la langue reconstruisent facilement les séquences originaires sujettes au détournement poétique, mais ne sont pas toujours conscients des mécanismes du défigement. Cependant, à notre avis, ces mécanismes doivent être pris en considération, parce qu'ils relèvent d'une activité métalinguistique.

Dans notre thèse, dont nous proposons ici une brève présentation, nous visons à étudier les énoncés qui résultent du défigement – un procédé linguistique qui consiste à détourner des syntagmes figés d'une langue (dans notre cas, de six langues), un procédé dont les poètes modernes se servent assez massivement.

La présente recherche est élaborée dans le cadre du projet espagnol du *Ministerio de Economía y Competitividad : Frasemas compositivas pragmáticas* (FFI-2010-15229). Le projet dans lequel nous sommes incorporée présente les aspects suivants :

- élaboration de dictionnaires électroniques de lexies et de phrasèmes ;
- formalisation linguistique des phrasèmes compositionnels dans une perspective contrastive
- espagnol, catalan, français, russe, biélorusse et polonais ;
- création et maintenance de grands corpus textuels ;
- analyse des processus de figement et de défigement ;
- réflexion de caractère traductologique sur les équivalences de traduction langues romanes - langues slaves.

Compte tenu des objectifs du projet, nous prenons pour objet de la recherche les énoncés défigés dans les textes poétiques contemporains pour les six langues du projet en une perspective contrastive plurilingue. Ainsi, les objectifs centraux du travail peuvent être formulés de la manière suivante :

- analyser un vaste corpus de poésie au moyen des outils conceptuels développés par les linguistes de la *Théorie Sens-Texte*, mais aussi, le cas échéant, par d'autres théories linguistiques lexicalistes ;
- identifier, recenser et systématiser les cas de défigement dans le corpus poétique de chacune des six langues ;
- fournir des éléments linguistiques formalisés pour définir les fonctions stylistiques du défigement dans l'œuvre d'un auteur ou d'un groupe d'auteurs en utilisant la terminologie de la stylistique et de la théorie littéraire.

Il est à noter que le défigement est largement utilisé aussi dans le langage de la presse et de la publicité. Pourtant, ses fonctions stylistiques et les particularités d'emploi dans les textes poétiques peuvent être différentes de celles des textes publicitaires. Nos observations concernent le défigement en tant que recours poétique.

Nous résumerons maintenant les aspects théoriques et méthodologiques de notre travail (cf. 2) et ensuite, nous procéderons à quelques illustrations pratiques (cf. 3).

2. Méthodologie

2.1. Cadre théorique

Les fondements théoriques de notre recherche peuvent être divisés en trois composantes importantes.

La première composante implique la théorie du figement. Le défigement est un terme très général pour définir un processus de déblocage de la contrainte dans les séquences dites figées. Il est hors de doute que le terme de défigement peut être clairement défini uniquement dans le cas où l'on spécifie ce qu'on comprend par figement. Quant à nous, dans notre travail nous prenons appui sur la notion de figement élaborée au sein de la *Théorie Sens-Texte* (cf. Žolkovskij et Mel'čuk 1965 ; Mel'čuk 1997, etc.). Nous empruntons aussi bien la classification des unités figées que son appareil conceptuel en général.

La deuxième composante est une méthode d'analyse des séquences défigées. Les travaux de Blanco (2012 et à paraître) abordent le sujet du défigement des collocations et des locutions utilisant les concepts de figement provenant de la *Théorie Sens-Texte* mais aussi du modèle de Gaston Gross. Cette approche descriptive et appliquée à des textes poétiques nous semble la plus appropriée pour notre type de recherche.

Troisièmement, d'autres expériences concernant le figement et surtout le défigement contribuent à notre réflexion autour de ce thème, comme les travaux de Gaston Gross (1996) ; García-Page (1989) ;

Pajdzińska (1993) ; Rastier (1997) ; Mejri (1997 et 2009) ; Cabasino (1999) ; Lecler (2006) ; Ben Amor (2006 et 2007) ; Català (2012), etc.

2.2. Concepts de base

Étant donné que notre recherche est basée sur la *Théorie Sens-Texte* (= TST), nous partons du concept de *phrasème*, tel qu'il est défini dans l'un des derniers articles de Mel'čuk (cf. Mel'čuk 2011). Les *phrasèmes*, selon Mel'čuk, sont des énoncés multilexémiques non libres. L'auteur les divise en *phrasèmes lexicaux* et *phrasèmes sémantico-lexicaux*, les premiers incluant les classes des *locutions* et des *collocations*, les derniers incorporant les *clichés*. Les classes des locutions et des collocations sont subdivisées à leur tour en sous-classes : *locutions fortes*, *semi-locutions* et *quasi-locutions*, d'une part ; *collocations standard* et *non standard*, d'autre part. Quant aux clichés, il s'agit, chez Mel'čuk, des *clichés* proprement dits, où viennent s'ajouter des *proverbes* avec un sens compositionnel, mais aussi des clichés contraints par la situation de leur emploi – les *pragmatèmes*.

Ainsi, selon le postulat établi dans notre travail, si on parle de défigement, on comprend le détournement d'un phrasème déterminé : soit d'une locution, soit d'une collocation, soit d'un cliché. En plus, on ne peut pas exclure de l'analyse les formules d'origines identifiables (citations) de même que les noms propres complexes, qui ne sont pas discutés dans les articles des auteurs de la TST. Cependant, aussi bien les citations que les noms propres sont une sorte de phrasèmes et, tout comme les phrasèmes « normaux », ils subissent le détournement poétique.

Afin que l'analyse de phrasèmes défigés soit moins confuse, on étudie séparément les *unités lexicales* (locutions, collocations¹ et noms propres complexes), les *unités phrastiques* (clichés, proverbes, pragmatèmes et citations) et les *unités textuelles* (prières, lettres formelles, certains types de pragmatèmes, citations, etc.).

Les premiers résultats qu'on a atteints consistent en la construction de corpus français et biélorusse et la rédaction à partir de ces corpus d'articles traitant le problème du défigement des collocations. Le corpus russe est maintenant en cours d'élaboration.

2.3. Construction des corpus

Nous travaillons les corpus de façon manuelle. Après avoir dépouillé des textes poétiques sur papier ou en ligne, les données – c'est-à-dire, les exemples des phrasèmes défigés – sont gardées dans les bases de données. Les phrasèmes défigés y sont classifiés selon leur classe et sous-classe. On précise les auteurs des exemples aussi bien que les titres des poèmes et on fait des annotations. Les poèmes où les cas de défigement ont été détectés sont recueillis, à leur tour, dans des fichiers à part.

Comme sources de dépouillement, nous nous servons d'anthologies, de recueils de différents poètes et de pages Web de poésie. En outre, nous avons régulièrement recours aux grands corpus textuels informatisés comme *Frantext* et *Национальный корпус русского языка* {*Corpus national de la langue russe*}² que nous utilisons afin de vérifier la fréquence et

1. Les collocations, quant à elles, forment plutôt deux unités lexicales qu'une.

2. Toutes les traductions des langues slaves en français sont mises entre les accolades.

de possibles contextes d'emploi de certaines séquences figées ou défigées rencontrées précédemment dans les recueils dépouillés.

En ce qui concerne la détection automatique des énoncés défigés, on ne l'exclut pas complètement, même si ce type d'accumulation des cas de défigement, pour l'instant, nous paraît moins efficace que la détection manuelle.

Pour lancer la recherche automatique, nous avons besoin d'une hypothèse – d'une formule préliminaire qu'on introduit dans le formulaire de recherche. Donc, pour trouver les cas de défigement, disons, d'une collocation de structure *Adj. comme Dét. N.* (comme dans *myope comme une taupe*), on doit prévoir comment cette structure peut être défigée, c'est-à-dire, transformée : p.ex. par l'addition ou l'insertion d'un mot, par la modification de telle ou telle propriété grammaticale, etc. Afin de rédiger les formules de requête qui pourraient donner des listes d'énoncés défigés, il faut analyser préalablement une multitude de cas de défigement et examiner tous les possibles indices de défigement de toutes les classes de phrasèmes. Nous sommes d'avis que cette analyse préliminaire peut être réalisée uniquement si l'on possède déjà une liste de données détectées manuellement. À partir de cette liste préliminaire, on peut commencer à suggérer des hypothèses, ce qui exige une recherche à part. Ainsi, les résultats de notre travail pourraient éventuellement servir de base à la recherche pour la détection automatique des séquences défigées.

En fait, nous envisageons, en perspective, de développer une application pour le logiciel de traitement textuel *NooJ* (cf. Silberztein

2002). En ayant introduit dans le programme le corpus des textes poétiques que nous sommes en train de construire, nous visons à élaborer quelques formules qui nous permettent de détecter des cas de transformation de certaines structures contraintes.

Pour poursuivre, nous commenterons les données obtenues pendant la première phase de notre recherche sur la systématisation du matériel de nos *corpus* en trois langues du projet – français, biélorusse et russe.

3. Défigement des phrasèmes : indices et exemples³

3.1. Indices du défigement

Lorsque la décision doit être prise, si une modification que le poète effectue au niveau phonique, lexical, syntaxique ou sémantique de l'énoncé est bien un cas de défigement, nous faisons appel à certains indices. En prenant appui sur nos corpus poétiques français, biélorusse et russe, à partir des critères de défigement, proposés par García-Page (1989 : 60), nous distinguons entre différents mécanismes qui servent d'indices principaux du défigement des phrasèmes (y compris les noms propres et les citations) :

- Substitution lexicale, qui se produit quand une des composantes d'un phrasème est remplacée par un ou des éléments unis à la composante

3. Dans cette division, nous reproduisons partiellement une partie de notre travail, présenté au colloque international FRADITRAD (Saint-Jacques-de-Compostelle, 29-31 octobre 2012)

originaires par des relations sémantiques (antonymie, synonymie, etc.), des relations de paronymie ou homophonie, etc. :

Songe d'une nuit d'hiver

(Queneau 2006 : 129)

- Addition ou insertion à un phrasème originaires d'un élément ou de plusieurs éléments détruisant l'unité sémantique et syntaxique du phrasème, en accentuant un deuxième sens d'une de ses composantes ou de tout l'énoncé :

invasions plus ou moins **barbares**

Souviens-toi du vase de Vix (Queneau 2006 : 113)

- Intersection de deux phrasèmes ou plus qui partagent une ou plusieurs de leurs composantes et qui se défigent l'un l'autre :

La loi d' la pesanteur **est dure** **mais c'est la loi**

Vénus Callipyge (Brassens 1963 : 162)

- Intersection d'un phrasème et d'un syntagme libre qui partagent une composante lexicale :

Je **bats les chiens** et **les tapis**

Comment tuer l'amant de sa femme quand on a été comme moi élevé dans les traditions (Brel 1998 : 331)

- Emploi d'un phrasème dans un contexte qui force l'activation d'un deuxième sens (soit d'une de ces composantes, éventuellement des deux). Ainsi, les composantes du phrasème deviennent des unités sémantiquement et syntactiquement indépendantes :

La dame **a perdu son sourire** dans le bois

Dimanche (Souppault in Décaudin éd. 1983 : 335)

- Commutation grammaticale, quand le défigement se produit à cause du changement de catégorie grammaticale d'un phrasème ou d'une de ses composantes ou bien à cause du changement de propriété grammaticale à l'intérieur de la même catégorie :

Пакусанья локці рыпцяць

Гострым прадметам (Жыбуль 2012 : 138)

{Les **doigts mordus** grincent}

- Altération de l'ordre des composantes, parfois accompagnée de changements de catégorie, de la suppression et/ou l'addition de nouveaux éléments (prépositions, déterminants, etc.) au phrasème originaires :

tu as eu raison d'abandonner le boulevard des paresseux, les estaminets des pisse-lyres, pour l'enfer des bêtes, pour les commerces des rusés et le **bonjour des simples**

Tu as bien fait de partir, Arthur Rimbaud ! (Char in Décaudin éd. 1983 : 439)

- Combinaison de plusieurs modifications à la fois :
Sur mon **brin de laurier** je **m'endors comme un loir**
Les trompettes de la Renommée (Brassens 1963 : 142)

Nous postulons que tous ces indices entrent en jeu à la seule condition que le lecteur (ou l'auditeur) reconnaisse le figement, c'est-à-dire, reconstruise à partir d'une séquence défigée un phrasème d'origine (cf. Lecler 2006 ; García-Page 1989 : 61).

Ayant montré les mécanismes de défigement détectés dans les corpus, examinons, de manière panoramique, les exemples extraits des textes de poètes contemporains francophones, biélorusses et russes.

3.2. Défigement des collocations

Nous citons ici quelques exemples des corpus afin d'illustrer nos explications et commençons par les cas de défigement des collocations. Dans la TST, les collocations sont décrites à l'aide des *fonctions lexicales* (= FL)⁴. Cet outil nous aide à compléter et préciser la classification des séquences défigées. En outre, nous considérons nécessaire de spécifier les parties du discours avec lesquelles les collocations et les locutions exercent les mêmes fonctions syntaxiques ; nous empruntons, donc, le

4. Dans cet article, nous omettons la description du système des FL, ainsi que quelques d'autres questions théoriques. Compte tenu de l'importance de la base théorique pour la compréhension du matériel ici présenté, nous renvoyons aux articles de Mel'čuk (2003 et 2011) et de Апресян ; Дяченко ; Лазурский ; Цинман (2007).

classement de Gaston Gross (1996) selon les catégories grammaticales.

Dans ces deux vers de Jacques Brel :

La maison qui se tirebouchonne
Et qui **pleure à grosses planches**
Le gaz (Brel 1998 : 321)

nous constatons le défigement d'une FL adverbiale, **Magn** (*pleurer*) = à gros sanglots. Dans la collocation d'origine, le premier actant sémantique est toujours humain. Par contre, dans la phrase défigurée par Brel les sanglots sont remplacés par les planches, puisque c'est la maison qui « pleure ».

Le vent
Debout
S'assoit
Sur les tuiles du toit
Accalmie (Prévert 1963 : 84)

Ici, l'adverbe DEBOUT s'introduit, d'abord, comme une composante de la collocation non standard vent debout avec le sens 'vent qui souffle face à la proue du navire' ou 'vent qui contrarie la marche d'un moyen de transport'. Cependant, dans la troisième ligne, l'élément s'assoit occasionne une collision du sens, que l'adverbe DEBOUT a en tant que composante de la collocation, avec un autre signifié du même adverbe. Le défigement, donc, est dû au double sens produit par le contexte. C'est un exemple caractéristique de la *syllapse de sens*, un procédé par lequel un mot est employé à la fois au sens propre et figuré (Dupriez 1984 : 434).

Pour le biélorusse et le russe, mentionnons les exemples suivants de collocations défigurées :

...падзеліцца нішчымнай
Апошняя **луста**
Радасці людскай
Ларыса Геніюш (Барадупін 2006 : 203)
{... elle partagera
la dernière maigre **tranche**
de joie humaine}

Луста {TRANCHE} est une parcelle de pain. Ce lexème ne se rencontre en biélorusse normatif qu'à côté du lexème хлеб {PAIN} dans la collocation, décrite par une FL nominale **Sing** (*хлеб*)= *луста* [~ a]. Le concret étant d'habitude plus suggestif que l'abstrait, le poète biélorusse établit un rapport d'analogie entre l'action abstraite de partage des derniers instants de joie avec l'action tout à fait concrète de partage de la dernière tranche de pain.

Тревога жиреет и жиреет,
жрет зачерствевший разум
Я и Наполеон (Маяковский 1955)
{**L'angoisse s'engraisse et s'engraisse,**
dévore la cervelle rassise}

En russe, la collocation d'origine est *тревога раскѐм* {L'ANGOISSE *croît*}. Cette collocation est décrite par la FL verbale standard complexe **IncepPredPlus**. Le collocatif *croître* est remplacé par le verbe s'ENGRAISSER qui pourrait être un synonyme plus riche de ce premier (Syn>(*croître*) = *s'engraisser*), seulement si on fait référence à un animal ou un humain. Ainsi, le collocatif substitué « personnifié » sa base nominale ANGOISSE et la séquence défigurée gagne en expressivité.

3.3. Défigement des locutions

Les locutions défigées sont classifiées dans le corpus selon le niveau de contrainte, c'est-à-dire, selon leur appartenance à une des sous-classes des locutions dans le système de Mel'čuk. De la même manière que dans le cas des collocations, on distingue entre les *locutions nominales, verbales, adjectivales, adverbiales*, outre la distinction qu'on fait entre les locutions fortes, quasi- et semi-locutions.

Le premier exemple constitue le défigement de deux locutions fortes à la fois : la locution forte verbale «FAIRE UN TROU À LA LUNE» ce qui veut dire 's'enfuir sans payer ses dettes', combinée avec la locution forte nominale «LUNE DE MIEL», désignant la période heureuse suivant les noces :

On **fait** force de **trous dans** ma **lune de miel**
Le cocu (Brassens 1963 : 106)

Les deux locutions coïncident dans leur composante en commun – *lune*. Par cette intersection des locutions, Brassens transmet un sens tout particulier : le poète se réfère à l'ingratitude des amants de la femme envers le *je* lyrique, des amants qui volent les plaisirs de sa vie conjugale, en abusant, en plus, de son hospitalité.

Dans le vers suivant, Prévert manipule la quasi-locution adjectivale française «À VENIR», synonymique des adjectifs FUTUR et PROCHAIN. Le poète la défige par l'addition du verbe S'EN ALLER, l'antonyme du verbe venir dans l'acception 'se déplacer dans la direction d'un lieu ou d'une personne de référence' :

les siècles à **venir** ou à **s'en aller**
Eaux-forts (Prévert 1980 : 216)

Pour les langues slaves, nous proposons les exemples ci-dessous :

Свирепствовали декаденты
В поэзии, точно чума,
Дарили такие моменты,
Что люди **сбегали с ума** !
Поэза упадка (Северянин in Евтушенко dir. 1995)
≈ Les décadents sévissaient
Dans la poésie, comme la peste,
Ils offraient de tels moments
Que les gens s'enfuyaient de leur raison !}

Il s'agit de la transformation de la semi-locution verbale russe «сойти с ума» {dans la traduction très imparfaite ≈ «quitter sa raison»}. Un lecteur russophone, sans doute, y déchiffrera le défigement produit moyennant le remplacement d'une des composantes par son synonyme approximatif : en russe, сбегать 'descendre en courant' (composante verbale de la séquence défigée) est sémantiquement lié à une des acceptions du verbe сойти 'descendre en marchant' (composante verbale de la locution d'origine). Une fois de plus, le poète oppose un élément stylistiquement neutre du phrasème d'origine à un élément de l'énoncé défigé qui n'est pas pertinent dans ce contexte et qui est, pour cette raison, stylistiquement marqué.

En biélorusse, le poète détourne la quasi-locution adverbiale «ўніз галавой» {«tête en bas»} :

і павісну на правадах **уніз безгалоўем**

« Я лётаю неўтаймаваным вясёлым полтэргайстам... » (Жыбуль 2003 : 14)

{≈ je m'accrocherai aux fils, **mon absence-de-tête en bas**}

En biélorusse, aussi bien qu'en français, le mot галава {TÊTE}, outre le fait de désigner une partie du corps humain, a d'autres significés, comme qualifier la faculté intellectuelle, intelligence, esprit. Donc, l'adjectif безгаловы {SANS TÊTE}, dont l'acception la plus habituelle est 'stupide', 'irréfléchi' ou 'oublieux', provient du lexème en question dans cette dernière acception. Toutefois, dans le poème cité, comme il s'agit d'un fantôme (plus précisément d'un « poltergeist »), l'adjectif *безгаловы* est aussi perçu dans le sens littéral, 'privé de tête, de corps matériel'. L'auteur fait dériver de cet adjectif le substantif безгалоўе (qui pourrait être traduit comme « tête de linotte » ou, littéralement, *absence de tête*) et insère cet hapax dans la locution « ўніз галавой ». L'effet de défigement repose ici sur la modification individuelle du signifiant d'une composante (*галава* → *безгалоўе* {*tête* → *absence-de-tête*}) et sur le croisement de deux acceptions de cette composante.

3.4. Défigement des noms propres

Les noms propres complexes ont, dans le défigement, un comportement similaire à celui des phrasèmes. Constitués de deux éléments lexicaux ou plus, ils imitent, en quelque sorte, la forme et les fonctions syntaxiques des locutions. Notons aussi que nous incluons les titres des œuvres littéraires dans la classe des noms propres complexes.

Le titre d'un des poèmes de Raymond Queneau (2006 : 139) *La grenouille qui voulait se faire aussi **ronde** qu'un œuf* nous sert d'exemple d'un nom propre défigé. Cette phrase masque, par le remplacement de l'adjectif et d'un nom, le titre de la fameuse fable de La Fontaine *La Grenouille qui veut se faire aussi **grosse** que le **Bœuf*** (1668). L'esprit ingénieux de Queneau, membre de l'Oulipo, est prodigieux en ce genre de substitutions lexicales qui s'inscrivent dans le programme oulipien de manipulations textuelles.

Le poème biélorusse évoque l'une des ballades du romantisme anglais John Keats, *La Belle Dame sans Merci* (1819), dont le titre est détourné de la manière suivante :

Джакондаў ў гагарынскім абліччы
ўсьміхаецца **la belle dame sans sourcils**
Першае спатканьне (Хадановіч 2004 : 38)
{Telle la Joconde au visage de Gagarin,
la belle dame sans sourcils sourit}

Comme le titre du poème de Keats est écrit en français, le poète biélorusse a recours au français pour défiger cette expression à l'intérieur de la phrase biélorusse. Par une seule substitution paronymique (*merci* → *sourcils*), le personnage romantique féminin de Keats se voit réinterprété dans le poème postmoderniste où, comparé à la Joconde et au premier cosmonaute, il perd son auréole sublime. Chez Khadanovitx, c'est la mort elle-même que le poète baptise de ce surnom badin de *la belle dame sans sourcils*.

En russe, c'est le nom de la ville de Saint-Pétersbourg qui est transformé :

Но я сирень без памяти люблю,
тем более – в Санкт-белонощном граде
Посвящение (Ахмадулина 2011 : 307)
{≈ J'aime éperdument le lilas,
Surtout à **Saint-bourg de nuits blanches**}

La collocation non standard *белые ночи* {NUITS blanches} (employée normalement au pluriel) dénote un phénomène de la nature qui est observable traditionnellement à Saint-Pétersbourg : les nuits blanches sont des nuits, dans les régions du nord, pendant lesquelles se maintient une certaine clarté crépusculaire. Dans le poème, cette collocation non standard russe remplace l'élément *-Peter-* (le nom du saint et du tzar qui a fondé la ville) dans le nom propre Saint-Pétersbourg.

3.5. Défigement des unités phrastiques

La dernière section des corpus est occupée par les unités phrastiques défigées : proverbes ou dictons, citations et pragmatèmes. Ces unités constituent des phrases qui sont formellement et syntaxiquement rigides (cela concerne surtout les proverbes et les pragmatèmes). À part quelques composantes lexicales qui forment le noyau sémantique de la phrase, parfois aussi sa structure syntactique et formelle doit être gardée afin que l'énoncé défigé puisse être reconstruit.

Les deux premiers exemples, l'un en français et l'autre en biélorusse, regroupent les citations, ou les formules attribuées à certains auteurs :

je meurs d'ennui auprès de la fontaine

Les fontaines ne chantent plus (Queneau 2006 : 34)

Ад смагі паміраю у Экс-Крыніцы

« Ад смагі паміраю у « Экс-Крыніцы » (Хадановіч 2008 : 48)

{**Je meurs de soif près de l'ancien « Krynitsa »**}

Ces énoncés résultent tous les deux du défigement d'un vers extrait de la *Ballade des contradictions* (1458) de François Villon : *Je meurs de seuf auprès de la fontaine...* Les deux séquences conservent la totalité ou, dans le second cas, presque la totalité d'éléments lexicaux de la citation de Villon, mais aussi imitent sa syntaxe et sa métrique. L'effet humoristique, dans le premier exemple de défigement, est assuré par la commutation du sens du verbe mourir provoquée par le remplacement *soif* → *ennui* : la quasi-locution «MOURIR DE SOIF» est quand même plus pathétique que la semi-locution «MOURIR D'ENNUI», qui signifie tout simplement 's'ennuyer beaucoup'. Dans la séquence défigée par le poète biélorusse, l'humour consiste dans le fait que *Krynitsa*, qui veut dire en biélorusse 'source', 'fontaine', est ici le nom du bar, et le lexème SOIF n'est pas compris comme besoin de l'organisme en eau, mais plutôt comme l'envie de consommer une boisson alcoolique.

Cette phrase, extraite d'un poème russe :

Поэт человеку – волк,
отсюда – размеры глаз и ушей,
зубов, когтей и стихов
« Заставлять слушать свои стихи... » (Павлова 2009 : 124)

{**Le poète** est un loup pour l'homme,

Cela explique le grand gabarit de ses yeux et de ses oreilles,
aussi des dents, des griffes et des vers}

contient un cas de défigement de la version russe *Человек человеку волк* {*L'homme est un loup pour l'homme*} du proverbe latin *Homo homini lupus est*. Seulement une composante du proverbe étant substituée par un substantif inapproprié, le lecteur ou l'auditeur reconnaît sans difficulté le dicton originaire. En fait, ici on a affaire à deux mécanismes de défigement : le remplacement lexical et l'impact du contexte. Dans le proverbe d'origine, le loup n'est qu'un symbole de l'hostilité que les humains peuvent éprouver l'un envers l'autre ; dans la phrase défigée, à cause du contexte, le proverbe est pris au pied de la lettre, avec, en plus, une claire allusion au conte traditionnel retranscrit par Perrault, *Le Petit Chaperon rouge*.

Et pour en finir avec les exemples, nous allons citer un pragmatème détourné :

L'amour **est mort vive** la haine
Et toi matériel déclassé
Va-t-en donc accrocher ta peine
Au musée des amours ratées
La haine (Brel 1998 : 112)

Les deux premières lignes de cet extrait cachent le pragmatème *le Roi/la Reine est mort(e), vive le Roi/la Reine !*. La situation qui détermine cette phrase comme pragmatème est la nomination d'un nouveau monarque ou d'une nouvelle reine. Le poète, substituant dans le phrasème la composante *roi* par les substantifs abstraits AMOUR et HAINE, proclame la guerre à son ex-amante. Ainsi, le pragmatème modifié caractérise une nouvelle situation dans le système poétique de Brel.

Les unités phrastiques défigées à part, on rencontre aussi des cas de défigement des unités textuelles, mentionnées avant. À cause des spécificités de ce type d'unités, nous n'abordons pas ce problème dans cet article.

4. En guise de conclusion

Le défigement des phrasèmes dans les textes poétiques est une sorte d'épreuve linguistique et esthétique, un moyen stylistique dont les poètes se servent afin d'éviter les formules communes, indésirables en poésie, et de libérer les sens initiaux des mots en mettant en évidence le caractère contraint des phrasèmes.

Au cours de la première phase de recherche réalisée, nous avons conclu que les mécanismes de défigement des phrasèmes sont semblables dans les textes en français, en biélorusse et en russe. Le défigement est stylistiquement marqué et pourrait être reconnaissable par le récepteur du message indépendamment du degré de compositionnalité (collocation, locution forte, etc.) et du fonctionnement syntaxique (unité lexicale ou phrastique) du phrasème originaire.

Comme nos exemples le démontrent, l'effet de défigement se produit soit par l'activation d'un deuxième sens des composantes d'un phrasème, soit par l'ambiguïté de toute l'unité, soit par la création d'un nouveau sémantisme dans une séquence défigée.

Les objectifs que nous avons établis pour la prochaine phase de recherche peuvent être résumés de la manière suivante : 1) compléter les corpus avec les données en espagnol, catalan et polonais ; 2) créer une application pour la détection automatique de quelques structures défigées à l'aide du logiciel *NooJ* ; 3) définir les fonctions stylistiques du défigement dans l'œuvre des auteurs modernes.

Enfin, les descriptions qu'on a faites et qu'on prétend faire en perspective sont d'ordre lexicologique et phraséologique. Pourtant, nous espérons que les résultats de l'étude seront applicables aussi à la stylistique et la théorie littéraire.

Bibliographie

- BEN AMOR (Th.) 2007, *Le jeu de mots chez Raymond Queneau*, Tunisie, Faculté des Lettres et des Sciences humaines de Sousse.
- BEN AMOR (Th.) 2006, « Figement, défigement et jeux de mots formés sur énoncés proverbiaux », in *Composition syntaxique et figement lexical*, J. François et S. Mejri dirs, Presses universitaires de Caen, pp. 261-272.
- BLANCO (X.) 2012, « Le défigement des locutions nominales comme trait de style dans la poésie de Mario Benedetti », in *Les locutions nominales en langue générale*, X.r Blanco, S. Fuentes, S. Mejri eds, Bellaterra, Servei de Publicacions de la UAB, pp. 33-60.
- BLANCO (X.) à paraître, « Le défigement des collocations comme recours stylistique », in *Cahiers de Lexicologie. Revue internationale de lexicologie et de lexicographie*, Paris, Classiques Garnier.
- BLANCO (X.) et YAKUBOVICH (Y.) 2012, « Desfraseologización de funciones léxicas en poesía », in *Meanings, Texts and Other Exciting Things. A Festschrift to Commemorate the 80th Anniversary of Professor Igor Alexandrovič Mel'čuk*, Jurij Apresjan et al. eds, Moscow, Yasiki slavianskoy kulturi, pp. 67-74.
- CABASINO (Fr.) 1999, « Défigement et contraintes syntaxiques. Une analyse comparée des presses française et italienne », in *Cahiers de lexicologie. Revue internationale de lexicologie et de lexicographie*, Paris, 74/1, pp. 99-147.
- CATALÀ (D.) 2012, « Figement et défigement des proverbes comme outil didactique du FLE », *Paremia*, Madrid, 21, pp. 59-66.
- DUPRIEZ (B.) 1984, *Gradus : Les procédés littéraires (Dictionnaire)*, Paris, Union générale d'éditions.
- GARCÍA-PAGE (M.) 1989, « Sobre los procesos de deslexicalización en las expresiones fijas », in *Español Actual*, Madrid, 52, pp. 59-79.
- GRESILLON (Al.) et MAINGUENEAU (D.) 1984, « Polyphonie, proverbe et détournement, ou un proverbe peut en cacher un autre », in *Langages*, Paris, 73, pp. 112-125.
- GROSS (G.) 1996, *Les expressions figées en français : Noms composés et autres locutions*, Paris, Ophrys.

LECLER (A.) 2006, « Le défigement : un nouvel indicateur des marques du figement? », in *Cahiers de Praxématique : Changements linguistiques : figement, lexicalisation, grammaticalisation*, M. Lecolle et S. Leroy dirs, Montpellier, 46, pp. 43-60.

MEL'ČUK (I.) 2011, « Tout ce que nous voulions savoir sur les phrasèmes, mais... », in *Cahiers de lexicologie. Revue internationale de lexicologie et de lexicographie*, Paris : <http://olst.ling.umontreal.ca/pdf/MelcukPhrasemes2011.pdf>.

MEL'ČUK (I.) 2003, « Collocations dans les dictionnaires », in *Les écarts culturels dans les Dictionnaires bilingues*, Thomas Szende réd., Paris, Honoré Champion, pp. 19-64 : <http://olst.ling.umontreal.ca/pdf/Collocations-Szende.pdf>.

MEL'ČUK (I.) 1997, *Vers une linguistique Sens-Texte, Leçon inaugurale* (faite le Vendredi 10 janvier 1997), Collège de France, Chaire internationale : <http://olst.ling.umontreal.ca/pdf/melcukColldeFr.pdf>

MEJRI (S.) 2009, « Figement, défigement et traduction. Problématique théorique », in *Figement, défigement et traduction = Fijación, desautomatización y traducción*, S. Meiri et P. Mogorrón Huerta eds, Alicante, Universidad de Alicante, pp. 153-163.

MEJRI (S.) 1997, « Défigement et jeux de mots », in *Études linguistiques*, Tunis, 3, pp. 75-92.

PAJDZINSKA (A.) 1993, *Frazeologizmy jako tworzywo współczesnej poezji*, Lublin, AWH.

RASTIER (Fr.) 1997, « Défigements sémantiques en contexte », in *La locution, entre langues et usages*, M. Martins-Baltar éd., Paris, ENS Editions Fontenay / Saint Cloud, pp. 305-329.

SILBERZTEIN (M.) 2002, *The NooJ Resources* [ressource électronique] : <http://nooj4nlp.net/pages/download.html>.

SCHAPIRA (Ch.) 2000, « Proverbe, proverbialisation et déproverbialisation », in *Langages*, Paris, 139, pp. 81-97.

ŽOLKOVSKIJ (A.) et MEL'ČUK (I.) 1965, « O vozmožnom metode i instrumentax semantičeskogo sinteza (On a possible method and instruments for semantic synthesis) », *Naučno-tehničkaja informacija*, Moscou, 5, pp. 23-28.

АПРЕСЯН (Ю.) ; ДЯЧЕНКО (П.) ; ЛАЗУРСКИЙ (А.) ; ЦИНМАН (Л.) 2007, « О компьютерном учебнике лексики русского языка », in *Русский язык в научном освещении*, Moscou, 14/2, pp. 48-112.

Recueils de poésie cités

BRASSENS (G.) 1963, *Georges Brassens par Alphonse Bonnafé*, Paris, Éditions Seghers.

BREL (J.) 1998, *Tout Brel*, Paris, Éditions Robert Laffont, 3^e éd.

DÉCAUDIN (M.) éd. 1983, *Anthologie de la poésie française du XX^e siècle: de Paul Claudel à René Char*, Paris, Gallimard.

PRÉVERT (J.) 1980, *Soleil de nuit*, Paris, Gallimard.

PRÉVERT (J.) 1963, *Histoires*, Paris, Gallimard.

QUENEAU (R.) 2006, *Courir les rues. Battre la campagne. Fendre les flots*, Paris, Gallimard.

АХМАДУЛИНА (Б.) 2011, *Стихотворения*, Москва, Эксмо.

БАРАДУЛІН (Р.) 2006, *Руны Перуновы*, Мінск, Радыёла-плюс.

ЕВТУШЕНКО (Е.) dir. 1995, *Строфы века. Антология русской поэзии*, Минск-Москва, Полифакт.

ЖЫБУЛЬ (В.) 2012, *Станеліі*, Мінск, Логвінаў.

ЖЫБУЛЬ (В.) 2003 *Дыяфрагма*, Мінск, Логвінаў.

МАЯКОВСКИЙ (Вл.) 1955, *Полное собрание сочинений в 30 томах*, Том 1, Москва, ГИХЛ.

ПАВЛОВА (В.) 2009, *На том берегу речи*, Москва, АСТ.

ХАДАНОВІЧ (А.) 2008, *Бэрлібры*, Мінск, Логвінаў.

ХАДАНОВІЧ (А.) 2004, *Лісты з-пад коўдры*, Мінск, Логвінаў.

La terminologie de l'environnement en vulgarisation scientifique : la famille lexicale de la régénération des forêts en portugais

Mariana Giacomini BOTTA
*Observatoire de linguistique
Sens-Texte (OLST)
Université de Montréal
Universidade de São Paulo*

Résumé

Cet article présente les résultats d'une étude sur les « mouvements migratoires » des termes entre les langues de spécialité (LSP) et la langue courante, et vice-versa. Nous concentrons notre attention sur les textes de vulgarisation publiés par la presse, en raison du fait que les journalistes scientifiques ne sont pas des spécialistes et sont susceptibles, pour cette raison, de comprendre et d'utiliser les termes différemment. On peut émettre l'hypothèse selon laquelle les termes utilisés dans les textes de vulgarisation n'ont pas le même statut que les termes employés dans les textes scientifiques. L'usage des termes par des non-spécialistes conduit normalement à des changements de sens, parce que, dans ce mouvement entre la LSP et la langue courante, les unités lexicales peuvent perdre certaines des caractéristiques qu'elles possèdent en langue de spécialité. Pour analyser ces mouvements, nous avons choisi de travailler sur des textes portant sur l'environnement et publiés dans la presse nationale brésilienne. Les données sont extraites de deux corpus compilés pour cette étude. Nous montrons que la motivation des termes (liée à leur introduction dans la langue de spécialité) peut être l'un des facteurs de leur « déterminologisation ».

Mots-clés :

terminologie, environnement, vulgarisation, déterminologisation, portugais, presse.

1. Introduction et objectifs

La présente étude porte sur les changements de sens auxquels les unités lexicales spécialisées sont soumises lorsqu'elles sont employées dans les textes de vulgarisation publiés dans la presse. C'est l'utilisation des termes des langues spécialisés par des journalistes qui a motivé cette étude, car les cas de mécontentement de scientifiques relatif à leurs articles ne sont pas rares¹.

Hernando (2002 : 3) rend compte que l'insatisfaction à l'égard du travail des journalistes scientifiques a déjà fait l'objet d'une étude de l'École de journalisme et des médias de l'Université du Minnesota. Cette recherche, dirigée par Tichenor (1970), a montré que la raison de ce mécontentement réside principalement dans le manque de précision de l'information publiée. Tichenor *et al.* (1970 : 677-678) ont recueilli les opinions des scientifiques sur les articles de presse portant sur différents sujets et ils ont montré que leur plus grande préoccupation est l'interprétation des données elles-mêmes, notamment les erreurs relatives à des concepts scientifiques.

Pourtant, dans le domaine du journalisme scientifique, on accorde assez d'importance à la question de la « traduction » de la langue de spécialité (LSP) vers la langue courante, de façon que le public non initié puisse comprendre la première. Le journaliste joue le rôle de médiateur entre le scientifique et le grand public mais, puisqu'il

n'est pas spécialiste, il comprend et utilise les termes différemment des experts du domaine.

En ce qui concerne le système lexical de langue, on sait que les terminologies varient selon *les acteurs, les échanges, les supports, les lieux où elles circulent et les pratiques qui les motivent, en bref, selon leurs conditions de production, de circulation et de réception* (Delavigne 2012 : 244). On peut, avec Dury (2008 : 231), considérer les termes comme des *entités « mobiles » qui peuvent migrer d'un domaine à l'autre ou se déplacer vers la langue générale.*

Dury affirme que la circulation langagière des termes s'accompagne presque toujours d'une modification de sens. Le sens du terme peut être soumis à un « resserrement » (il est adopté avec une partie seulement de son sens) ou, selon le cas, à un « élargissement » (à son sens premier sont ajoutées de nouvelles composantes sémantiques, propres à la discipline). Dans ce mouvement entre la LSP et la langue courante, le terme peut perdre certaines des caractéristiques qu'il possède en langue de spécialité, mais *quelle que soit l'ampleur de la circulation d'un terme, et le nombre de transgressions sémantiques qu'il subit, il subsiste toujours un sens noyau, qui indique son origine* (Dury 1999 : 22).

Pour rendre compte de ces changements de sens, nous proposons d'étudier les « mouvements migratoires » des termes entre la langue de spécialité (LSP) et la langue courante. Notre objectif est de décrire les « mécanismes de migration » de termes mis en œuvre par la presse et d'essayer de répondre aux questions suivantes :

1. Cette question a été récemment abordée par le journaliste brésilien André Trigueiro, expert et professeur de journalisme environnemental, dans l'article « NYT fecha editoria de meio ambiente.E daí? », Disponible sur : <http://g1.globo.com/platb/mundo-sustentavel/2013/01/22/nyt-fecha-editoria-de-meio-ambiente-e-dai/>. Consulté le 10 fév. 2013.

- a) Que font les journalistes (quels sont les mécanismes) pour rendre les termes « plus clairs » au public non initié ?
- b) Y a-t-il des raisons linguistiques ou discursives qui peuvent justifier ces changements de sens ?
- c) Les expressions linguistiques utilisées dans la presse correspondent-elles aux termes employés par les scientifiques ?

On peut illustrer et mieux détailler ces questions en prenant comme exemple un texte de la presse brésilienne portant sur l'environnement. À partir de l'extrait reproduit ci-dessous, nous pouvons formuler plusieurs remarques à propos de l'usage des unités lexicales *recomposição* (restauration) et *reflorestamento* (reboisement) :

[...] o percentual de **recomposição** estabelecido é chamado de “escadinha”, já que quanto maior a propriedade, maior o **reflorestamento** exigido [...]. (OGLOBO_08/09/2012).

[...] le pourcentage de **restauration** établi [par la loi] est appelé « petit escalier », puisque le **reboisement** obligatoire sera proportionnel à la taille de la propriété [...] (notre traduction).

Il est connu que *reflorestamento* est un terme de l'environnement, mais nous doutons que la presse l'utilise avec le sens précis qu'il véhicule dans le domaine. S'il y a une variation diaphasique (situationnelle), faut-il la décrire dans les dictionnaires et banques de données terminologiques ?

Dans le même exemple, il n'est pas possible de déterminer si *recomposição* et *reflorestamento* sont employées pour désigner un même concept ou s'ils correspondent à des termes distincts. On remarque

aussi que les journalistes emploient d'autres unités lexicales, telles que *regeneração* et *recuperação*, dans des contextes semblables. Sont-elles aussi des termes ? Est-ce que la presse fait appel à d'autres termes pour expliquer les concepts du domaine ou a-t-elle davantage recours à des unités lexicales de la langue courante ?

Nous proposons des éléments de réponse à toutes ces questions à partir d'une analyse comparative des données issues de deux corpus du portugais brésilien, compilés pour cette étude. Le premier est composé de textes de la presse nationale, extraits des rubriques *Natureza* (nature), *Planeta* (planète) et *Meio ambiente* (environnement), et le deuxième rassemble des textes scientifiques (thèses et articles) portant sur la régénération des forêts.

2. Linguistique, terminologie et vulgarisation

Le discours de vulgarisation, aussi appelé, selon l'auteur, *discours d'information scientifique et technique* ou *discours de transmission de connaissances*, est l'une des modalités de diffusion des connaissances. La vulgarisation vise à répandre des idées et des connaissances dans le grand public, il s'agit d'une diffusion hors d'une communauté auprès des non-spécialistes. Selon Delavigne (2001 : 34), [...] *tout à la fois information sur le monde scientifique et transmission de connaissances spécifiques – et non de savoirs –, la vulgarisation rend tout simplement possible un regard sur la science et permet la construction d'une culture.*

D'après cette auteure, les pratiques vulgarisatrices combinent une dimension sociale, la connaissance et la langue. La vulgarisation est donc une affaire de mots, parce qu'elle permet, par le biais de la langue, de construire un type de connaissance. Selon elle, la vulgarisation peut passer par des expositions et conférences, mais *la presse apparaît comme le média prototypique de la vulgarisation* (Delavigne 2001 : 26).

Sujet de quelques études linguistiques depuis les années 1980, la vulgarisation est fréquemment assimilée à une traduction et son analyse *se focalise sur les divers procédés de 'traduction' de ces terminologies absconses* (Mortureux 1987 : 9). Pour cette auteure, qui a travaillé sur le fonctionnement des discours de vulgarisation, principalement sur la problématique de la paraphrase et de la traduction, ces discours sont appréhendés comme une « réénonciation paraphrastique », dans un registre non savant, du discours spécialisé. Delavigne (2012 : 233) affirme que *les énoncés de vulgarisation gagnent à être envisagés non comme des traductions, mais comme des productions singulières*, ce que nous proposons dans cette étude.

En ce qui concerne le changement de sens des termes dans la migration vers la langue courante, Dury (1999 : 26) cite Wierzbicka (1985) à propos de l'existence d'un « concept minimal » et d'un « concept maximal » du terme : *le spécialiste détiendrait en toute logique le concept maximal d'un terme [...], et le non-spécialiste, par contre, serait celui qui accéderait au concept minimal, minimum vital, mais suffisant pour communiquer*.

Dans ses recherches sur ce sujet, Dury (2008 : 230) s'appuie également sur les travaux de Meyer et Mackintosh (2000a ; 2000b) qui explorent le processus linguistique qu'elles appellent *déterminologisation*, qui sera également utile pour notre analyse. Ce phénomène se produit quand un terme d'un domaine est utilisé par des non-spécialistes et subit des changements sémantiques, grammaticaux et pragmatiques. Selon les auteurs, il y a déterminologisation lorsque l'utilisation du terme par les non-spécialistes contribue à la dilution de son sens terminologique original.

Ainsi, on peut émettre l'hypothèse selon laquelle les termes utilisés dans les textes de vulgarisation n'ont pas le même statut que les termes employés dans les textes scientifiques. Dans le mouvement entre la LSP et la langue courante, le terme peut perdre certaines des caractéristiques qu'il possède en langue de spécialité. Il subit, alors, cette tension appelée *déterminologisation*.

3. Méthodologie

Les analyses présentées ici sont basées sur la méthode terminologique développée par L'Homme (2004 ; 2008), qui commence par l'élaboration des corpus. Nous avons constitué un premier corpus de textes écrits de la presse nationale brésilienne, qui compte 312 articles et 139 943 occurrences. Nous avons recueilli des textes disponibles en ligne dans les sites web des trois principaux journaux (les plus vendus et plus influents) de la presse nationale brésilienne, *Folha*

de *S. Paulo, O Estado de S. Paulo et O Globo*, entre le 1^{er} août et le 30 septembre 2012.

Nous avons aussi créé un corpus de textes spécialisés en portugais, qui regroupe 50 textes scientifiques portant sur la régénération des forêts. Ces textes ont été publiés au Brésil entre 1981 et 2012 et sont disponibles en ligne. Ce dernier corpus compte 136 910 occur-

**TABLEAU 1. CARACTÉRISTIQUES DES DEUX CORPUS
CONSTITUÉS POUR CETTE RECHERCHE**

	NOMBRE D'OCCURRENCES	NOMBRE D'ARTICLES	PÉRIODE
CORPUS PRESSE	139 943	312	1/08/2012 – 30/09/2012
CORPUS SCIENTIFIQUE	136 910	50	1981 – 2012

rences. La constitution de ces deux corpus a été nécessaire pour que nous puissions analyser les mouvements et changements de sens entre les différents discours.

La deuxième étape de l'analyse était la sélection des termes. Selon L'Homme (2004 : 64-66), un terme est une unité lexicale qui a un sens lié à un domaine de spécialité et *le statut terminologique d'une unité lexicale se définit en fonction du lien qu'on peut établir entre son sens et un domaine de spécialité* (L'Homme 2004 : 64). Ce lien est plus facile à établir pour des unités dénotant des objets concrets.

A fin de confirmer le caractère spécialisé d'une unité lexicale, particulièrement des unités prédicatives, l'auteure propose autres trois

critères « lexico-sémantiques » : a) la vérification de la nature des actants sémantiques (pour les unités à sens prédicatif) ; b) la parenté morphologique (accompagnée d'une parenté sémantique) ; et c) des relations paradigmatiques.

Pour la sélection des termes, nous avons traité nos deux corpus séparément au moyen de l'extracteur de termes *TermoStat*, développé par Drouin (2003). À partir des listes de candidats termes produites par l'extracteur, nous avons retenu les unités lexicales appartenant au champ sémantique de la régénération des forêts. Ce choix a été motivé par le fait que *recomposição* apparaît parmi les 20 premiers candidats termes du corpus de textes de la presse. Il s'agit d'une unité qui, à première vue, ne semble pas être un terme de l'environnement, parce qu'elle a un sens très général dans la langue courante. Dans cette variété de discours, elle a le même sens que le mot français *recomposition* : « action de recomposer ; composer (ce qui est décomposé, défait) ; réunir les éléments de » (*Le Nouveau Petit Robert* 2013).

Dans les résultats du traitement du corpus de textes scientifiques par le même logiciel, *recomposição* est le neuvième parmi les candidats termes, mais nous trouvons aussi dans ce groupe les unités *regeneração, recuperação, reflorestamento* et *manejo*, qui appartiennent également au champ sémantique de la régénération des forêts. Dans cet article, nous analyserons seulement les quatre premières unités susmentionnées de cette famille.

Dans le tableau ci-dessous, nous présentons les principaux candidats termes de chaque corpus, extraits de la liste générée par *TermoStat* :

**TABEAU 2. LES PRINCIPAUX CANDIDATS TERME
DU CORPUS PRESSE ET DU CORPUS SCIENTIFIQUE**

CORPUS PRESSE			CORPUS SCIENTIFIQUE		
CANDIDAT TERME	FRÉQUENCE	SPÉCIFICITÉ	CANDIDAT TERME	FRÉQUENCE	SPÉCIFICITÉ
código florestal	265	126.93	espécie	1007	186.57
meio ambiente	173	102.45	floresta	599	178.23
desmatamento	161	98.5	mata	311	124.04
gelo	226	97.49	plantio	205	122.47
florestal	148	94.71	manejo	198	120.86
mudança climática	122	84.16	semente	156	94.82
ártico	92	74.52	regeneração	135	92.69
floresta	209	73.87	recomposição	130	90.35
derretimento	80	68.12	vegetação	129	84.76
aquecimento	110	59.84	florestal	86	80.82
ambiental	55	57.4	muda	93	79
efeito estufa	54	56.87	reflorestamento	72	73.34
carbono	95	56.6	ecossistema	66	56.45
recomposição	73	56.29	árvore	146	55.5
monitoramento	48	53.55	preservação	91	51.76

Les étapes suivantes de la méthode ont été : a) la collecte de contextes où ces unités sont utilisées ; b) l'examen des microcontextes, afin d'isoler les actants liés à chaque terme (au moyen du concordancier *Intercorpus*²) ; c) la prise en compte de distinctions sémantiques

2. *Intercorpus* est un concordancier mis à la disposition des chercheurs de l'OLST (Observatoire de linguistique Sens-Texte), de l'Université de Montréal.

(entre plusieurs acceptions et usages d'un terme) et des liens lexicaux des termes présents dans chaque corpus. Enfin, nous avons comparé les résultats obtenus de l'analyse de chaque unité dans le corpus de presse et dans celui de la LSP.

4. L'analyse des données

4.1. L'usage de *recomposição*

Dans le corpus de la LSP nous avons trouvé 141 occurrences du terme *recomposição*, qui rend compte du rétablissement complet d'un écosystème dégradé (les eaux des rivières, les animaux, la subsistance de la population) et pas seulement la replantation d'arbres. Il désigne aussi la tentative de restauration de l'aspect original d'une région dont l'une des étapes est la replantation d'espèces végétales natives.

Recomposição est un exemple de terme qui a été introduit dans la LSP par l'appropriation d'un mot de la langue courante : c'est la nominalisation du verbe *recompôr*. La création de ce terme résulte d'une néologie sémantique par modification du sens d'un lexème existant, c'est-à-dire, la spécialisation d'un mot courant. Il s'agit, selon Mortureux (1985 : 836), d'une homonymie fondée sur une « extension » de sens, un processus très productif dans les discours scientifiques. Selon cette auteure, il s'agit d'un homonyme de lexème courant, car sa motivation (la relation perceptible entre sa forme et son sens) est d'ordre sémantique (Mortureux 1995 : 18-19).

Dans la LSP, l'unité lexicale *recomposição* reçoit des traits spécifiques du domaine et désigne le processus de restauration de l'aspect primitif d'une région dégradée (avec la replantation d'espèces végétales natives, la dépollution des eaux, la protection des animaux et l'établissement de moyens de subsistance de la population). Parmi les principales réalisations d'actants de ce terme dans les textes spécialisés, nous avons trouvé : *de floresta/florestal, de mata, da paisagem, da vegetação, da flora, da fauna, artificial et espontânea*. Voici ci-dessous un aperçu des résultats du traitement du corpus Scientifique dans le concordancier *Intercorpus*, utilisé pour analyser les microcontextes et sélectionner des actants liés à cette unité lexicale :

seguinte Lei: Art. 1º É obrigatória a recomposição florestal, pelos proprietários, nas áreas : contribuir, de forma técnica, para a recomposição de matas ciliares não só na região oeste d ao da área a ser recuperada; • Para a recomposição, as árvores mais indicadas para a sua regi i do projétil alojado na coxa e para a recomposição da asa direita. O gavião real está na list itores bióticos, pode-se partir para a recomposição da área, que pode consistir dos seguintes resultado das ações executadas para a recomposição e revegetação dos barrancos que apresentam manescente; 2. Criar condições para a recomposição espontânea da mata; 3. Evitar a poluição a segue-se também criar condições para a recomposição espontânea da mata, bem como fazer a recom o equilíbrio dos ecossistemas. Para a recomposição da mata nativa devem ser usadas somente es into sustentável para a região. Para a recomposição completa de sua mata ciliar, estima-se que .va é de que o Mais Ambiente promova a recomposição de 23 milhões de hectares de matas ciliare ío ribeirinha que reside no local, e a recomposição de mata ciliar às margens do Urucum Mas um njunto de leis visando a proteção e a recomposição das florestas nativas brasileiras, conform no. 7381 de 19/03/2010 Dispõe sobre a recomposição da cobertura vegetal, das matas ciliares n . de abastecimento de água, urge que a recomposição da vegetação nativa seja prioridade. Final : no exposto acima, considera-se que a recomposição de áreas ciliares degradadas em locais que seus recursos naturais”, que envolve a recomposição da mata ciliar do Córrego do Jardim Bandei isão CO2, seqüestra Carbono, promove a recomposição da flora e da fauna, permite projetos agro

FIGURE 1 : EXEMPLE DE RÉSULTATS
DU CONCORDANCIER INTERCORPUS – CORPUS SCIENTIFIQUE

Dans le corpus de la presse, il y a 72 occurrences de cette unité lexicale. Elle désigne la restauration d'une zone qui a été illégalement

ou irrégulièrement déboisée et elle est utilisée comme quasi-synonyme de *reflorestamento* (reboisement). Dans le discours de vulgarisation, le focus de ce concept est la replantation d'arbres, natifs ou fruitiers. Les principaux actants de ce terme dans les textes de la presse sont : *de áreas, de matas et de hectares*. La figure 2 montre une partie des résultats obtenus par le traitement du corpus Presse par l'*Intercorpus* :

de água de até 10 metros de largura, a recomposição de mata ciliar será de 15 metros. A temente da largura do curso de água, a recomposição ficará entre 20 a 100 metros de APP ntida, para as propriedades maiores, a recomposição máxima de 100 metros de mata. No en e médias, de 4 a 15 módulos fiscais, a recomposição de áreas desmatadas será de 15 metr s médias, de 4 a 15 módulos fiscais, a recomposição de áreas desmatadas será de 15 metr derá efeito o artigo que trata sobre a recomposição das APPs. CPI do Cachoeira O presid lidde de os estados legislarem sobre a recomposição de áreas desmatadas em vez da União lidde de os estados legislarem sobre a recomposição de áreas desmatadas em vez da União o e eventuais divergências se refere a recomposição daquilo que foi desmatado ilegalmen ntares aprovaram emenda determinando a recomposição obrigatória mínima de 15 metros de e lei do Código Florestal não trata da recomposição de mata ciliar dos rios urbanos, al e lei do Código Florestal não trata da recomposição de mata ciliar dos rios urbanos, al e rejeitado pelo governo, trata-se da recomposição de áreas de preservação permanentes ncada ruralista, que exigia redução da recomposição nas margens de rios. No final, o go exigência aos médios produtores de uma recomposição maior de áreas desmatadas, conforme

FIGURE 2 : EXEMPLE DE RÉSULTATS
DU CONCORDANCIER INTERCORPUS – CORPUS PRESSE

La vérification des actants sémantiques est l'un des critères « lexico-sémantiques » proposés par l'Homme (2004) pour confirmer le caractère spécialisé d'une unité lexicale, mais il nous semble qu'il est aussi important pour la comparaison de l'usage des unités dans des discours distincts. Dans le cas de *recomposição*, la présence de l'actant *da fauna* dans la LSP signale son sens plus large, appliqué à tous les organismes animaux et végétaux d'un écosystème. Grâce à

l'étude des actants, nous nous rendons compte que, dans le discours de vulgarisation, cette unité est utilisée dans un sens plus restreint (la faune/les animaux ne figurent pas parmi les actants).

Elle a aussi un sens « juridique », lié au contexte du droit de l'environnement et aux exigences de la loi. C'est l'actant *hectares*, qui désigne une mesure de superficie, qui indique cet usage différent. Puisque cette dernière acception n'apparaît pas dans les textes scientifiques sur la régénération des forêts, nous pouvons conclure qu'il s'agit de deux termes, de deux domaines différents, mais liés entre eux : l'environnement et la gestion de forêts ou le droit environnemental³.

Dans notre corpus de textes de la presse, nous trouvons des usages de cette unité lexicale liés à ces deux sens. Dans le discours journalistique, il semble que cette unité lexicale ne soit pas reconnue en tant que terme de l'environnement et les non-spécialistes perçoivent leur sens terminologique de manière moins précise. En raison de ce fait, elle est employée avec un sens moins spécifique et subit une tension appelée *déterminologisation* (Meyer et Mackintosh 2000a ; 2000b). Selon ces auteurs, il s'agit de la dilution du sens terminologique causé par l'utilisation du terme par les non-spécialistes.

Mais nous n'observons pas dans le discours de vulgarisation le même comportement lorsque cette unité est employée dans l'acception

3. *Recomposição* semble correspondre en français au terme *restauration*, qui est décrit dans le domaine « écologie » du site *Termium* : « opération visant à rendre un écosystème dégradé ou une population dégradée à son état d'origine, à redonner à un lieu son aspect initial ». Dans le même site, il est possible de trouver un autre sens, très proche, dans le domaine de la « gestion environnementale/sécurité » : « [La restauration d'un lieu] consiste à remettre un lieu dans son état original ou dans un état conforme aux exigences de la loi ou d'un contrat. »

liée à la gestion de forêts ou au droit environnemental. Dans ce cas, elle semble maintenir le sens de la LSP.

4.2. L'usage de *regeneração*

Dans le corpus de la LSP nous avons trouvé 161 occurrences de ce terme, qui désigne l'opération dont le but est de redonner à un lieu ses caractéristiques d'origine (qui ont été perdues). Dans les contextes du corpus, il est toujours lié aux idées de restauration et de renouvellement d'un écosystème dégradé.

Dans le corpus de textes de la presse, nous avons trouvé une seule occurrence de cette unité, dans la forme *regeneração*. Dans ce contexte, l'unité lexicale désigne aussi le processus de restauration de l'aspect primitif d'une région dégradée :

Esses dados poderão mostrar, por exemplo, as características da vegetação em áreas que foram desmatadas no passado, mas que agora encontram-se em processo de regeneração. (OESP_06/08/2012).

Ces données peuvent montrer, par exemple, les caractéristiques de la végétation des zones qui ont été déboisées dans le passé, mais qui sont maintenant dans un processus de régénération. (*notre traduction*).

Les dictionnaires du portugais décrivent cette unité en citant comme synonymes « reconstitution, restauration » (Houaiss 2009, *notre traduction*) et « réorganisation » (Aulete 2008, *notre traduction*). Dans ces deux ouvrages lexicographiques, on trouve la description de

plusieurs sens spécialisés, dans des domaines tels que la biologie, l'électronique, l'informatique, la physique nucléaire, la chimie et la religion, mais il n'y a aucune mention de l'acception liée à l'environnement.

Ces données nous montrent que, dans la langue courante ainsi que dans la presse, cette unité n'est pas perçue en tant que terme. Sa faible fréquence montre qu'elle n'est pas encore populaire et pour cette raison, il n'est pas possible d'identifier un processus déterminologisation. Elle semble être utilisée dans le discours de vulgarisation avec le même sens que celui véhiculé par la LSP.

Cependant, cette unité lexicale a une fréquence importante dans le domaine de la régénération des forêts, où elle a comme actants principaux d'autres termes de l'environnement (*florestal/de floresta, natural, espontânea, artificial, de espécies, da vegetação*). Nous pouvons voir une certaine coïncidence d'actants avec le terme *recomposição*, à l'exception de *da paisagem* et *da fauna*, ce qui montre que ces deux termes ne peuvent être employés comme variantes que dans certains contextes. La préférence de la presse pour l'utilisation de *recomposição* pourrait s'expliquer par la possibilité combinatoire plus large de ce terme dans le domaine ou par l'accent au caractère « réglementaire », obligatoire de cette action.

4.3. L'usage de *reflorestamento*

Dans notre corpus de la LSP, ce terme a 83 occurrences et il peut désigner soit la replantation d'**espèces végétales non natives**, soit une méthode spécifique de replantation. C'est un terme qui a été

créé pour désigner un concept de l'environnement, l'action de refaire une forêt. Nous voyons qu'il a une motivation morphologique, c'est-à-dire que sa compréhension repose sur la connaissance du sens de ses éléments.

D'après Mortureux (1995 : 19), les considérations [...] sur la motivation des termes ont paru justifiées dans la mesure où elles éclairent les mécanismes de leur interprétation dans la communication. Cette affirmation nous amène à supposer que la motivation des termes peut avoir un rapport avec leur déterminologisation dans la vulgarisation.

Dans la presse, l'unité lexicale *reflorestamento* a 11 occurrences et désigne n'importe quel type de processus artificiel de replantation d'arbres dans des zones dévastées. Les dictionnaires du portugais décrivent *reflorestamento* comme l'action de *reflorestar*, « restaurer la couverture forestière (d'une zone déboisée) en plantant des semenciers ou des gaules d'arbre » (Houaiss 2009, *notre traduction*). Le type d'arbre (natifs, fruitiers) n'est pas mentionné. Ce trait ne semble pas être aussi important dans la langue courante que dans la presse. Nous vérifions la perte de traits spécifiques et concluons qu'il y a là déterminologisation.

4.4. L'usage de *recuperação*

Dans nos deux corpus, l'unité lexicale *recuperar* a comme sens général : promouvoir la restauration de, le rétablissement de, reconstituer. Il est important de signaler que, dans la langue courante, ce

mot n'a pas exactement le même sens que *recupérer* en français ; en portugais, il y a deux verbes distincts, dont les dictionnaires indiquent qu'ils ont la même origine : *recuperar* et *recobrar*. C'est ce dernier, correspond au sens de récupérer en français.

Dans le corpus spécialisé, on retrouve 134 occurrences de cette unité lexicale et, dans de corpus de la presse, on n'en relève que 37 occurrences. Tant dans le premier corpus que dans le deuxième, cette unité est utilisée dans le même sens que dans la langue courante, pour remplacer soit *recomposição*, soit *reflorestamento*, dans des contextes où les traits spécifiques de ces termes ne sont pas nécessaires. Elle n'est pas « terminologisée », ne reçoit pas de traits spécifiques, mais elle a comme actants, dans les deux discours, d'autres termes de l'environnement, tels que : *de matas, de florestas, de áreas degradadas, de ecossistema, de recursos, de bacias hidrográficas, de rios et natural*.

5. Conclusions et considérations finales

Nous avons mené une étude comparative sur l'usage des unités lexicales *recomposição, regeneração, reflorestamento* et *recuperação* dans le domaine spécialisé de l'environnement et dans les textes de la presse nationale brésilienne. Nous avons pu confirmer que les termes se comportent différemment dans les textes de vulgarisation, parce que les journalistes ne sont pas des spécialistes et qu'ils perçoivent le sens terminologique de manière moins précise. Ils emploient les termes avec un sens moins spécifique que les experts, et cela dépend de l'interprétation qu'ils font du sens du terme.

Dans les deux variétés de discours étudiées, les quatre unités lexicales analysées présentent des comportements distincts: a) *reflorestamento* est le seul terme de cette famille qui a été créé pour désigner un concept de l'environnement, sa motivation est morphologique et, dans la presse, il subit la tension appelée *déterminologisation* et perd des traits spécifiques; b) *recomposição* et *regeneração* sont des termes qui ont été créés à partir de mots existants dans la langue courante, auxquels se sont ajoutés des traits spécifiques dans la LSP (« terminologisation »), mais *recomposição* est fréquent dans la presse et parfois employé comme synonyme de *reflorestamento* (déterminologisation), contrairement à *regeneração*, qui n'est pas fréquent dans les textes de la presse, où il conserve ses traits spécifiques (il n'y a pas de déterminologisation) ; c) *recuperação* est utilisé dans le domaine spécialisé dans le même sens que dans la langue courante, il n'y a pas d'ajout ni perte de traits spécifiques.

La double description de la structure lexicale des unités (telles qu'elles sont utilisées par les spécialistes et par les journalistes) nous a permis de déceler les différences de sens selon leur emploi dans chaque discours. La comparaison du fonctionnement dans la LSP et dans le langage de vulgarisation a rendu possible l'observation des « mouvements migratoires » des mots et aide à mieux comprendre les mécanismes de terminologisation et de déterminologisation. Ce travail a montré également que l'origine du terme (la motivation), ainsi que ses propriétés morpho-sémantiques, peuvent faire partie des facteurs de déterminologisation.

Remerciements

La recherche décrite dans cet article est financée par le programme de bourses de recherche postdoctorale du gouvernement du Canada - MAECI (Ministère Affaires étrangères et Commerce international Canada). Nous tenons à remercier également Marie-Claude L'Homme et les relecteurs anonymes qui ont parcouru ce texte.

Bibliographie

- DELAVIGNE (V.) 1995, « Approche socioterminologique des discours du nucléaire », in *Meta : journal des traducteurs*, Montréal, 40/2, pp. 308-319.
- DELAVIGNE (V.) 2001, *Les mots du nucléaire : contribution socioterminologique à une analyse des discours de vulgarisation*. Thèse de doctorat en linguistique de l'Université de Rouen.
- DELAVIGNE (V.) 2012, « Peut-on “traduire” les mots des experts ? Un dictionnaire pour les patients atteints de cancer », in *Dictionnaires et Traduction*, M. Heinz éd., Berlin, pp. 233-263.
- DICIONARIO ELETRÔNICO HOUAISS da língua portuguesa, 2009, São Paulo, Editora Objetiva.
- DROUIN (P.) 2003, « Term Extraction Using Non-technical Corpora as a Point of Leverag », in *Terminology*, 9/1, Amsterdam, pp. 99-115.
- DURY (P.) 2008, “The rise of carbon neutral and compensation carbone : A diachronic investigation into the migration of vocabulary from the language of ecology to newspaper language and vice versa”, in *Terminology*, 14:2, Amsterdam, pp. 230-248.
- DURY (P.) 1999, « Les variations sémantiques en terminologie : étude diachronique et comparative appliquée à l'écologie », in *Sémantique des termes spécialisés*, V. Delavigne et M. Bouveret éd., Rouen, Publications de l'Université de Rouen, pp. 17-32.

HERNANDO (M. C.) 2002, “La divulgación científica y los desafíos del nuevo siglo”, in *Primer Congreso Internacional de Divulgación Científica en la USP* (Conferencia). Disponible sur: <http://www.museudavida.fiocruz.br/media/La%20Divulgacion%20Cientifica%20y%20los%20Desafios%20del%20Nuevo%20Siglo%20-%20Manuel%20Calvo%20Hernando.pdf>. Consulté le 5 fév. 2013.

iDicionário AULETE, 2008 (version web). Disponible sur : http://aulete.uol.com.br/site.php?mdl=aulete_digital. Consulté le 20 jan. 2013.

L'HOMME (M.-Cl.) 2008, « Le DiCoInfo. Méthodologie pour une nouvelle génération de dictionnaires spécialisés », in *Traduire*, Paris, 217, pp. 78-103.

L'HOMME (M.-Cl.) 2004, *La terminologie : principes et techniques*, Montréal, Presses de l'Université de Montréal.

LE GRAND ROBERT de la langue française, 2001 (version numérique), Paris, Dictionnaires Le Robert.

LE PETIT ROBERT de la langue française, 2013 (version numérique), Paris, Dictionnaires Le Robert.

MEYER (I.) ; MACKINTOSH (K.) 2000a, « L'étirement du sens en terminologie : aperçu du phénomène de la déterminologisation », in *Le sens en terminologie*, H. Béjoint et Ph. Thoiron eds, Lyon, Presses universitaires de Lyon, pp. 198-217.

MEYER (I.) ; MACKINTOSH (K.) 2000b, « When terms move into our everyday lives: An overview of de-terminologization », in *Terminology*, 6/1, Amsterdam, pp. 111-118.

MORTUREUX (M.-Fr.) 1995, « Les vocabulaires scientifiques et techniques », in *Les Carnets du Cediscor*, Paris, 3, Presses Sorbonne nouvelle, pp. 13-25.

MORTUREUX (M.-Fr.) 1994, « Traitement des termes et “simplification” dans le discours de vulgarisation scientifique », in *Cahiers du français contemporain : Simple – Simplification*, Lyon, 1, ENS Éditions, pp. 213-225.

MORTUREUX (M.-Fr.) 1987, « Traduction et vulgarisation scientifique : un transfert de problématique ? », in *DISCOSS* (Discours contrastif, Sciences et sociétés), Paris, 3, pp. 7-21.

MORTUREUX (M.-Fr.) 1985, « Linguistique et vulgarisation scientifique », in *Social Science Information*, 24/4, Londres, Beverly Hills et New Delhi, pp. 825-845. Disponible sur : <http://ssi.sagepub.com/content/24/4/825.citation>. Consulté le 20 nov. 2012.

MORTUREUX (M.-F.) 1982, « Paraphrase et métalangage dans le dialogue de vulgarisation », in *Langue française*, Paris, 53, pp. 48-61. Disponible sur : http://www.persee.fr/web/revues/home/prescript/article/lfr_0023-8368_1982_num_53_1_5115. Consulté le 17 nov. 2012.

TICHENOR (P. J.), OLIEN (C. N.), HARRISON (A.) et DONOHUE (G.) 1970, “Mass Communication Systems and Communication Accuracy in Science News Reporting”, in *Journalism & Mass Communication Quarterly*, University of North Carolina, 47, pp. 673-683.

TERMIUM Plus (banque de données terminologiques et linguistiques). Disponible sur : <http://www.termiumplus.gc.ca/tpv2alpha/alpha-fra.html?lang=fra>. Consulté le 20 janvier 2013.

WIERZBICKA (A.) 1985, *Lexicography and conceptual analysis*, Ann Arbor, Karoma.

Équivalences 40/1-2 - 2013
ISSN 0751-9532