**Journées du CRTT 2012**

# *La néologie en langue de spécialité*

# *Détection, implantation et circulation des nouveaux termes*

Sous la direction de

Pascaline Dury, José Carlos de Hoyos, Julie Makri-Morel,

François Maniez, Vincent Renner, María Belén Villar Díaz

**Centre de Recherche en Terminologie et Traduction**

**Université Lumière Lyon 2**

# Sommaire

Présentation.....3

PARTIE 1

**OUTILS ET METHODES DE DETECTION DES NEOLOGISMES**

*L'activité prise au mot : le traitement des néologismes en entreprise*..... 13
Erlos F., Université Paris 3

*Augmenting terminology by crawling new term translation pairs from textual corpora: Some theoretical background and applicational directions* ..... 37
Kageura K., University of Tokyo

*Detección, análisis y clasificación de neologismos en corpus especializados: marcadores de creación léxica*..... 51
Roldán Vendrell M., Universidad de Jaén

*A view into retronymy as a source of neology*..... 75
Xydopoulos G. J. et Lazana I., University of Patras

PARTIE 2

**IMPLANTATION, CIRCULATION ET DIFFUSION DES NEOLOGISMES**

*Pour une néologie sociale*.....101
Quirion J., Université d'Ottawa

*Analyser la néologie dans le domaine de l'ingénierie nucléaire* .....121
Candel D. et Calberg-Challot M., Université Paris Diderot

*Linguistic and sociolinguistic factors that influence the detection, implantation and circulation of natural terminology in academic uses of Basque*.....141
Zabala I., San Martín I. et Lersundi M., University of the Basque Country

*Des communiqués de presse du Cnes à la presse généraliste. Vers un observatoire de la diffusion des termes*.....165
Condamines A. et Picton A., Université Toulouse 2 Le Mirail et Université de Genève

*Implantation of English terms including the -ING morpheme in French, Spanish and Italian: A corpus-based study of the debates of the European Parliament*.....189
Maniez F., Université Lumière Lyon 2

*Neologismos, políticas lingüísticas feministas y norma académica en los glosarios en línea Gentyll del ámbito del deporte*.....203
Bengoechea M., Universidad de Alcalá

## PARTIE 3

### APPROCHES DIACHRONIQUES

*Les néologismes techniques dans la traduction florentine de 1464 de l'*Opus agriculturae *de Rutilius Aemilianus Palladius*.....231
Campetella M., Université Catholique de Lyon

*Neologismos de ida y vuelta. De la* Doctrina de la neurona *de Cajal al* Principles of Neural Sciences *de Kandel*.....245
García Palacios J. et Sánchez Ibáñez M., Universidad de Salamanca

*Le traitement de la néologie technique dans la neuvième édition du* Dictionnaire de l'Académie Française .....267
Rey C. et Delacroix A., Université de Picardie Jules Verne

*La Tulipe « royne » : prolifération et circulation néologiques aux XVI^e^-XVII^e^ siècles*.....279
Selosse P., Université Lumière Lyon 2

Index des auteurs.....297

# Présentation

> *The story begins at the turn of the twentieth century, when Herbert J. Webber was in charge of the Plant-Breeding Laboratory of the United States Department of Agriculture. […] Webber was searching for a scientific term to denote those plants that are propagated vegetatively by buds, grafts, cuttings, suckers, runners, slips, bulbs, tubers, etc. […] He emphasized that he was looking for a term that could be used and understood by the general public. It should, therefore, be short, euphonious, phonetically spelled, easily pronounced and different from any other word in ordinary use, that it would not suggest any meaning other than the one desired. He suggested the word 'clon', meaning a twig, spray or slip, such as is broken off for propagation. 'Clon' (plural 'clons', adjective 'clonal') was to be applied to varieties of apples, pears, strawberries, etc., that are not propagated from seeds, to distinguish them from plants such as wheat and corn that are grown from seed. The word was to be pronounced with a long o, as in Greek.* (Mittwoch, 2002 : 381).

C'est par cette entrée en matière au caractère rétrospectif, sur la création du terme *clone,* que nous souhaitons présenter le volume qui suit sur la néologie en langues de spécialité. Dans cette citation, beaucoup de choses sont dites sur les termes nouveaux qui apparaissent dans des contextes spécialisés : sur leurs auteurs, sur les raisons de leur création, sur leur forme, mais aussi sur leur usage « anticipé » ou « programmé ».

Beaucoup de ces points d'ailleurs sont encore matière à réflexion en terminologie, bien que la néologie ait déjà suscité un volume de recherche considérable en langue générale, comme en langue de spécialité. De nombreuses pistes ont été proposées pour observer et analyser les termes nouveaux, beaucoup d'outils ont été élaborés pour les détecter, les repérer, des bibliographies sur le sujet longues et détaillées ont été produites[1].
Il suffit par exemple de parcourir le sommaire du numéro spécial 183 consacré par la revue *Langages* à la néologie en 2011 pour se rendre compte de l'extraordinaire richesse de travail qu'offre l'étude des termes nouveaux. On y rencontre ainsi des articles sur la néologie lexicale, la néologie syntaxique, la néologie combinatoire, la néologie rétrospective, le traitement des néologismes dans les dictionnaires, l'adaptation morphologique des emprunts néologiques, la cybernéologie, etc.

Malgré l'importance des recherches terminologiques consacrées à la néologie, beaucoup reste encore à faire car le champ d'étude est vaste ; d'ailleurs l'intérêt que portent les terminologues aux mots nouveaux ne semble pas faiblir. Ainsi,

---

[1] John Humbley (2012 : 12) évoque notamment la bibliographie de la néologie commencée par Jean-Claude Boulanger en 1970 et continuée dans le numéro 2 de la revue *Neologica* en 2008. On peut également consulter la liste des publications compilées par Jean-François Sablayrolles sur la néologie disponible à l'adresse suivante : http://www-ldi.univ-paris13.fr/index.php?option=com_contact&view=contact&id=39:jean-francois-sablayrolles&catid=82&Itemid=473

pour ne citer que deux sources récentes, peut-on évoquer deux numéros spéciaux de revues consacrés à la néologie : celui de *Terminology* sur les néologismes dans la communication spécialisée, paru en 2012 et dirigé par Maria Teresa Cabré Castellvi, Rosa Estopà Bagot et Chelo Vargas Sierra, et celui de *Neologica*, paru en 2012 également, qui porte essentiellement sur la néologie de traduction.

Le présent volume n'a pas pour ambition de proposer un état de l'art sur la question de la néologie, tant le champ à étudier est vaste, nous l'avons dit. Il ambitionne par contre de présenter un ensemble de réflexions variées mais complémentaires sur certains aspects liés à l'apparition de termes nouveaux dans un lexique donné. Les articles qui composent ce volume trilingue (français, anglais, espagnol) sont tous issus des travaux présentés lors du colloque international organisé à l'Université Lumière Lyon 2 les 2 et 3 juillet 2012 par le CRTT (Centre de Recherche en Terminologie et Traduction), sur le thème suivant : « la néologie en langue de spécialité : détection, implantation et circulation des nouveaux termes ».

Ce colloque avait pour objectif de lancer une réflexion sur divers aspects, parfois peu abordés, de la néologie dans les langues de spécialité, et notamment sur :

– les méthodes et les outils utilisés pour repérer la création lexicale et l'émergence d'une nouvelle conceptualisation,
– la place qu'occupent les corpus écrits et oraux dans la détection des néologismes,
– les méthodes et les outils utilisés pour mesurer et évaluer l'implantation d'un nouveau terme dans une langue de spécialité et sa circulation vers d'autres langues de spécialité,
– l'influence des politiques linguistiques sur l'implantation et la circulation des nouveaux termes,
– les évolutions diachroniques auxquelles sont sujets les néologismes,
– le traitement des nouveaux termes dans les dictionnaires généraux et spécialisés.

Dans la première partie du volume, la question des outils et des méthodes de détection des néologismes, ainsi que leur traitement, sera notamment abordée. Dans la première contribution, Frédéric Erlos démontre l'utilité et la singularité d'une approche textométrique pour détecter les néologismes spécialisés au sein de la production textuelle d'une entreprise et détaille les différentes étapes du processus : collecte des données, traitement du corpus par un logiciel de textométrie, repérage et étude des formes candidates au statut de néologisme. Le dispositif choisi permet de faire émerger, par contraste avec la norme que représente le vocabulaire de la série textuelle, des formes graphiques inédites ou particulièrement caractéristiques pour chaque nouveau texte ajouté au corpus,

et il a l'avantage de documenter de manière extrêmement fine l'apparition des néologismes au fil des discours.

Kyo Kageura aborde ensuite la question de la détection des néologismes sous un autre angle : depuis les années 1990, de nombreux travaux ont été produits afin d'améliorer l'extraction semi-automatique bilingue de néologismes dans des corpus textuels construits à cet effet, mais d'après l'auteur bien que ces travaux aient beaucoup contribué, notamment sur le plan technique, à faire progresser la détection des termes nouveaux, le passage du stade de la détection de ces termes en corpus à la construction d'une terminologie reste problématique. Il propose donc, dans sa contribution, une méthode de détection et d'extraction qui facilite la création de glossaires bilingues, ici japonais-anglais.

Dans l'article suivant, Mercedes Roldán Vendrell détaille la méthodologie utilisée pour la détection, l'analyse et la classification des néologismes telle qu'elle y a eu recours dans le cadre du projet *Olivaterm* (2008-2012), lequel visait à élaborer un dictionnaire terminologique multilingue du domaine scientifique et socioéconomique de l'huile d'olive. Une attention particulière est portée aux marqueurs lexico-syntaxiques non seulement utiles pour le repérage des néologismes mais qui permettent également d'établir et de préciser les relations lexico-sémantiques qu'entretiennent les néologismes et les autres unités appartenant au même domaine de spécialité (en particulier l'hypéronymie, la synonymie et l'antonymie) ; l'aboutissement de ces recherches est un modèle de terminologie multilingue dont le contenu est aussi bien adapté à la transmission de connaissances pointues en la matière qu'à la traduction spécialisée.

George J. Xydopoulos et Irene Lazana terminent cette première partie en décrivant le mécanisme de la rétronymie, en insistant notamment sur ses aspects sémantiques et sur sa contribution au domaine de la néologie. Après une définition détaillée du phénomène, ils se livrent à un réexamen des principales caractéristiques de la rétronymie et l'envisagent sous l'angle des relations taxonomiques d'(auto)-hyponymie et d'antonymie. Les phénomènes de polarité et de récurrence sont également analysés. Ils se livrent ensuite à un examen du statut morphologique des rétronymes et des néonymes et se penchent sur le processus d'abréviation des expressions rétronymiques, qu'ils considèrent comme l'indice d'un transfert du vocabulaire spécialisé vers le vocabulaire de la langue générale. Le mécanisme d'importation des rétronymes et néonymes de l'anglais en grec est également étudié, ainsi que quelques cas de rétronymie et de néonymie propres au grec.

La deuxième partie de l'ouvrage se consacre à l'implantation, la circulation et la diffusion des termes nouveaux, trois étapes consécutives mais étroitement liées dans le cycle de vie d'un néologisme. Comme le fait remarquer Jean Quirion (p.88), « implicitement, l'implantation est liée à la circulation, en ce sens qu'une création qui prend de la vigueur voyage nécessairement au sein de divers réseaux ». C'est d'ailleurs son article qui ouvre cette partie, une contribution qui

rappelle à quel point la néologie est sociale, dans tous les sens du terme : la terminométrie employée au fil des années pour mesurer l'implantation des termes nouveaux et pour établir des pronostics d'acceptation de tel ou tel néologisme a montré qu'en matière d'aménagement linguistique l'accent doit être mis sur la langue plutôt que sur un terme en particulier. Enfin, une participation accrue des locuteurs au processus néologique devrait être favorisée, notamment en pratiquant l'ouverture aux réseaux sociaux et l'externalisation massive (le recours à la créativité des internautes, par l'entremise d'un site Web, pour résoudre un problème, développer un projet, ou proposer un terme nouveau).

La nécessité de prendre en compte les utilisateurs des néologismes se trouve également au cœur de la seconde contribution de cette partie. Dans leur étude sur la néologie dans le domaine de l'ingéniérie nucléaire, Danielle Candel et Marie Calberg-Challot soulignent la nécessité d'un travail en relation étroite avec les spécialistes du domaine et détaillent leur double expérience, au sein de la Commission générale de terminologie et de néologie de la Délégation générale à la langue française et aux langues de France (DGLFLF) et en immersion dans une grande entreprise du nucléaire, Areva. Elles insistent notamment sur la relative instabilité des néologismes du domaine, sur un mouvement cyclique d'apparition-disparition-réapparition de certaines dénominations « problématiques » et concluent sur le fait que la mise en avant de termes nouveaux ne peut se faire sans la prise en compte des réactions du public, sans son approbation.

Igone Zabala, Itziar San Martin et Mikel Lersundi se penchent quant à eux sur les obstacles potentiels à la description, à l'implantation et au figement de la terminologie utilisée dans l'enseignement universitaire en langue basque. Dans la mesure où l'utilisation du basque comme langue véhiculaire à l'université détermine les connaissances des futurs spécialistes et leur utilisation adéquate de la terminologie, son développement naturel et son autorégulation en milieu universitaire constituent un enjeu crucial. Les auteurs décrivent la création du projet *Terminologia Sareak Ehunduz* (TSE), dont l'objectif est précisément de remédier aux carences susceptibles de nuire à l'implantation et au figement de cette terminologie (manque de réseaux de communication fluides entre experts, inaccessibilité de certains textes utilisés pour la communication en milieu universitaire). Le corpus créé dans le cadre du programme TSE et le recours aux opinions des experts recueillies dans divers forums et séminaires leur donnent l'occasion de décrire divers problèmes liés à la politique linguistique appliquée dans le cadre du système universitaire.

L'article proposé par Anne Condamines et Aurélie Picton s'intéresse, lui, à la circulation des termes nouveaux et notamment au phénomène de la dé-terminologisation, c'est-à-dire aussi bien au passage qu'à l'usage des termes spécialisés dans la langue dite générale. La problématique s'est élaborée à partir d'une demande du Centre National d'Etudes Spatiales (Cnes), qui souhaitait

voir comment le spatial « imprègne » la connaissance commune. Cette demande a été ré-interprétée par les auteurs en mettant en place un « observatoire » permettant d'examiner, à partir de la constitution de corpus et l'utilisation de méthodes outillées pour l'analyser, les phénomènes de dé-terminologisation. Les auteurs proposent quelques résultats déjà obtenus grâce à l'analyse comparée (statistique, lexicographique et sémantique) dans les corpus de l'observatoire, sur le fonctionnement de certains termes-candidats (particulièrement ceux construits à partir de *satellite*) et exposent les perspectives d'analyses, nombreuses, qui pourraient être menées grâce à la constitution de cet observatoire, sur les phénomènes de glissement des termes vers la langue générale.

Dans sa contribution, François Maniez s'intéresse aux anglicismes se terminant par le morphème *-ing* dans trois langues romanes – le français, l'espagnol et l'italien – et, à partir de données extraites du corpus *Europarl*, il étudie les choix de traduction des interprètes du Parlement européen afin de déterminer dans quelle mesure ceux-ci privilégient les emprunts directs à l'anglais ou font l'effort de proposer des équivalents construits en n'ayant aucunement recours à des morphèmes ou lexèmes anglais. Il remarque notamment que la longueur des anglicismes semble conditionner leur degré d'implantation et conclut son étude en soulignant que les interprètes contribuent à l'implantation de certains anglicismes dans les trois langues, même si la fréquence d'utilisation de ces termes décroît avec le temps au profit d'équivalents, et parfois – dans le cas du français – à celui d'officialismes, et que le phénomène varie d'une langue à l'autre, l'italien répugnant moins que le français et l'espagnol à l'anglicisation d'une partie de son lexique.

Pour conclure cette seconde partie du volume, Mercedes Bengoechea expose et examine les difficultés de créer un outil linguistique servant à intégrer les néologismes exprimant la valeur 'agent', homme ou femme, du domaine sportif (exemple : *delantero centro*, *delantera centro*). La proposition faite tient compte d'une réflexion sur le statut du genre et d'une analyse des politiques linguistiques féministes. Le résultat de ce projet de recherche se concrétise dans le glossaire en ligne Gentyll, consultable à l'adresse <http://gentyll.uah.es/glossaries.html>, qui regroupe des informations à la fois sur l'espagnol et sur l'anglais.

Selon John Humbley (2011 : 51), « la néologie est reconnue depuis longtemps comme composante essentielle de la terminologie, mais la prise en compte de sa dimension diachronique est bien plus récente ». Le but de la troisième partie de ce volume est justement de valoriser la dimension diachronique dans les recherches sur la néologie, au moyen de quatre contributions, et notamment celle de Moreno Campetella, qui utilise comme corpus la traduction florentine datant de 1464 de *l'Opus Agriculturae* de Rutilius Aemilianus Palladius (Vème siècle), version « vulgaire » de ce texte technique de l'antiquité latine tardive qui constitue la première traduction italienne de l'œuvre de Palladius. À travers les

nombreux néologismes lexicaux et sémantiques de ce texte, il décrit le processus de formation de la terminologie technique agricole italique, et plus généralement de la future « langue littéraire italienne commune ». Parmi les nouveautés lexicales observées, qui sont souvent le résultat d'une grande littéralité propre aux traductions italiennes du XVème siècle, il relève certains néologismes lexicaux, mais aussi des néologismes sémantiques et des calques. La traduction est un reflet de l'attitude de l'*intelligentsia* florentine du XVème siècle, dont la langue littéraire avait déjà atteint son apogée depuis au moins deux siècles, mais dont le lexique technique était encore en formation.

Joaquín García Palacios et Miguel Sánchez Ibáñez, quant à eux, décrivent la création néologique dans le domaine des neurosciences à deux moments différents de l'histoire, à savoir, le passage du XIXe au XXe siècle, d'une part, et l'époque contemporaine, d'autre part. Leur analyse des mots retenus pour véhiculer les notions essentielles de « neurone » et de « synapse » leur permet non seulement de décrire le processus d'implantation des néologismes étudiés (de l'indétermination initiale, caractérisée par la concurrence de plusieurs formes, parfois syntagmatiques, à la sélection d'une d'entre elles, acceptée par la totalité de la communauté scientifique), mais aussi de rendre compte de l'existence d'un important changement de perspective entre les deux époques évoquées. Selon eux, en effet, le multilinguisme favorisant la circulation de néologismes issus de différentes langues, dont l'espagnol, propre aux articles de la fin du XIXe siècle disparaîtra progressivement pour laisser place à un monolinguisme axé sur l'anglais, caractérisant la néologie scientifique actuelle.

Dans leur article, Christophe Rey et Amandine Delacroix abordent la question du traitement de la néologie technique au sein du répertoire bien particulier que constitue le Dictionnaire de l'Académie Française, ouvrage caractérisé par un rythme de publication particulièrement prudent, mais ayant aussi historiquement fait le choix de ne pas traiter les lexiques spécialisés. Les analyses des auteurs portent, dans ce cadre, sur la neuvième édition du dictionnaire (publiée à partir de 1992), la dernière à ce jour et toujours en cours de rédaction.

Enfin, en clôture de l'ouvrage, Philippe Selosse, en évoquant la Tulipe « royne », propose une réflexion très riche sur la dimension diachronique de la néologie. Selon lui (p.240), « pour étudier, en diachronie, la formation des néologismes, leur circulation et leur évolution dans un discours de spécialité, il faut un poste d'observation qui permette à la fois de cerner avec certitude l'apparition des néologismes et de disposer d'un corpus quantitativement représentatif du phénomène de création, diffusion et amplification de ceux-ci ». C'est le cas en effet pour la tulipe, au centre de cet article : l'auteur décrit comment sa découverte en Turquie au XVIème siècle a suscité en un siècle (1560-1660) la création, puis la diffusion de nombreux néologismes en langue vernaculaire (par les jardiniers, par les fleuristes et les amateurs) tout comme en langue de spécialité (par les botanistes).

Le comité d'organisation du colloque (Pascaline Dury, José Carlos de Hoyos, Julie Makri-Morel, François Maniez, Vincent Renner, María Belén Villar Díaz) tient, pour conclure, à remercier chaleureusement les membres du comité scientifique pour les lectures et relectures attentives et avisées des propositions de communication reçues, qui ont permis de proposer un programme riche et de qualité.

Nous tenons également à remercier sincèrement Kyo Kageura, de l'université de Tokyo, et Jean Quirion, de l'université d'Ottawa, dont les contributions figurent dans cet ouvrage, et qui ont accepté de présenter les conférences plénières avec beaucoup de talent et de disponibilité. Nous souhaitons enfin également témoigner notre reconnaissance au Conseil scientifique de l'université Lyon 2, à la Région Rhône-Alpes et à la ville de Lyon, ainsi qu'à la Délégation générale à langue française et aux langues de France (DGLFLF) pour le soutien financier apporté à ce colloque.

Pascaline Dury

## Références

BOULANGER J.-C. (2008), Chronologie raisonnée des bibliographies de la néologie précédée de quelques miscellanées, *Neologica*, 2, 185-200.
CABRE CASTELLVI M. T., ESTOPA R., VARGAS SIERRA C. (2012), Neology in specialized communication, *Terminology*, numéro spécial, 18 (1).
WEBBER H. J. (1900), Work of the United States Department of Agriculture on plant hybridization, *Journal of the Royal Horticultural Society*, 24, 128-145.
HUMBLEY J. (2012), Présentation, *Neologica*, 6, 9-16.
MEJRI S., SABLAYROLLES J.-F. (dir.) (2011), Numéro spécial sur la néologie, *Langages*, 183.
MITTWOCH U. (2002), 'Clone': The History of a Euphonious Scientific Term, *Medical History*, 46, 381-402.
*Neologica*, Revue internationale de terminologie, 2012, 6.

# Partie 1

## OUTILS ET METHODES DE DETECTION DES NEOLOGISMES

# L'activité prise au mot : le traitement des néologismes en entreprise

Frédéric Erlos[1]

## Résumé

Le sociolecte en usage dans les organisations mêle différents degrés de spécialisation de la langue et se présente rarement sous la forme d'un lexique formalisé faisant autorité. Afin de prendre la mesure de ces caractéristiques, nous proposons de limiter à certaines situations de communication la collecte des nouveautés relatives à un tel code fonctionnel. En effet, l'utilisation récurrente des mêmes genres du discours produit des normes de sélection lexicale, dont les évolutions peuvent faire l'objet d'un suivi dans le temps. Les logiciels de textométrie permettent alors d'outiller l'exploitation de corpus organisés sous la forme de séries textuelles chronologiques, en vue du repérage de formes candidates au néologisme. Ces formes sont ensuite évaluées depuis trois points de vue complémentaires : leur place dans le référentiel servant de cadre à la communication, leur fonctionnement comme support d'un objet de discours, et enfin, leur appartenance à l'une des composantes du parler d'entreprise. Une collecte de néologismes, réalisée durant plusieurs mois à partir d'un corpus de circulaires internes, sert à illustrer le propos et met en lumière l'importance du renouvellement de la terminologie d'une entreprise.

## Mots clés

Néologisme ; néonyme ; entreprise ; langue de spécialité ; parler d'entreprise ; terminologie d'entreprise ; référentiel ; genre du discours ; objet de discours ; corpus ; linguistique de corpus ; série textuelle chronologique ; textométrie.

## Introduction

Une entreprise crée et modifie en permanence les objets qui sont utiles à son fonctionnement : infrastructure, organisation, production et commercialisation de produits et de services. Les connaissances nécessaires à la réalisation de cette activité sont transmises à l'intérieur de l'organisation sur les supports et avec les formats les plus divers : présentations orales, publications, messagerie électronique, sites intranets, espaces de travail collaboratifs. Le néologisme, défini à ce stade comme « *une unité nouvelle, de nature lexicale, dans un code linguistique défini* [2]», manifeste et véhicule le mouvement impulsé par l'activité, en même temps qu'il représente un élément perturbateur pour l'état des connaissances

[1] SYLED-CLA2T, Université Paris 3, ferlos@orange.fr

[2] A. REY (1976 : 4)

précédant son apparition. Dès lors, le repérage et le référencement de ces unités apparaissent comme étant de nature à renforcer la clarté et l'efficacité de la communication nécessaires à l'activité. En effet, durant la période d'instabilité qui accompagne l'apparition d'un mot nouveau, des tâches telles que l'explicitation d'une notion, l'association d'un sigle à son développé ou encore la description d'une entité sont étroitement associées à la mise en cohérence des contenus, à l'harmonisation terminologique, ainsi qu'à la traduction. Cependant, nous constatons que ces traitements terminographiques se heurtent à plusieurs difficultés portant à la fois sur le choix des unités néologiques et sur la spécification de la nouveauté qui les caractérise. Nous décrivons une méthodologie destinée à surmonter ces obstacles, au moins partiellement, et à asseoir sur des bases consistantes un dispositif textométrique opérationnel, que nous présentons et dont nous discutons les résultats, obtenus à partir d'un corpus de documents réservés à l'usage interne.

## 1. Le choix des unités néologiques dans le contexte de l'entreprise

La mise en place d'un traitement des néologismes en entreprise nécessite qu'un périmètre de collecte adapté soit défini. Pour cela, il faut au préalable tenter de situer le sociolecte propre à une organisation entre les deux limites représentées par le lexique de la langue commune et l'idiolecte de chaque locuteur. D'autre part, le traitement des unités nouvelles ayant pour but de faciliter les échanges, il convient de veiller à ce qu'il soit réalisé au plus près du contexte réel des communications verbales.

### 1.1. Détermination de la portée de la collecte

#### 1.1.1. Néologismes et spécificités d'un sociolecte

Les discours produits au sein d'une entreprise mobilisent différentes strates de la langue : au fonds commun s'ajoutent, d'une part, les langues spécialisées en usage dans les secteurs où l'activité se déploie et, d'autre part, la terminologie propre à toute organisation. Dardo de Vecchi a proposé de rendre compte de cette réalité à l'aide de la notion de « parler d'entreprise »,

> pour désigner l'ensemble des processus linguistiques qui actualisent les répertoires linguistiques des membres d'une communauté, définie en fonction de l'appartenance à une entreprise. Autrement dit, la cristallisation linguistique de tout moyen de communication mis à la disposition d'une entreprise, pour des conceptualisations ayant des origines diverses[3].

L'adoption du parler d'entreprise comme norme de référence pour la collecte des néologismes conduit à privilégier un code fonctionnel situé[4]. Certes, un tel sociolecte agrège des composantes générales – langue commune et certaines de

[3] D. DE VECCHI (1999 : 316).
[4] A. REY (1976 : 15).

ses spécialisations –, mais sa spécificité s'enracine dans un foyer générateur de termes dont les productions deviennent centrales, en raison même du particularisme qu'elles manifestent. Les unités caractéristiques du parler d'entreprise doivent par conséquent faire l'objet d'une attention particulière lors de la collecte.

Cependant, le contexte de l'entreprise n'est généralement pas très favorable au développement d'une tradition qui aurait préalablement mis au net sa terminologie spécifique. Certes, le mouvement permanent que l'activité imprime à toutes les tâches qu'elle mobilise n'empêche pas l'apparition de glossaires ou de lexiques. Cependant, il faut constater que ces instruments de travail sont émiettés en fonction de besoins *ad hoc* et qu'ils sont traités avec un soin et un niveau de détail très variables. Non seulement ils ne sont que très rarement à jour, mais surtout, aucun d'entre eux ne peut prétendre représenter un état satisfaisant du parler d'entreprise. Or, l'hétérogénéité des normes vient compliquer le repérage des néologismes, qui nécessite que l'on puisse s'appuyer, entre autres, sur une image du code en usage pour une période donnée. Dans le but de réduire cette complexité, nous proposons, dans un premier temps, de renoncer au cadre constitué par une organisation prise dans sa globalité.

### 1.1.2. Néologismes et référentiel

Afin de répondre à cette exigence, il nous a semblé fructueux d'aborder la question à partir d'un modèle facilement adaptable et ouvert aux problématiques liées à l'activité. C'est ce qu'offre la notion de référentiel proposée par l'épistémologue Ferdinand Gonseth :

> (…) le mot *référentiel* en vient à signifier l'ensemble (explicite ou implicite) des préalables faute desquels telle ou telle activité systématique ne pourrait pas avoir lieu. Le référentiel peut alors s'offrir comme un cadre où cette activité prendra place. C'est ainsi que tout ce qui fait pour nous, en telle ou telle circonstance, la « réalité » d'un terrain de jeu peut servir de référentiel commun au joueur et au spectateur[5].

Le cadre offert par le référentiel pouvant s'appliquer à toute forme d'activité, il permet de délimiter un contexte de collecte plus fin que celui offert par l'entreprise ou par les secteurs d'activité. Cette notion possède également la vertu de mettre en avant l'articulation nécessaire entre les référentiels individuels et collectifs, ce qui permet de poser la question du cadre de la collecte en termes de connaissances mobilisées par et pour la réalisation de l'activité.

En effet, quelle que soit la situation, la réalisation de l'activité présuppose l'existence d'un ensemble de connaissances partagées, au moins jusqu'à un certain point. Par conséquent, toutes les unités faisant appel à la compétence cognitive des acteurs, plus qu'à leur compétence linguistique[6], doivent faire

---

[5] F. GONSETH (1975 : 22).

[6] J. HUMBLEY (2006 : 122).

l'objet d'un traitement terminographique dans le cadre d'un référentiel donné. Dès lors que l'enjeu cognitif prédomine, la collecte des unités doit être étendue dans deux directions. La première concerne, comme nous l'avons vu plus haut, les différentes strates linguistiques dont le parler d'entreprise est composé. La seconde porte sur la diversité des natures d'unités susceptibles d'être collectées : lexie, syntagme lexicalisé, terme simple ou complexe, mais aussi unité brachygraphique[7] et dénomination propre[8]. Les critères de sélection des néologismes, que ceux-ci correspondent à une nouvelle unité – les emprunts faits à une autre langue et les changements de nature grammaticale sont rangés dans cette catégorie –, ou à la modification d'un concept ou d'une description déjà associés à une unité existante, doivent intégrer cette diversité et favoriser sa restitution.

## 1.2. Spécification de la nouveauté des néologismes

Dès lors que l'actualité des échanges impose son rythme à la collecte, celle-ci se déroule à l'échelle d'une micro-diachronie ayant pour toile de fond l'évolution d'un code fonctionnel situé. Nous avons vu que celui-ci ne fait pas l'objet d'un recensement susceptible de matérialiser la norme linguistique pour une période donnée. En revanche, il est possible d'aborder cette question d'un autre point de vue, en recherchant le moyen de capter l'évolution du code au fur et à mesure qu'elle se produit. Les néologismes apparaissant dans les discours, cela revient à se demander dans quelle mesure la sélection lexicale et phraséologique opérée lors de la production des discours peut également fonctionner comme substitut d'un code fonctionnel de référence. Cependant, le vocabulaire des discours ne peut être utilisé comme norme que si nous pouvons avoir prise, d'une part, sur les facteurs extralinguistiques qui sont susceptibles de le modifier et, d'autre part, sur la manière d'utiliser les propriétés d'un tel vocabulaire, afin qu'il fonctionne à son tour comme une norme.

### 1.2.1. Le genre du discours comme norme située

Un référentiel sert de cadre à des communications verbales dont les caractéristiques principales varient en fonction de la situation, des diverses compétences des interlocuteurs, des propriétés de l'univers du discours et du genre du discours utilisé. Celui-ci permet d'articuler une spécification fine des critères extralinguistiques caractérisant la situation de communication avec l'observation des usages linguistiques :

> L'utilisation de la langue s'effectue sous la forme d'énoncés concrets uniques (oraux et écrits) qui émanent des représentants de tel ou tel domaine de l'activité humaine. L'énoncé reflète les conditions spécifiques et les finalités de chacun de ces domaines, non seulement par son contenu (thématique) et son style de

[7] R. KOCOUREK (1991 : 93-94).

[8] I. TAMBA, B. BOSREDON, (1994 : 123) : « (…) *toutes les formes fixes de dénominations d'un référent unique, françaises ou étrangères, sans considération de statut grammatical* (…) ».

> langue, autrement dit par la sélection opérée dans les moyens de la langue – moyens lexicaux, phraséologiques et grammaticaux –, mais aussi et surtout par sa construction compositionnelle. Ces trois éléments (contenu thématique, style et construction compositionnelle) fusionnent indissolublement dans le tout que constitue l'énoncé, et chacun d'eux est marqué par la spécificité de la sphère d'échange. Tout énoncé pris isolément est, bien entendu, individuel, mais chaque sphère d'utilisation de la langue élabore ses types relativement stables d'énoncés, et c'est ce que nous appelons les genres du discours[9].

Cependant, si tous les discours produits[10] dans une telle sphère d'échange sont, en théorie, dignes d'intérêt pour la collecte des néologismes, les contraintes pratiques du recensement et de la description obligent à privilégier les échanges qui, d'une part, laissent des traces facilement exploitables, donc plutôt écrites, et d'autre part, empruntent des formes relativement stabilisées dans la durée. En effet, c'est à travers ses réalisations discursives que le code fonctionnel mobilisable dans le cadre d'un référentiel donné peut être observé. Si les conditions de production des discours remettaient en jeu à chaque échange la totalité d'un parler d'entreprise, c'est la possibilité même de l'observation des néologismes qui serait menacée de dissolution. Afin de prendre en compte ces contraintes, le repérage des néologismes nécessite la mise en place d'un dispositif qui repose, en définitive, sur l'identification de foyers d'énonciation stabilisés observables à partir de leurs productions écrites.

Par ailleurs, l'intérêt présenté par tel ou tel genre du discours varie en fonction des finalités de la collecte[11]. De manière générale, et pour les néologismes apparaissant dans un parler d'entreprise en particulier, les genres du discours utilisés pour diffuser des informations devant passer d'une population initiée à une autre qui l'est moins, constituent de bons postes d'observation. Il peut s'agir, par exemple, des informations internes à transmettre d'une communauté de locuteurs vers d'autres métiers, d'une maison mère vers ses filiales ou de l'entreprise vers des destinataires externes (pouvoirs publics, analystes, journalistes, clientèle, etc.). Les indications explicites destinées à signaler les nouveautés sont plus souvent présentes dans les genres du discours d'interface que dans ceux utilisés pour les échanges entre pairs. D'autre part, le genre du discours d'autorité vise à imposer à ses lecteurs l'obligation de prendre en compte l'information nouvelle en même temps que la bonne manière de la dénommer. À défaut de présenter un terrain riche en termes concurrents, ces discours livrent une version officielle de la dénomination, qui est destinée à occuper le statut de vedette dans les référentiels terminologiques. Il apparaît ainsi que l'attention prêtée au genre du discours permet de situer la collecte et de prévoir ses lacunes, mais aussi d'observer le fonctionnement des unités dans

---

[9] M. BAKHTINE (1984 : 265).
[10] A. BORZEIX, B. FRANCKEL (2005 : 231-261).
[11] J. PEARSON (1998), A. PICTON (2009).

leur cotexte et de privilégier, comme nous le verrons plus loin, celles dont le référent fonctionne comme un objet de discours[12].

1.2.2. La série textuelle chronologique comme norme endogène

Les garanties offertes par la sélection d'un ou de plusieurs genres du discours adaptés gagnent à être renforcées pour que le vocabulaire d'un corpus puisse fonctionner comme une norme. En effet, les occurrences d'un type d'énoncé peuvent respecter plus ou moins leur modèle ; elles sont également susceptibles de provenir de pôles différents au sein d'une même sphère d'échanges ; enfin, leur production peut être plus ou moins espacée dans le temps. Ce sont là encore autant de propriétés externes dont les variations sont susceptibles de perturber l'étude des effets de la variable temps sur l'utilisation d'un vocabulaire. Il convient donc de les maîtriser dans la mesure du possible. C'est pourquoi les corpus que nous utilisons ont été constitués sous la forme de séries textuelles chronologiques :

> Nous réserverons donc l'appellation de série textuelle chronologique aux séries périodiques homogènes constituées par des textes produits dans des situations d'énonciation similaires, si possible par un même locuteur (individuel ou collectif) et présentant des caractéristiques lexicométriques comparables[13].

Dans ces corpus, chaque texte est en relation avec la norme que constitue le vocabulaire de la série textuelle, c'est-à-dire l'ensemble formé par tous les mots attestés au moins une fois dans l'un des textes de la série. Lorsqu'un nouveau texte est ajouté à un corpus de suivi (*monitoring corpus*), cette relation peut revêtir des aspects variables répartis entre deux extrêmes théoriques : une absorption complète, le nouveau texte ne contient que des mots déjà utilisés, et une originalité totale, la dernière occurrence de la série n'est constituée que de mots nouveaux. Un tel dispositif promeut le vocabulaire mobilisé dans le corpus au statut de norme de référence pour un contexte précis. Les mots qui viennent enrichir la norme revêtent alors le statut de formes candidates pour le repérage des néologismes, soit en raison de leur nouveauté, soit parce que leur fréquence dans le texte ajouté à la série est particulièrement élevée par rapport à la fréquence attendue pour l'un quelconque des autres textes du corpus.

Pour cette étude, nous avons constitué un corpus ouvert de circulaires internes, diffusées par l'organe central d'un groupe bancaire vers d'autres entités

---

[12] J.-B. GRIZE (1996), F. SITRI (1998). Dans la logique naturelle de Jean-Blaise Grize, l'objet de discours est à la fois une notion associée dans les discours à une ou plusieurs expressions nominales et le résultat d'une schématisation qui « (…) *est une représentation discursive orientée vers un destinataire de ce que son auteur conçoit ou imagine d'une certaine réalité*. » (1996 : 50). Si l'attention de cet auteur se porte principalement sur les opérations logiques utilisées dans l'élaboration des schématisations, d'autres auteurs, dont Frédérique Sitri, ont cherché à mettre en lumière les moyens linguistiques mis en œuvre pour la construction de ces objets, dans et par les discours.

[13] A. SALEM (1988 : 107).

appartenant au même groupe, à raison de deux cents par an environ[14]. Ces circulaires viennent modifier, ajuster ou enrichir les processus en cours durant l'année. Élaborées par les différents services de cet organe central, elles se présentent sous la forme de quelques pages obéissant à un formalisme assez strict (cartouche introductif normalisé, différentes relectures avant publication, signature de l'émetteur).

## 2. Un filet à papillons textométrique

L'exploitation d'une série textuelle chronologique sous la forme d'un corpus ouvert commence par le formatage des textes devant faire l'objet des traitements informatiques. Nous nous limitons ici à mentionner les opérations textométriques[15] fondamentales nécessaires et suffisantes pour le but poursuivi. Celles-ci peuvent ensuite être complexifiées en fonction des finalités recherchées et des possibilités des logiciels[16].

### 2.1. Évolution diachronique du vocabulaire dans un corpus ouvert

Lorsqu'il a été rendu exploitable par le logiciel de textométrie, le corpus est segmenté en formes graphiques, c'est-à-dire en suites de caractères encadrées par deux espaces ou par des caractères délimiteurs, comme les signes de ponctuation. Toutes les formes graphiques recensées à ce stade sont associées à une fréquence, correspondant au nombre de leurs occurrences dans le corpus, et à un numéro d'ordre lexicométrique, obtenu en triant les fréquences par ordre décroissant. En fonction de la partition retenue pour le corpus, dans le cas le plus simple chaque texte représente une partie, le logiciel construit le tableau lexical entier. Dans ce tableau à double entrée, le lexique des formes graphiques représente la bordure verticale gauche et la succession des parties la bordure horizontale haute. Le tableau offre sur chaque ligne les fréquences d'une forme graphique dans les différentes parties du corpus, et dans chaque

---

[14] Dans le cadre d'un autre travail, nous avons utilisé un corpus de rapports d'activité dont la publication est annuelle. Les exigences réglementaires qui imposent ce type de bilan annuel valent aussi bien pour l'extérieur que pour l'intérieur des organisations de taille importante. F. ERLOS (2009).

[15] La textométrie recouvre le champ de la lexicométrie, celle-ci étant définie comme un « *ensemble de méthodes permettant d'opérer des réorganisations formelles de la séquence textuelle et des analyses statistiques portant sur le vocabulaire d'un corpus de textes* » (L. LEBART et A. SALEM, 1994 : 314).

[16] Liste non limitative de logiciels de textométrie : http://adcost-elliadd.univ-fcomte.fr/astadiag/ (Astadiag), *http://www.dtmvic.com/05_SoftwareF.html* (DmtVic), http://ancilla.unice.fr/?redirected_from=ancilla.unice.fr/ ~brunet/pub/hyperbase.html (Hyperbase). *http://www.tal.univ-paris3.fr/lexico/* (Lexico 3), http://www.ling .uqam.ca/ato/sato/ (Sato 4.2), http://textometrie.ens-lyon.fr/ (plateforme TXM 0.6), http://www.xaira.org/ (Xaira 1.25). On pourra consulter pour d'autres approches, par exemple P. DROUIN et al. (2006) et A. PICTON (2009).

colonne, les occurrences des formes graphiques dans l'une des parties du corpus.

| Ordre lexicométrique | Formes | Partie 1 | P 2 | P 3 | P 4 | P 5 |
|---|---|---|---|---|---|---|
| 1 | **de** | 20 | 36 | 14 | 21 | 21 |
| 2 | **la** | 6 | 10 | 3 | 15 | 16 |
| 3 | **des** | 2 | 22 | 9 | 22 | 7 |
| 4 | **et** | 4 | 15 | 5 | 16 | 6 |
| 5 | **à** | 3 | 7 | 4 | 4 | 15 |

*Tableau 1 : Extrait du tableau lexical entier (TLE) du corpus de circulaires 2011-2012 présentant les cinq premières formes et les cinq premières parties*

À partir de ces données fondamentales, l'exploitation textométrique d'un corpus fournit des indications, à la fois générales et particulières, sur la manière dont la temporalité affecte l'évolution d'un vocabulaire. Pris dans sa globalité, le vocabulaire du corpus s'enrichit de formes nouvelles au fur et à mesure que de nouveaux textes sont ajoutés. L'exploitation informatique de ces données permet de préciser dans quelle proportion se fait cet accroissement.

*Graphique 1 : Accroissement du vocabulaire d'un corpus de circulaires*

Commentaire : le graphique ci-dessus représente la manière dont les 10 000 formes graphiques du lexique d'un corpus ouvert de circulaires (année 2011 et les cinq premiers mois de 2012), se répartissent sur les 160 000 vocables dont se composent les textes du corpus. En mai 2012, le corpus comportait 269 parties, ce qui représente une moyenne de 600 occurrences par texte. Le graphique met en évidence un tassement de l'accroissement du vocabulaire au fur et à mesure que le nombre d'occurrences s'accroît. Alors que 40 000 occurrences ont été suffisantes pour faire apparaître 5000 formes différentes, plus de 100 000 occurrences supplémentaires ont été nécessaires pour doubler cette quantité. Le graphique indique également, pour les cinq premiers mois de l'année 2012, une progression d'un millier de formes nouvelles réparties sur 83 textes, soit en moyenne une douzaine de formes nouvelles par circulaire. Le calcul a été réalisé avec le logiciel Lexico 3.

La mesure de l'accroissement du vocabulaire fournit ainsi un ordre de grandeur pour la quantité de formes nouvelles à laquelle il faut s'attendre dans chaque

partie nouvelle ajoutée au corpus, compte tenu de sa position au sein de la série textuelle chronologique. Cet accroissement participe d'un mouvement plus profond, caractéristique des séries textuelles et qu'André Salem a proposé de nommer « temps lexical ». Il consiste en un renouvellement progressif du vocabulaire du corpus :

> tout émetteur produisant des textes sur une période de temps assez longue utilise sans cesse de nouvelles formes de vocabulaire qui viennent supplanter, du point de vue fréquentiel, d'autres formes dont l'usage se raréfie. (…) Il s'ensuit que les vocabulaires de parties correspondant à des périodes consécutives dans le temps présentent en général plus de similitudes entre eux que les vocabulaires correspondant à des périodes séparées par un intervalle de temps plus long[17].

Dans le corpus de circulaires que nous utilisons, le « texte » est à considérer ici globalement comme une année de production de circulaires. L'année 2011 sert de référence pour l'année 2012 en cours. Il aurait également été possible de descendre au niveau de chaque service émetteur de circulaire mais, compte tenu du but poursuivi, l'émetteur considéré ici est l'organe central, pris comme locuteur collectif s'adressant aux autres entités du groupe bancaire.

Par ailleurs, une année d'antériorité s'est révélée suffisante pour accéder aux variations du temps lexical de la série textuelle chronologique prise comme exemple. En effet, les variations lexicales ont des causes multiples que l'on peut essayer de ramener à deux des caractères du genre du discours proposés par Mikhaïl Bakhtine : le contenu thématique et la sélection opérée dans les moyens de la langue. Le premier est susceptible de varier avec chaque circulaire, mais nous observons qu'un grand nombre de sujets ont un caractère cyclique, correspondant dans la majorité des cas à un rythme annuel. Une année d'antériorité permet ainsi de repérer les changements, y compris lorsqu'ils se font à l'intérieur d'un cadre préconstitué depuis plusieurs années. La seconde cause de variation ajoute ses effets à la précédente. Elle dépend cette fois de chaque rédacteur. Celui-ci introduit dans le texte de la circulaire son propre usage de la langue, ainsi que des choix lexicaux dépendant de sa plus ou moins grande maîtrise des codes du genre et de la terminologie du secteur concerné. Il en résulte, entre autres, l'apparition de variantes dont il faut veiller à contenir la portée, compte tenu du renouvellement des rédacteurs. Là encore, le choix d'une année d'antériorité pour le repérage des néologismes a constitué un compromis adapté à ce genre du discours. Enfin, comme nous le montrons plus loin, si la norme de vocabulaire peut être ramenée à une année d'antériorité lors du repérage des néologismes, les dimensions diachronique et générique du corpus gagnent à être étendues avant l'enregistrement du néologisme dans un référentiel terminologique.

---

[17] A. SALEM (1988 : 107).

### 2.2. Formes originales, formes spécifiques et néologismes

Au fur et à mesure que de nouvelles circulaires sont diffusées et intégrées au corpus, le dispositif mis en place permet d'interroger l'apport d'un nouveau texte à partir de deux mesures complémentaires : les formes originales et les formes spécifiques. Nous présentons et discutons chaque méthode à partir des néologismes collectés dans le corpus de circulaires.

#### 2.2.1. Les formes originales ne sont pas toutes des néologismes et réciproquement

La première méthode utilisée pour le repérage et l'établissement des néologismes consiste à isoler, dans le vocabulaire du nouveau texte, les formes originales, c'est-à-dire celles qui lui appartiennent exclusivement par rapport au reste de la série. Elles constituent, en quelque sorte, l'apport net du nouveau texte à l'accroissement du vocabulaire du corpus et, à ce titre, elles sont de bonnes candidates pour le repérage des néologismes. Ces formes sont obtenues en testant les lignes du tableau lexical entier. Celles pour lesquelles la somme des fréquences observées dans l'ensemble des parties du corpus, la dernière partie ajoutée comprise, est égale à la fréquence observée pour cette dernière partie de la série, correspondent à des formes graphiques originales. Les données retenues, à titre d'exemple, proviennent d'un échantillon de circulaires émises durant le premier trimestre de l'année 2012, soit cinquante-quatre textes. Nous avons vu plus haut que chaque partie nouvelle ajoutée au corpus apporte en moyenne une douzaine de formes nouvelles pour une longueur moyenne de 600 occurrences. Nous observons que les formes originales ne sont pas toutes associées à la même fréquence. Sans qu'il soit possible de dire qu'il s'agit d'une règle générale, l'observation montre que les formes originales possédant la fréquence la plus élevée sont fréquemment associées à des objets de discours, alors que les hapax relèvent plutôt de la mention simple. Si ces formes candidates se révèlent être des néologismes, leur fonctionnement dans les textes comme dénomination d'un objet de discours permet d'établir la nouveauté avec plus de certitude, dès ce premier stade. En effet, le cotexte fournit dans ce cas des indications utiles pour élaborer une définition ou une description associée à l'unité terminologique. Dans l'exemple ci-dessous, la dénomination « Amplia » correspond à un nouveau produit, et les formes « Bonne » et « Compréhension » font partie du terme complexe « Recueil de Bonne Compréhension », désignant un formulaire, rempli par le souscripteur d'un produit financier, et permettant à l'établissement bancaire de recueillir la preuve que son client a bien compris la nature du produit vendu et les risques qui lui sont associés.

| Formes originales, partie 187 | Fréquence |
|---|---|
| Amplia | 3 |
| AMPLIA | 2 |
| Bonne | 1 |
| Compréhension | 1 |
| Première | 1 |

*Tableau 2 : Formes originales de la partie 187*

Cependant, les formes originales détectées par le test ne sont pas toutes pertinentes pour le repérage des néologismes, loin s'en faut. Ces formes ont diverses provenances, qui peuvent être réparties en fonction de la typologie suivante :

– les coquilles ; elles sont rares, mais elles ne sont pas absentes, malgré les correcteurs d'orthographe et les relectures ;

– les variantes graphiques (capitale attribuée à un mot dans le cours d'une phrase, voire écriture du mot en majuscules) ; elles sont plus nombreuses que les coquilles, en raison de l'absence de normalisation du corpus sur ce point, mais elles sont pertinentes pour le repérage des néologismes, les rédacteurs utilisant fréquemment ce type d'artifice typographique pour signaler une nouveauté. Les graphies « MEMO » et « OBJECTIVITE » sont, par exemple, utilisées pour désigner, dans le premier cas, un format de fiche explicative destinée à la clientèle, et dans le second, un engagement commercial. Dans le tableau 2, la forme « Première » doit sa majuscule à la position occupée par le mot en tête de phrase.

– les coordonnées spatio-temporelles et celles relatives à l'organisation apportent une certaine quantité de dénominations propres, présentes sous la forme d'une simple mention (personnes, lieux, unités organisationnelles).

– enfin, la thématique propre à la circulaire entraîne la sélection de mots qui font à cette occasion une apparition ponctuelle. Par exemple, pour la présentation d'un service proposé aux collectivités publiques, les mots « usagers », « habitants » et « loisirs » font leur entrée dans le vocabulaire du corpus, mais cela ne manifeste rien d'autre qu'un usage banal de la langue commune. Il s'agit alors d'un phénomène de rareté relative, liée à la taille réduite du corpus, ainsi qu'au genre du discours auquel les circulaires appartiennent.

En deuxième lieu, il apparaît que ce test doit être complété par la recherche des formes spécifiques car, d'une part, il ne permet pas de détecter toutes les

formes pertinentes pour le repérage des néologismes (un quart des néologismes relevés dans l'échantillon lui ont échappé), et d'autre part, il ne fournit pas d'indications très riches pour guider l'exploration des contextes où les formes apparaissent. Or, cette vérification des contextes est assez contraignante : il faut la recommencer au moins une dizaine de fois par circulaire, alors que 40% des textes se révèlent improductifs en matière de néologisme.

2.2.2. Néologismes et calcul des spécificités

La seconde mesure utilisée pour repérer et établir les néologismes correspond à un jugement porté en probabilité sur la fréquence des formes présentes dans une nouvelle partie ajoutée au corpus. Le calcul des spécificités permet d'évaluer si la fréquence réelle observée pour telle forme dans telle partie du corpus correspond à une surreprésentation, une sous-représentation ou à une valeur proche de la valeur modale, c'est-à-dire de la valeur la plus probable[18]. Pour un seuil donné en deçà duquel on estimera sortir de la banalité, par exemple 5% de chances, une forme spécifique positive correspondra à une surreprésentation, une forme spécifique négative équivaudra à une sous-représentation et une forme banale sera une forme dont la fréquence est statistiquement attendue. L'intérêt de la spécificité positive est d'attirer l'attention sur le suremploi d'une forme graphique (originale ou non). Celui-ci fonctionne alors comme un bon indice pour repérer un changement intervenant, soit dans la signification d'un mot ou d'un syntagme lexicalisé existant, soit dans la description d'une entité, ou bien encore, pour détecter l'apparition d'un terme ou d'une dénomination propre complexe[19].

De façon plus générale, le calcul des spécificités fait émerger les principaux éléments thématiques caractéristiques de la partie étudiée[20]. Ce sont alors les formes graphiques dotées des coefficients de spécificité les plus élevés qu'il est intéressant de retenir. Dans tous les cas issus de l'échantillon de textes retenus pour cette étude, le calcul des spécificités a fourni les formes candidates pour le repérage des néologismes qui n'avaient pas été détectés à partir des seules formes originales. Ce calcul sélectionne également les formes originales, comme le montre l'exemple donné ci-dessous, mais il a tendance à les noyer parmi les autres formes spécifiques positives. Les limites de ce calcul, lorsqu'il est

---

[18] Les probabilités sont calculées à partir du modèle hypergéométrique qui prend en compte la taille du corpus (le nombre total d'occurrences), la taille de la partie étudiée, la fréquence totale de la forme retenue et la fréquence de celle-ci dans la partie étudiée.

[19] Voir, entre autres, F. ERLOS (2009 : 659-764).

[20] Cette méthode est également utilisée pour repérer les différences pouvant affecter les réseaux de co-occurrents associés à une forme. Pour ces explorations complémentaires, les logiciels de textométrie offrent en outre la possibilité d'observer finement les cotextes dans lesquels une forme apparaît. Les classiques concordances sont alors sollicitées, de même que les possibilités offertes par des subdivisions plus fines que les textes, comme les paragraphes, afin de mettre en lumière des marqueurs discursifs destinés à signaler la nouveauté.

appliqué au repérage des néologismes, sont discutées dans la troisième section. Nous constatons que lors de la phase de repérage, cet inconvénient est compensé par le fait que les formes graphiques possédant un coefficient de spécificité positif élevé entrent pratiquement toujours dans la caractérisation du néologisme, c'est-à-dire dans la définition du terme ou la description de l'entité nommée. L'indication fournie par la fréquence de certaines formes originales, afin d'identifier les unités servant à désigner un objet de discours, trouve ici un complément utile pour l'exploration du cotexte où ces formes apparaissent. Dans l'exemple ci-dessous, nous avons grisé les formes utilisées comme unités pour la définition du néonyme Amplia, qui désigne un fonds à formule éligible au CTO (compte-titre ordinaire), au PEA (plan d'épargne en actions), ainsi qu'à l'assurance-vie.

| **Forme** | **Frq. Tot.** | **Frq. Par.** | **Coeff.** |
|---|---|---|---|
| PEA | 45 | 14 | 20 |
| CTO | 57 | 14 | 18 |
| fonds | 101 | 16 | 17 |
| Assurance | 55 | 12 | 15 |
| Vie | 54 | 11 | 14 |
| formule | 28 | 8 | 12 |
| commercialisation | 69 | 9 | 10 |
| Une | 66 | 8 | 8 |
| **Amplia** | 3 | 3 | 7 |
| 10 formes spécifiques (…) | | | |
| **AMPLIA** | 2 | 2 | 5 |
| 39 formes spécifiques (…) | | | |
| **Bonne** | 1 | 1 | 3 |
| **Compréhension** | 1 | 1 | 3 |
| 22 formes spécifiques (…) | | | |
| **Première** | 1 | 1 | 3 |

*Tableau 3 : Formes spécifiques positives de la partie 187*

Commentaire : le tableau ci-dessus représente un extrait des formes spécifiques positives de la partie 187 du corpus (colonne Forme). Le nombre de formes spécifiques séparant les formes originales mentionnées dans le tableau 2 a été indiqué sur une ligne coupant les colonnes. Fréquence totale (Frq. Tot.) correspond à la fréquence de la forme dans l'ensemble du corpus, Fréquence dans la partie (Frq. Par.) correspond à la fréquence observée dans cette partie du corpus, et Coeff. correspond à la puissance de 10 utilisée pour la notation scientifique de la probabilité attribuée par le calcul hypergéométrique à la fréquence d'une forme dans la partie. Par exemple, pour une probabilité de 0,000005, notée 5 E-6, l'exposant « 6 » est reporté dans la colonne « coefficient ». Plus la probabilité est faible, plus la puissance de 10 est élevée et plus la fréquence de la forme est jugée caractéristique de la partie étudiée. Dans le tableau, seuls les résultats du calcul hypergéométrique ayant produit une probabilité inférieure à 5% (0,05 ou 5 E-02) pour la fréquence observée sont affichés. Le coefficient étant positif, il s'agit ici de spécificités positives correspondant à la présence de fréquences plus élevées que celle de la valeur modale. Les spécificités négatives ne sont pas affichées. Le calcul a été réalisé avec le logiciel Lexico 3.

La complémentarité des deux approches (formes originales et formes spécifiques) est illustrée à l'aide de l'exemple donné ci-dessous.

| **Forme** | **Frq. Tot.** | **Fréquence** | **Coeff.** |
|---|---|---|---|
| collectivités | 37 | 13 | 20 |
| transactions | 12 | 8 | 16 |
| e | 24 | 9 | 15 |
| publiques | 26 | 7 | 11 |
| 69 formes spécifiques (…) | | | |
| loisirs | 1 | 1 | 3 |
| 24 formes spécifiques (…) | | | |
| Collectivités | 14 | 2 | 3 |

*Tableau 4. Formes spécifiques positives de la partie 194*

Le néologisme « e-transactions Collectivités publiques », qui correspond au nom d'un service commercialisé par le groupe bancaire, n'avait pas été détecté lors du test des formes originales. De fait, toutes les formes qui le composent sont déjà assez bien représentées dans le reste du corpus. Cependant, le calcul des spécificités pour la partie 194 a mis en exergue leur forte spécificité positive. Le repérage a alors d'autant plus de chances d'être efficace que le néologisme fonctionne dans le texte sélectionné comme un objet de discours, par opposition à la simple mention, déjà signalée à propos des coordonnées spatio-temporelles ou organisationnelles.

## 3. Présentation des résultats et discussion

Nous pouvons récapituler les points acquis dans les deux précédentes sections en proposant, pour le contexte étudié, une définition du néologisme librement inspirée de celle donnée par Alain Rey[21] : il s'agit d'une unité appartenant à un parler d'entreprise, dont la forme signifiante ou la relation signifiant-signifié n'était pas réalisée au stade immédiatement antérieur du code fonctionnel utilisé dans le cadre d'un référentiel donné. Ce code est assimilé au vocabulaire d'une ou de plusieurs séries textuelles chronologiques, constituées à partir d'occurrences textuelles relevant de genres du discours en usage dans la sphère d'échanges où l'activité de l'entreprise se réalise. De cette conception découle le dispositif textométrique destiné au repérage de formes candidates, mais aussi l'évaluation du caractère néologique des unités correspondant à ces formes. En effet, le dispositif de repérage ne permet pas, à lui seul, de vérifier les principaux critères habituellement adoptés, afin d'identifier le caractère néologique d'une unité[22]. La prise en compte de la diachronie, à travers l'évolution du vocabulaire des textes rassemblés dans une série textuelle chronologique, permet d'obtenir des formes candidates, mais elle ne fournit pas encore d'unités néologiques. Afin d'atteindre ce but, le dispositif mis en place doit permettre la vérification des autres critères que sont l'absence de recensement dans les sources lexicologiques et terminologiques de référence, l'instabilité formelle et sémantique de l'unité, ainsi que la perception de sa nouveauté chez les locuteurs[23].

### 3.1. Les néologismes d'un parler d'entreprise

#### 3.1.1. Critères de sélection des néologismes

Le repérage des formes candidates (formes originales et formes spécifiques positives associées à un coefficient élevé pour un texte donné), sert de point de départ à une série de vérifications destinées à sélectionner les unités néologiques. Ces vérifications sont combinées entre elles afin d'étayer le jugement porté sur le caractère néologique de l'unité évaluée. Trois niveaux d'analyse sont pris en compte. Le premier, de caractère pragmatique, traite de manière indirecte le sentiment de nouveauté éprouvé par le locuteur, en le rapportant au référentiel de la situation de communication. Le second critère tient compte du fonctionnement discursif de l'objet désigné par l'unité étudiée.

---

[21] A. REY (1976 : 17) : « *Ainsi, le néologisme est une unité du lexique, mot, lexie ou syntagme, dont la forme signifiante ou la relation signifiant-signifié, caractérisée par un fonctionnement effectif dans un modèle de communication déterminé, n'était pas réalisée au stade immédiatement antérieur du code de la langue.* »

[22] M.-T. CABRÉ (1998 : 252 sqq.).

[23] On sait que L. GUILBERT (1975 : 36-40) a renvoyé dos à dos, avec des accents chomskyens, l'interrogation de la compétence des locuteurs, jugée trop variable, et l'exploration des produits de la performance, jugés incommensurables, au profit de locuteurs-auditeurs idéaux, incarnés par les lexicologues.

Enfin, un dernier critère s'appuie sur la place occupée par l'unité étudiée parmi les strates du parler d'entreprise. Son appartenance à la langue commune, aux terminologies des secteurs d'activité ou à la terminologie de l'entreprise déclenche, pour chaque cas, des investigations différentes dans les ressources disponibles.

Dans son déroulement, l'analyse se place à chacun des trois niveaux en suivant, sans ordre prédéfini, les indications les plus immédiatement accessibles et en cherchant à les valider ou à les rejeter à l'aide des informations tirées des autres critères. Par exemple, le mot « compréhension », qui apparaît dans les tableaux 2 et 3, n'est pas étudié isolément, mais en tant que partie du terme complexe « recueil de bonne compréhension ». L'identification de la position occupée par cette unité dans les strates du parler d'entreprise va de pair avec l'évaluation de la position de l'objet dans le référentiel ainsi que l'observation de son fonctionnement discursif dans le texte étudié. Cette circulation entre les trois critères, réalisée par l'analyse, est difficile à restituer de façon linéaire. C'est pourquoi, pour plus de clarté, nous présentons les trois critères séparément.

*3.1.1.1. Centralité de l'objet désigné et actualisation du référentiel*

L'interrogation de la place occupée par l'objet désigné – à l'aide d'une unité néologique –, dans le référentiel associé à la situation de communication, repose sur l'hypothèse que les producteurs des textes du corpus souscrivent au principe de coopération gricéen, ainsi qu'aux maximes conversationnelles qui lui sont associées[24]. De ce point de vue, les maximes de manière, qui portent sur la façon de dire ce que l'on a à dire (en règle générale « soyez clair » et en particulier, « évitez de tenir des propos obscurs ou ambigus, soyez bref et méthodique »), si tant est qu'elles soient toujours respectées, contraignent l'énonciateur à renseigner son interlocuteur sur un nouvel objet qu'il n'est pas censé connaître. Par conséquent, nous supposons que la prise en charge énonciative de la nouveauté aura d'autant plus de chances d'être explicite, et donc visible, que la place occupée par l'objet en question sera centrale. La caractérisation de cette position, comme centrale ou périphérique, est établie par rapport au référentiel associé à la circulaire. L'activité dont il est question est alors adoptée comme repère. Pour reprendre l'exemple précédent, le « recueil de bonne compréhension » étant étroitement associé à la commercialisation d'un produit financier, et cette forme d'activité étant centrale dans le contexte de production de la circulaire étudiée, la position de cet objet a été jugée comme étant plutôt centrale. À l'inverse, « loisirs », qui apparaît dans le tableau 4, a été jugé périphérique. En effet, le mot renvoie en l'occurrence aux loisirs courants, qui occupent une place secondaire par rapport au service de paiement à distance dont il est question dans la circulaire étudiée.

[24] H.P. GRICE (1975).

*3.1.1.2. Type de fonctionnement discursif de l'objet de discours et prise en charge de la nouveauté*

Le second critère tient compte du fonctionnement discursif de l'objet de discours auquel le néologisme candidat a fourni une expression nominale. L'évaluation porte alors sur la richesse des informations contenues dans la classe-objet ouverte par l'objet de discours. Cependant, ce n'est pas tant l'ouverture d'une classe qui importe ici que le fait qu'une telle ouverture pourrait difficilement passer sous silence la nouveauté de l'objet lorsqu'elle est avérée. Dans la schématisation qui est proposée par le discours d'une circulaire, certains objets peuvent n'être que mentionnés, alors que d'autres ouvrent une classe-objet englobant divers aspects de l'objet évoqué. Ces classes se composent d'ingrédients et d'objets qui, à leur tour, sont susceptibles d'ouvrir de nouvelles classes. Cette description du fonctionnement discursif, telle qu'elle est proposée par la logique naturelle, doit être articulée avec les dispositifs textuels destinés à prendre en charge la nouveauté, ainsi que les unités concernées. Celles-ci sont susceptibles d'être reformulées à l'aide des mécanismes de la cohésion textuelle, telle que l'anaphore. Lorsque l'unité candidate initie une classe-objet dans un texte de circulaire, les informations qui lui sont associées permettent, d'une part, de valider ou d'invalider le jugement de néologisme (présence d'expressions telles que « nouveau produit », « nouvelle activité »), et d'autre part, de disposer des matériaux utiles pour la construction d'une définition ou d'une description. En revanche, une simple mention ne permet pas de trancher et suspend la qualification du néologisme aux résultats fournis par les autres critères. Par ailleurs, les unités dont les occurrences sont signalées par des signes typographiques, tels que parenthèses ou guillemets, ou bien celles qui sont associées à une périphrase descriptive, connaissent un traitement destiné à faire le départ entre une prise de distance de l'énonciateur et un signalement de la nouveauté, réalisé à l'aide d'une citation ou de tout autre procédé.

*3.1.1.3. Strates du parler d'entreprise et choix des normes de référence*

Enfin, le troisième critère s'appuie sur la position occupée par l'unité candidate dans les strates du parler d'entreprise. Le rattachement de l'unité à l'une de ces strates entraîne des vérifications différentes. Les tableaux 2, 3 et 4 fournissent des exemples d'unités appartenant à chacune de ces strates : « Amplia » est une dénomination propre de produit forgée par l'entreprise ; « loisirs » correspond au sens courant qu'il a dans la langue commune ; « recueil de bonne compréhension » est une unité terminologique relevant du secteur de la conformité et appliquée à la commercialisation des produits bancaires, financiers ou d'assurance. Lorsqu'une unité telle qu'« Amplia » répond positivement aux différents critères (forme originale, objet occupant une position centrale dans le référentiel de la circulaire et fonctionnement discursif sous la forme d'une classe-objet), son appartenance à la terminologie de l'entreprise est jugée suffisamment établie pour l'attribution de la qualité de

néologisme. En revanche, « loisirs », dont on a vu plus haut qu'il occupe une place périphérique par rapport au référentiel, appartient sans ambiguïté à la langue commune (« *Par métonymie. Activité, distraction à laquelle on se consacre pendant les loisirs*.[25] »). Comme par ailleurs cet hapax n'est pas utilisé pour ouvrir une classe-objet, il semble aller de soi que son originalité doit être attribuée à sa rareté dans la série textuelle, compte tenu des sélections appelées par le genre du discours auquel les circulaires appartiennent. La distinction qui doit être opérée entre rareté et nouveauté s'applique systématiquement pour les unités néologiques relevant de la terminologie des secteurs d'activité comme, par exemple, le « recueil de bonne compréhension » dont il a déjà été question. Elle nécessite d'étendre dans deux directions le périmètre des normes de référence.

Une première extension consiste à accroître de quelques années la période prise en compte pour les circulaires (il s'agit de la période allant de 2005 à 2010), et à élargir le champ de la norme générique, afin de prendre en compte des corpus rassemblant des textes appartenant à des genres du discours voisins, produits par le même locuteur collectif. Cela permet de s'assurer du caractère néologique du terme à partir de données textuelles plus étendues que le corpus utilisé pour le repérage. La seconde extension prend en compte les ressources terminologiques disponibles, internes et externes, afin de distinguer les néologismes à proprement parler des simples variantes de termes existants. Ces vérifications peuvent ensuite donner lieu à deux sortes d'arbitrages. Dans le premier cas, lorsqu'un néonyme candidat est bien attesté dans les corpus textuels de vérification, il n'est pas enregistré, car son implantation dans le parler d'entreprise est jugée suffisante, quel que soit son statut dans les ressources terminologiques existantes. En revanche, dans le cas inverse où le candidat est absent des corpus de vérification, mais présent dans les ressources terminologiques, il est considéré comme un néologisme pour le code fonctionnel en usage dans le référentiel pris pour cadre. Dans ce cas, il y a une sorte de « prime » au statut cognitif incertain du terme pour le contexte étudié. Le même raisonnement est appliqué lorsque les néonymes candidats sont absents des deux sources complémentaires de vérification.
Nous illustrons ce processus de qualification, qui fait passer par étapes une unité du statut de forme ou de groupe de formes à celui de néologisme, à l'aide de la dénomination « e-transactions Collectivités publiques », présentée dans le commentaire du tableau 4. La forme « transactions » relève *a priori* de la langue commune où elle bénéficie déjà d'une certaine polysémie : acte par lequel on transige, l'accord qui en résulte, accord contractuel, opération commerciale ou boursière, etc., sont mentionnés dans le *Trésor de la Langue Française*. Mais la présence d'un « e » parmi les formes spécifiques et un début d'observation co-textuelle montrent qu'il s'agit de l'unité composée « e-transactions ». Celle-ci est déjà présente dans le corpus et, par ailleurs, elle fait l'objet d'un recensement, depuis 2006, dans le *Grand dictionnaire terminologique* de l'Office Québécois de la Langue Française. Pour celui-ci, « e-transaction » est à proscrire (il s'agit d'un

[25] *Trésor de la langue française*, article « Loisir ».

terme déconseillé), et la vedette « transaction électronique », dotée de cinq concurrents, y est définie comme une « transaction sécurisée qui est effectuée, lors d'un achat ou d'un paiement en ligne, par l'intermédiaire du réseau Internet. » Dans le corpus de repérage, la nouveauté correspond à la dénomination d'un service, « e-transactions Collectivités publiques », défini comme une « offre de paiement à distance sécurisé, effectué par carte bancaire et à destination du marché des collectivités publiques ». Absente des corpus de vérification, cette nouvelle dénomination a été enregistrée comme néologisme.

3.1.2. Résultats obtenus

Les tests réalisés à l'aide du dispositif de type textométrique qui a été présenté plus haut ont permis de sélectionner des formes candidates en raison de leur originalité ou de leur fréquence plus élevée dans chaque nouveau texte ajouté à une série textuelle chronologique. Dans l'échantillon de circulaires retenues pour cette étude, 34 néologismes ont été collectés dans 20 textes différents, soit 1,7 néologisme par texte productif, ou 0,6 en moyenne si l'on prend l'ensemble de l'échantillon de 54 textes. Parmi les néologismes collectés, 25 ont pu l'être à partir du seul test des formes originales, soit 75% environ[26], le reliquat l'ayant été à partir du test des formes spécifiques.

Parmi les néologismes recensés, les dénominations propres sont les mieux représentées (18 sur 34). Elles se répartissent en noms de produits, noms d'entités ou d'artefacts. Les 16 termes simples et complexes (ces derniers étant classiquement majoritaires) se répartissent à parts égales entre termes bancaires et financiers appartenant à ces secteurs compris au sens large (recueil de bonne compréhension, prêt pour la modernisation de la restauration, escompte des billets à ordre électroniques, seuil des incidents significatifs, prélèvement SEPA / SEPA Direct Debit / SDD, paiement express sans contact / sans contact, décomptabilisation, opération de refinancement à long terme / LTRO / long term refinancing operation)[27], et termes propres à l'entreprise. La répartition des unités collectées par rapport aux principales strates du parler d'entreprise est la suivante : les deux tiers relèvent directement de la terminologie de l'entreprise (22 néologismes sur 34), un petit tiers de la terminologie des secteurs d'activité, et une unité correspond à un nom de série télévisée à la mode, qui a été associé à la vente d'un produit.

---

[26] Il faut préciser que pour les unités complexes combinant au moins deux formes graphiques, nous avons compté qu'une forme originale a permis l'identification d'une unité complexe inclusive en examinant le cotexte à l'aide de concordances. Le caractère néologique a été évalué pour l'unité complexe prise comme un tout.

[27] Dans certains cas, la nouveauté réside dans l'utilisation d'un terme anglo-saxon ou français, là où seule l'une de ces deux langues était déjà attestée. Dans ce cas précis, la durée inhabituelle du refinancement (trois ans au lieu de trois mois), a introduit un flottement entre le développé habituellement retenu pour LTRO, « longer term », et celui qui a semblé plus approprié à la nouveauté « long term ».

Ces résultats sont à rapporter au périmètre restreint qui leur a servi de cadre et qui résulte de la combinaison d'une série textuelle chronologique avec un référentiel situé. Les instances discursives rassemblées dans la série textuelle sont représentatives des échanges validés – par opposition aux mises au point effectuées dans les groupes de travail – entre un organe central et une partie au moins des entités qui lui sont rattachées. Les circulaires officialisent des informations, des recommandations ou des instructions et, à ce titre, elles constituent une source pertinente pour le repérage des termes vedettes, validés par les instances de direction de l'entreprise. Dans la perspective d'une collecte du parler d'entreprise qui distinguerait des sous-groupes d'émetteurs au sein de l'organisation, les apports d'un tel corpus gagneraient à être complétés, d'une part, à l'aide d'autres sources moins formelles, afin d'enregistrer les variantes ou les expressions parallèles propres aux métiers concernés, et d'autre part, à partir des travaux préparatoires, qui comportent fréquemment les expressions alternatives utilisées par les premiers initiés pour désigner un même objet (nom de code, surnom, noms d'essais ou provisoires, etc.). Si un tel approfondissement de la collecte était envisagé, il serait probablement nécessaire de limiter son périmètre à tel ou tel secteur précis de l'entreprise.

## 3.2. L'approche textométrique : un filet adapté mais devant être utilisé avec précaution

### 3.2.1. Les avantages de la méthode utilisée

L'approche textométrique a été retenue dans la mesure où les textes rassemblés en séries textuelles chronologiques constituent la matière première utilisée pour le repérage et l'établissement des néologismes. En effet, nous avons vu qu'ils jouent à la fois le rôle de norme de référence pour les formes candidates et celui de contexte d'apparition des unités néologiques. Lorsqu'une unité nouvelle est utilisée pour désigner un objet de discours, le contexte fournit, pour les définitions, les caractéristiques du concept, et pour les descriptions, un premier ensemble d'informations de nature encyclopédique.

Par ailleurs, la segmentation du texte en formes graphiques brutes, sans étiquetage morphosyntaxique, présente des avantages pour la collecte des unités nouvelles. De cette manière, non seulement certains usages différenciateurs pour les néologismes sont conservés (orthographe particulière des majuscules initiales, mots en capitales, préférence pour les pluriels, dénominations propres déguisées en syntagmes lexicalisés), mais aussi, l'attribution d'étiquettes erronées pour le marquage des parties du discours ne vient pas perturber l'interprétation. La robustesse de cette technique permet en outre de passer d'une langue à l'autre, sans contrainte particulière. Elle assure également une très bonne portabilité d'un corpus à l'autre, les traitements reposant sur l'exploitation de données similaires.

Enfin, les logiciels de textométrie fournissent un outillage déterminant pour les deux principales étapes de l'identification des néologismes : l'identification de

formes candidates dans un cadre de micro-diachronie, ainsi que l'établissement du caractère néologique des unités étudiées. Dans le premier cas, les outils statistiques permettent de disposer d'une norme endogène pour des données textuelles représentatives d'un genre du discours. Depuis cette approche macroscopique, il est ensuite possible de descendre les paliers textuels successifs allant du groupement de textes d'une même série à un texte particulier, des sections d'un texte, tels que les paragraphes, au cotexte d'une occurrence de forme graphique. Cette navigation facilite grandement le travail du néologue, qui a besoin, pour étayer son jugement, aussi bien d'indications globales que d'attestations précises.

3.2.2. Limites de l'exploitation statistique des données brutes

La manipulation des formes graphiques brutes présente des avantages qui ont été rappelés plus haut, mais elle doit être utilisée avec précaution, aussi bien pour le repérage des formes originales que pour le calcul des formes spécifiques. Dans le premier cas, étant donné que le comptage des formes est réalisé en distinguant les variantes graphiques, un même mot est susceptible d'être compté plusieurs fois, comme dans l'exemple donné plus haut avec la dénomination « Amplia », dont les cinq occurrences sont réparties sur deux graphies (tableau 2). Il est raisonnable de penser que pour les formes originales cet inconvénient est limité, dans la mesure où il y a de bonnes chances pour qu'au moins l'une des graphies soit signalée par le test. Le repérage des variantes graphiques est plus délicat avec le calcul des spécificités. Dans le tableau 4, les deux formes du mot « collectivités » ont certes été sélectionnées toutes les deux par le calcul, mais elles sont séparées par quatre-vingt quatorze formes spécifiques positives. Cet inconvénient peut être surmonté par une lecture attentive de la liste et en procédant à des regroupements. En revanche, le calcul hypergéométrique utilisé pour la sélection des formes spécifiques peut, dans certains cas, ne pas signaler des formes qu'il aurait été utile d'obtenir pour le repérage des néologismes.

En effet, Pierre Lafon[28] estime que le calcul hypergéométrique est particulièrement bien adapté pour toutes les fréquences, à quelques exceptions près. Il attire l'attention sur le fait que ce calcul peut ne plus être utile pour des fréquences inférieures ou égales à trois. Or, nous avons vu qu'il est courant de rencontrer des formes candidates dotées de faibles fréquences. Dans le cas des hapax, le calcul devient alors directement dépendant de la taille de la partie du corpus où ils apparaissent.[29] Il est par conséquent nécessaire de tenir compte des caractéristiques des corpus utilisés, afin de déterminer un seuil de probabilité qui ne soit ni trop restrictif, ni trop permissif. En effet, dans le premier cas, il est possible de passer à côté de formes candidates pertinentes, dans l'autre, un résultat trop bruyant risque d'être difficile à exploiter. Les

[28] P. LAFON (1980 : 127-165).
[29] L. LEBART, A. SALEM (1994 : 177).

corpus que nous utilisons pour le repérage des néologismes, du fait de leur organisation sous la forme de séries textuelles chronologiques, sont composés de très nombreuses parties possédant des caractéristiques textométriques homogènes. À cette condition, les hapax (formes originales de fréquence 1), sont signalés comme spécifiques, y compris lorsque le seuil de probabilité est fixé à 5%.

Néanmoins, le calcul hypergéométrique peut également être pris en défaut dans ce type de corpus textuel, lorsque les occurrences d'une forme graphique sont organisées sous la forme d'une progression ou d'un reflux étalés sur plusieurs parties. Dans ce cas, le jugement statistique a tendance à banaliser les fréquences intermédiaires présentes dans les parties situées entre les deux extrémités de la série. Mais cette restriction tombe lorsque les parties étudiées sont la première ou la dernière de la série. En effet, dans ces deux cas, la comparaison des vocabulaires oppose l'une ou l'autre de ces parties à l'ensemble du corpus. Cela permet de tester une fréquence particulièrement faible ou élevée par rapport à toutes les autres valeurs prises par la fréquence d'une forme donnée dans le reste du corpus[30]. Dans l'approche que nous avons retenue, c'est toujours la dernière partie ajoutée au corpus qui est prise en compte, puisque c'est elle dont il faut tester le vocabulaire pour détecter les formes candidates. Il n'en reste pas moins que le non respect des précautions qui viennent d'être exposées pourrait laisser échapper quelques papillons, que nous aurions cru attraper avec le filet textométrique.

## Conclusion

Dans cette étude nous avons adopté une approche résolument pragmatique de la nouveauté, pour laquelle il a été nécessaire de spécifier le caractère relatif de la notion de néologisme lorsqu'elle est appliquée au sociolecte caractérisant une entreprise. Dans ce but, nous avons proposé d'évaluer la nouveauté d'une unité par rapport au vocabulaire habituellement utilisé pour la production de discours relevant d'un même genre et organisés en corpus, sous la forme de séries textuelles chronologiques regroupant des textes aux propriétés homogènes. Un dispositif textométrique s'appuyant sur deux méthodes éprouvées (calcul des formes originales et des formes spécifiques), permet de faire émerger, par contraste avec la norme que représente le vocabulaire de la série textuelle, des formes graphiques nouvelles ou particulièrement caractéristiques pour chaque texte nouveau ajouté au corpus. Ces formes candidates sont alors utilisées pour repérer les unités néologiques, qui peuvent être de différentes natures : expressions nominales simples ou complexes, terminologiques ou non, unités brachygraphiques et dénominations propres. Pour chacune de ces unités candidates, nous procédons à des vérifications au cours desquelles les niveaux pragmatique, discursif et linguistique sont interrogés. En effet, nous cherchons

[30] Pour une explication détaillée et des propositions de solution pour traiter cette question, voir A. SALEM (1988 : 122-129).

à identifier la prise en charge discursive de leur nouveauté au regard, d'une part, de l'importance de l'objet qu'elles désignent dans le référentiel servant de cadre à la communication et, d'autre part, du traitement de cet objet dans les discours. Par ailleurs, nous évaluons le caractère avéré de leur nouveauté par rapport aux normes de référence disponibles pour la strate du parler d'entreprise dont elles relèvent (lexiques de la langue commune, terminologies des langues spécialisées, corpus faisant usage de la terminologie d'entreprise). Une telle approche permet de documenter de façon opérationnelle l'apparition des néologismes au fil des discours. Cependant, elle repose sur des choix qui sont de nature à limiter sa portée, au moins dans un premier temps. D'une part, elle suppose que l'on puisse constituer dans tous les cas les séries textuelles chronologiques adaptées au besoin et, d'autre part, elle pose la question d'une mise en commun des fruits des collectes réalisées à partir de corpus distincts. C'est pourquoi, pour ce qui concerne ce dernier point, nous avons proposé ailleurs de consolider les résultats obtenus à partir de telles collectes au sein de référentiels terminologiques adaptables au contexte.

## Bibliographie

BAKHTINE M. (1984), *Esthétique de la création verbale*, traduit du russe par A. Aucouturier, Paris, Gallimard.

BORZEIX A., FRAENKEL B. (dirs.) (2005), *Langage et travail – Communication, cognition, action*, Paris, CNRS (1ère éd. : 2001).

CABRÉ M. T. (1998), *La Terminologie – Théorie, méthode et applications*, traduit du catalan et adapté par M. C. Cormier et J. Humbley, Paris / Ottawa, Armand Colin / Presses de l'Université d'Ottawa.

DROUIN P., PAQUIN A., MÉNARD N. (2006), Extraction semi-automatique des néologismes dans la terminologie du terrorisme », In VIPREY J.-M. (dir.), *Actes des 8es Journées internationales d'analyse statistique des données textuelles*, 279-288, Besançon, Presses Universitaires de Franche-Comté.

ERLOS F. (2009), *Discours d'entreprise et organisation de l'information – Apports de la textométrie dans la construction de référentiels terminologiques adaptables au contexte*, Thèse préparée sous la dir. d'André Salem, Université Paris 3.

GONSETH F. (1975), *Le Référentiel univers obligé de médiatisation*, Lausanne, L'Âge d'Homme.

GRICE H. P. (1975), Logic and Conversation, In COLE P., MORGAN J. L. (eds.), *Syntax and Semantics 3 : Speech Acts*, 41-58, New York, Academic Press.

GRIZE J.-B. (1996), *Logique naturelle et communications*, Paris, Presses Universitaires de France.

GUILBERT L. (1975), *La Créativité lexicale*, Paris, Larousse.

HABERT B., NAZARENKO A., SALEM A. (1997), *Les linguistiques de corpus*, Paris, Armand Colin / Masson.

HUMBLEY J. (2006), Terminologie et nom propre, In BLAMPAIN D., BRACOPS M. (dirs.), *Des arbres et des mots – Hommage à Daniel Blampain*, 107-124, Bruxelles, Éditions du Hasard.

KOCOUREK R. (1991), *La Langue française de la technique et de la science – Vers une linguistique de la langue savante*, Wiesbaden, Brandstetter.
LAFON P. (1980), Sur la variabilité des formes dans un corpus, *Mots*, Vol. 1, 127-165.
LEBART L., SALEM A. (1994), *Statistique textuelle*, Paris, Dunod.
PEARSON J. (1998), *Terms in context*, Amsterdam, Benjamins.
PICTON A. (2009), *Diachronie en langue de spécialité. Définition d'une méthode linguistique outillée pour repérer l'évolution des connaissances en corpus. Un exemple appliqué au domaine spatial*, Thèse préparée sous la dir. d'Anne Condamines, Université Toulouse 2.
REY A. (1976), Néologisme : un pseudo-concept ?, *Cahiers de lexicologie*, Vol. 28 (1), 3-17.
REY A. (1992), *La Terminologie. Noms et notions*, Paris, Presses Universitaires de France.
SALEM A. (1988), Approches du temps lexical – Statistique textuelle et séries chronologiques, *Mots,* Vol. 17, 105-143.
SITRI F. (1998), *Un modèle d'objet de discours dialogique, entre thématisation et reprise*, Thèse préparée sous la dir. de Sophie Moirand, Université Paris 3.
TAMBA I., BOSREDON B. (1995), Titres de tableaux et noms propres, In NOAILLY M. (dir.), *Nom propre et nomination, Actes du colloque de Brest, 21-24 avril 1994*, 123-135, Paris, Klincksieck.
VECCHI D. M. de (1999), *La Terminologie en entreprise – Formes d'une singularité lexicale*, Thèse préparée sous la dir. de John Humbley, Université Paris 13.
VECCHI D. M. de (2007), La nomination et son « suivi » : entreprises et pragmaterminologie, *Neologica*, Vol. 1, 51-66.

# Augmenting terminology by crawling new term translation pairs from textual corpora: Some theoretical background and applicational directions

Kyo Kageura[1]

## Summary

In this paper we propose a "terminology-driven" framework for augmenting bilingual terminologies. Since the 1990s, much work has been carried out on bilingual term translation pair extraction from textual corpora. While this has produced many technical advances that can be used in a variety of text-oriented applications, there remains a lacuna, in-between term extraction and terminology construction. The framework I propose in this paper is designed to fill this lacuna from the terminology side. Throughout the paper, I will assume Japanese-English bilingual terminology, although the essential framework should be valid for any language.

## Keywords

terminology, terminological structure, bilingual terms

## 1. Introduction

Language practitioners such as translators constantly face the problem of translating technical terms which are not listed in existing terminological dictionaries. This is a reflection of the fact that the number of terms in many domains is constantly growing, and the revision and augmentation of terminological dictionaries cannot keep up, even though the problem of limited coverage imposed by paper-based dictionaries can now be avoided. While many of what translators regard as "new terms" may not really be new from the point of view of domain experts – or, for that matter, of theoretical linguistics – who are carrying out research on the cutting edge, they are still "new" for many language practitioners such as terminologists or translators. As such, we can postulate this problem language practitioners face as a practical issue related to neology.

As new terms are created not in terminological dictionaries but in texts or in conversation, one reasonable approach to dealing with this issue is to collect up-to-date corpora and extract terms. Since the early 1990s, therefore, much work has been devoted to monolingual and bilingual automatic term extraction

[1] Graduate School of Education, University of Tokyo, kyo@p.u-tokyo.ac.jp

from corpora. Focusing only on research dealing with bilingual term extraction, a substantial amount has been carried out so far, both extraction from parallel corpora and from comparable corpora (Dagan and Church, 1997; Fung and McKeown, 1997; Gaussier, 1998; Chiao and Zweigenbaum, 2002; Kwong et al., 2004; Daille and Morin, 2005; Tonoike et al., 2005; Bernhard, 2006; Robitaille et al., 2006; Tonoike et al., 2006; Mitkov and Corpas, 2008; Saralegi et al., 2008; Lefever et al., 2009; Prochasson et al., 2009; Laroche and Langlais, 2010; Daille and Morin, 2012). In the EU, research into corpus-based term extraction has culminated in a large-scale EU project (TTC, 2012).

While such corpus-based term extraction has shown steady advances, we observe that there still remains a gap between the results obtained by these methods and the requirements of translators. This is reflected in the fact that, from our experience of running an online translation-aid system, terminological resources created by using bilingual term extraction are not favoured by translators. The present paper analyses the issue from translators' point of view, and proposes a method and framework for collecting new bilingual term pairs which will partially fill this gap. Note that some of the arguments remain unpolished, and are based sometimes on oversimplification. As such, the author would like this paper to be used as a starting point for discussion in the hope that it will further vitalise the issue of automatic terminology management.

## 2. Translators, New Terms and Terminological Resources

### 2.1. Status of new terms for translators

To clarify some of the possible characteristics required for useful terminological resources, let us start here by stratifying the "new term-ness" for translators in the process of translation. Roughly speaking, this can be put as follows:

a) terms the translation of which translators cannot find in existing terminological resources;
b) terms the translation of which translators cannot find in relevant and reliable parallel data;
c) terms for which translators can find a translation candidate in corpora although it cannot be confirmed as accurate with available information;
d) terms for which translators cannot find any translation candidates in corpora.

Looking for translation candidates for new terms in this order is important for translators, as using proper technical term translations is critical in making translations meaningful, especially for technical or domain-dependent documents.

For translators in particular and language practitioners in general, new terms are concrete and individual terminological units or *substances* which they face for the first time. On the other hand, researchers focusing on neology tend to address *potential forms* of terms that are likely to be created, or possible *mechanisms* that lead to the creation of new terms (cf. Cabré et al., 2012). The theoretical target of studies dealing with neology is thus to describe forms, formations, functions or roles of what could be accepted as new terms, while translators are concerned first and foremost with actual terms which are new to them or the terminological resources they have. If we extrapolate the theoretical target of the range of terms translators are to deal with, it can probably be described as the *realistic possibility of existence* of terms (Kageura, 2012).

While fully appreciating that this argument involves oversimplified generalisation of the range of studies dealing with neology as well as the acts of translators, this contrast is still useful for characterising the issue of new terms translators and language practitioners face and connecting that to the task of automatic term extraction in general.

### 2.2. Automatic term extraction and terminology

Let us summarise here the basic framework that most automatic term extraction methods adopt. Assuming bilingual term extraction, it can be described as follows:

1. prepare parallel or comparable corpora relevant to the domain;
2. identify word sequences that follow the patterns that are likely to be terms as term candidates, typically by using POS patterns and/or statistical weighting;
3. make correspondences between term candidates in the source and target languages identified in step 2.

The order of proceeding may vary, as well as the methods used for making interlingual correspondences, but this summarisation is sufficient for the immediate purposes of our discussion. An important point here is that the automatic term extraction in this framework is concerned with individual terms and corpora. While the set of extracted terms may constitute (part of) terminologies, as a method this framework bypasses the theoretical level of terminologies and the practical level of terminological dictionaries. As such, the contribution of automatic term recognition as defined in this framework maps smoothly to categories b) and c) of "new term-ness" introduced in section 2.1.

Though standard automatic term recognition is important and potentially useful in addressing the issue of new terms, if we wish to deal with the problems translators face with "new terms" by addressing category a), i.e. by augmenting existing terminological dictionaries, the existing automatic term

recognition methods may not be sufficient because they do not explicitly deal with the level of terminologies and terminological dictionaries. While some in regards terminological dictionaries are an archaic legacy of the paper age, the concept of terminology as a coherent set and of terminological dictionaries as its materialisation have an importance in and of themselves which cannot be dissolved into the relationship between corpora and the terms used in them.

The importance of the coherency of terminological dictionaries was discussed in Kageura and Abekawa (2011). Here we briefly look at the theoretical status of terminology as a set and its relation to individual terms, following the arguments set out by Kageura (2002; 2012). While what are immediately and empirically available to us are individual terms and terminological dictionaries consisting of these individual terms, it is the *concept of terminology* which enables us to talk about terms and identify terms as empirical objects; without the pre-existing concept of "terminology," lexical items cannot be perceived as terms. Dividing terms and terminology on the one hand and conceptual space and empirical space on the other, we can visualise this situation as shown in Figure 1 (adapted from Figure 1.3 in Kageura (2012)), in which the empirical path starts from individual terms we actually observe, which is in turn supported by the empirical path that flows from the concept of terminology.

*Figure 1. The relationship between terminology and terms*

This implies that having terminologies and terminological dictionaries in mind as a target of new term collection can have a theoretical status different from that which standard automatic term recognition research has had in mind (cf. Wilks et al., 1996). From the application point of view, targeting terminological dictionaries matches the stratified classification of "new term-ness" perceived by translators. Upon this understanding, we can postulate a hypothesis concerning the range of new terms:

The range of new terms to be created is controlled and bound by the structure and the structural dynamics of *existing terminology*.

Note that the range of new terms estimated on the basis of the modelling of the structural dynamics of existing terminology as a set for each target domain

should be one step more concrete than the description of possible forms of new terms; it corresponds to the concept of the *realistic possibility of existence*.

Under this hypothesis, we can pose two questions:

1. Is it possible to establish a dynamic model which enables us to estimate the realistically possible range of terms in a given terminology, by using the information contained in the terminology itself?
2. Is it possible to take advantage of such a model in some automatic methods to enhance and augment existing terminologies?

In the sections that ensue, I would like to answer these questions positively and demonstrate the basic ideas and directions necessary to do so, although full technical explorations are yet to be completed.

## 3. Modelling the growth of terminologies

Let us start from the description of term formation patterns or patterns of complex terms, which are most commonly exploited in automatic term recognition due to the high percentage of complex terms in most terminologies in most languages and the importance of complex terms in the creation of new terms (Ishii, 1987; Sager, 1990; Cerbah, 2000). Term formation patterns are most typically described by means of the features of constituent elements and possibly the type of relationships among constituent elements. POS is the most frequently used feature of constituent elements, though semantic or conceptual types are also sometimes used. Figure 2 shows this framework (adapted from Figure 6.1 of Kageura (2012)). Note that the descriptions in this framework provide the possible form of complex terms, i.e. well-formedness.

*Figure 2. Basic framework of describing term formation patterns*

In terminology, the creation of new terms is bound not only by syntagmatic regularities but also by paradigmatic or conceptual systematicity which comes from the subsystem of terminology. For instance, the term "morphological analysis" is not only created because this formation is suitable for representing the concept of <morphological analysis>, but also because there are other, parallel, constructions such as "syntactic analysis" or "discourse analysis" in the same terminology. To take into account the terminological/paradigmatic factor in term formation, we can simply put the framework illustrated in Figure 2 into

the terminological structure, as shown in Figure 3 (from Figure 6.1 of Kageura (2012)), in which "subsystem" indicates conceptual/terminological subsystem.

*Figure 3. Framework of describing terminological formation*

To transform the target of description from possible forms to the range of possible new terms which have the realistic possibility of existence, it is necessary to increase the granularity of description. This can, as a first approximation, be simply achieved by using individual constituent elements as separate features, i.e. instead of introducing some grammatical or conceptual features that characterise constituent elements, we can just regard constituent elements as the most fine-grained features to characterise the elements. This is illustrated in Figure 4, in which CE indicates constituent elements.

*Figure 4. Framework of describing terminological formation with concrete granularity*

Taking this as a point of departure and taking into account the currently available mathematical and descriptive methodologies, the descriptive study of terminological formation can be pursued in the following two directions:

1. Describing how existing constituent elements are used and how new constituent elements are introduced in constructing new terms. This can most simply be carried out by removing all the relations between constituent elements and reducing the terminological space to a bag-of-morphemes. Then we can resort to statistical methods of estimating

the growth of constituent elements (Baayen, 2001). Kageura (2002), adopting this approach in analysing and modelling the patterns of terminological growth, showed that this approach is useful and valid for estimating the range of constituent elements used in new terms which can realistically exist.

2. Describing how the different status of existing constituent elements affects their use in the formation of new terms. This can be achieved by (a) defining the terminological network in which terms constitute vertices and the existence of the common constituent elements constitutes edges as well as by (b) defining the network of constituent elements in which constituent elements constitute vertices and the co-occurrence of constituent elements in a term constitutes edges. By analysing these networks, we can observe how the different status of terms or morphemes affects the creation of new terms related to these terms or by using these morphemes. Kageura (2012) reports the result of the embryonic enterprise in this direction, although it falls short of fully exploring this perspective to reveal factors affecting term formation in the totality of terminology.

As a theoretical enterprise, these two directions should ultimately be integrated. In order to do so it is necessary to introduce a fine-grained model which describes and estimates how constituent elements are combined in individual terms. Though theoretically possible, this task is expected to be difficult to carry out due to the predominance of constituent elements which only occur with very low frequency.

On the application front, on the other hand, to what extent we can exploit the information provided by terminological structure in enhancing and augmenting existing terminologies is an intriguing and important task, which will complement the conventional automatic term extraction approach by addressing level a) of the "new term-ness" as perceived by translators and language practitioners. Let us now turn to this issue.

## 4. Terminology-driven crawling of new term translation pairs

In the previous section we explicitly introduced the terminological system or dynamics as a binding factor in term formation and showed possible ways of exploiting this idea to model terminological formation or the growth of terminology. In this section, we will sketch a general framework for exploiting the same idea in order to enhance and augment existing terminologies. Before illustrating the methodological framework and basic ideas behind it, we will first clarify the position of standard automatic term recognition within the social arrangement of terminological dictionaries and corpora.

### 4.1. Terminology, textual corpora and automatic term recognition

Now that we have seen the conceptual primacy of terminologies in relation to individual terms, we can clarify the relationship between terminology and domain corpora or between the terminological sphere and the textual sphere. Roughly, it can be illustrated as in Figure 5. On the one hand, we have domain texts which come from utterances made by domain experts, subject specialists or language practitioners and recorded and distributed with some sort of social filtering. If we observe the process of producing texts, we have general language lexicon and grammar that support the utterances in general and terminological dictionaries that support specialised utterances. Admitting fully that this is an oversimplification of the reality surrounding domain corpora and terminology, it suffices for the immediate purpose of clarifying the position of automatic term recognition and contrasting it with what we call "terminology-driven crawling," which we propose and illustrate below.

*Figure 5. The relationship between terminology and texts*

The standard process of automatic term recognition can be illustrated as in Figure 6. Note that here the main driving factors in identifying terms or term pairs are formal patterns and the calculation of weight or degree of alignment, which are based on information contained in textual corpora. Note that while in the case of bilingual term extraction compositional translation methods are used in some studies, their aim is to make correspondences between candidate terms or to validate target terms, and the identification of term candidates is still carried out by using information obtained from textual corpora (cf. Tonoike, 2005; Tonoike, 2006; Daille and Morin, 2012).

*Figure 6. The standard process of automatic term recognition*

We can superimpose Figure 6 on Figure 5 to show the position and status of the conventional automatic term recognition task within the relationships between the textual sphere and the terminological sphere (Figure 7).

*Figure 7. The position of automatic term recognition between texts and terminology*

Assuming the independent status of terminology – which is supported both by the theoretical argument we gave in section 2.2 and by the hierarchy of "new term-ness" perceived by translators which we set out in section 2.1 – we can identify a lacuna in between conventional automatic terminology recognition and the enhancement and augmentation of terminological dictionaries with new terms. In the next section we will propose a basic framework to fill this lacuna.

### 4.2. Augmenting existing terminologies through terminology-driven crawling: A sketch of a framework

Assuming that we can theoretically estimate the range of new term candidates which can realistically exist by means of exploring the structure and dynamics of existing terminologies, we can define a "generate and validate" framework for augmenting terminological dictionaries, as follows:

1. generate the range of possible new terms based on existing terminologies;
2. validate whether or not each of the term candidates exists in reality by using corpora.

Figure 8 illustrates this framework; we can label the method "terminology-driven crawling."

*Figure 8. The framework of "terminology-driven crawling"*

Research within this framework can pursue the creation of suitable models for estimating and generating potential terms by using information and collecting informative and effective corpora for validation, tasks that we are currently carrying out. Full technical details of our actual methods, implementations and evaluations will be reported separately. Here, we will limit ourselves to illustrating our first, more heuristic approach. The following is one of the simplest practical approaches to generating potential terms. Here we assume a bilingual situation:

1. decompose complex terms in terminologies into constituent elements;
2. establish inter-lingual correspondences between constituent elements;
3. generate head-modifier pairs in one of the language pairs;
4. generate a bipartite graph based on the head-modifier pairs of that language;
5. partition the bipartite graph (to avoid a combinatory explosion in the next step);
6. for each connected component of the bipartite graph, take the direct product of the head and modifier to generate extended head-modifier pairs;
7. for each newly created head-modifier pair, generate a corresponding head-modifier pair for the other language by using the correspondence between constituent elements of the two languages;
8. run each pair through a search engine, either together or separately, and validate the status of the extended head-modifier pairs.

Figure 9 illustrates the steps 4-6, and Figure 10 illustrates steps 7-8.

*Figure 9. Generating extended head-modifier pairs*

*Figure 10. Extending candidates to bilingual pairs and validating them*

## 5. Prospects

Through the observation of the recognition of "new term-ness" by translators as language practitioners and theoretical examination of the status of terminology and terms, we first identified the primary position of terminology as a set. We then clarified the position of conventional automatic term recognition within the relationship between textual corpora and terminologies, and identified a lacuna in the task of enhancing and augmenting existing terminologies or terminological dictionaries. Finally, we briefly sketched a framework within which this lacuna can be addressed, and illustrated a first rough idea for tackling this issue in concrete terms.

Technically, a number of issues remain unaddressed in this paper. These include the following:

1. The approach sketched in section 4.2 can only deal with potential new terms which consist of existing constituent elements. In reality, however, constituent elements which have not been used in terms in existing terminologies may well be used.
2. The approach only deals with head-modifier pairs, i.e. complex terms consisting of two constituent elements. However, there are a non-negligible number of complex terms consisting of three or more constituent elements.
3. The approach only deals with complex terms. However, as in the case of Japanese neology, some new terms are created by means of borrowing rather than compounding.

Some of these issues (1 and 2) can be dealt with in the proposed framework by refining the methods, while others (3) require a different approach. For a certain range of potential new terms, conventional automatic term recognition methods may be more suitable. As such, what we have proposed in this paper constitutes, both theoretically and practically, an approach complementary to, rather than in competition with, existing standpoints, approaches and methodologies. The optimal division of labour between different standpoints and approaches should therefore also be identified, in addition to simply refining separate approaches and individual methods to extract new terms or enhance terminological dictionaries.

## Acknowledgements

This study is partly supported by Challenging Exploratory Research of JSPS Grant-in-aid for scientific research "Automatic augmentation of bilingual terminological dictionaries by means of proactive crawling of candidate term pairs from the web." The "terminology-driven crawling" mechanism described in Section 4 is being developed in cooperation with Dr. Koichi Takeuchi of Okayama University and members of his laboratory, especially Mr. Koichi Sato. I appreciate their competence and enthusiasm in pursuing this research topic. I would like to thank all the people who contributed their time to organising the exciting conference "Neology in Specialized Languages: Detection, Implantation and Circulation of New Terms," especially Dr. Pascaline Dury.

## References

BAAYEN R. H. (2001), *Word Frequency Distributions*, Dordorecht: Kluwer.

BERNHARD D. (2006), "Multilingual term extraction from domain-specific corpora using morphological structure," *EACL 2006*.,171-174.

CABRÉ M. T., BAGOT R. S. and SIERRA C. V. (eds.) (2012), *Neology in Specialized Communication: Special Issue of Terminology*, 18(1).

CERBAH F. (2000), "Exogeneous and endogeneous approaches to semantic categorization of unknown technical terms," *COLING 2000*, 145-151.

CHIAO Y-C and ZWEIGENBAUM P. (2002), "Looking for candidate translational equivalents in specialized, comparable corpora", *COLING 2002*, 1208-1212.

DAILLE B. and MORIN E. (2005), "French-English terminology extraction from comparable corpora", *IJCNLP 2005,* 707-818.

DAILLE B. and MORIN E. (2012), "Revising the compositional method for terminology acquisition from comparable corpora", *COLING 2012,* 1797-1810.

DAGAN I. and CHURCH K. (1997), "Termight: Coordinating humans and machines in bilingual terminology acquisition", *Machine Translation.* 12:89-107.

FUNG P and MCKEOWN K. (1997), "Finding terminology translations from non-parallel corpora", *Proceedings of the 5th Workshop on Very Large Corpora,* 192-202.

GAUSSIER E. (1998), "Flow network models for word alignment and terminology extraction from bilingual corpora", *COLING/ACL 1998*, 444-450.

ISHII M. (1987), "Economy in Japanese scientific terminology," In CZAP H. and GALINSKI C. (eds.) *Terminology and Knowledge Engineering*, Frankfurt: Indeks Verlag, 123-136.

KAGEURA K. (2002), *The Dynamics of Terminology: A Descriptive Theory of Term Formation and Terminological Growth*, Amsterdam: John Benjamins.

KAGEURA K. (2012), *The Quantitative Analysis of the Dynamics and Structure of Terminologies*, Amsterdam: John Benjamins.

KAGEURA K. and ABEKAWA T. (2011), "On the concept of 'comprehensiveness' in information services: The case of the online translation aid and hosting service Minna no Hon'yaku", *Asia-Pacific Conference on Library and Information Education and Practice*.

KWONG O. Y. et al. (2004), "Alignment and extraction of bilingual legal terminology from context profiles", *Terminology*, 10(1):81-99.

LAROCHE A. and LANGLAIS P. (2010), "Revisiting context-based projection methods for term-translation spotting in comparable corpora", *COLING 2010*, 617-625.

LEFEVER E. et al. (2009), "Language-independent bilingual terminology extraction from a multilingual parallel corpus", *EACL 2009*, 496-504.

PROCHASSON E. et al. (2009), "Anchor points for bilingual lexicon extraction from small comparable corpora", *MT Summit XII*.

ROBITAILLE X. et al. (2006), "Compiling French-Japanese terminologies from the web", *EACL 2006*, 225-232.

SAGER J. (1990), *A Practical Course in Terminology Processing*, Amsterdam: John Benjamins.

SARALEGI X., SAN VICENTE I. and GURRUTXAGA A. (2008), "Automatic extraction of bilingual terms from comparable corpora in a popular science domain", *Proceedings of the 1st BUCC Workshop*, 27-32.

TONOIKE M. et al. (2005), "Effect of domain-specific corpus in compositional translation estimation for technical terms", *IJCNLP 2005*, 116-121.

TONOIKE M. et al. (2006), "A comparative study on compositional translation estimation using a domain/topic-specific corpus collected from the web", *Proceedings of the 2nd Web as Corpus Workshop,* 11-18.

TTC – Terminology Extraction, Translation Tools and Comparable Corpora. 2010-2012. http://www.ttc-project.eu/

WILKS Y., SLATOR B. M. and GUTHRIE L. M. (1996), *Electronic Words: Dictionaries, Computers, and Meanings*, Cambridge, Mass: MIT Press.

# Detección, análisis y clasificación de neologismos en corpus especializados: marcadores de creación léxica

Mercedes Roldán Vendrell[1]

## Resumen

La detección, análisis y clasificación de neologismos es uno de los objetivos del proyecto "Olivaterm" ("Terminología del aceite de oliva y comercio: China y otros mercados internacionales"), iniciado en 2008 en la Universidad de Jaén con el objetivo de desarrollar un trabajo terminológico sistemático plurilingüe en el área científica y socioeconómica del olivar y el aceite de oliva. En este trabajo, en primer lugar, explicamos los métodos utilizados en el proyecto "Olivaterm" para la detección de neologismos, con especial atención al análisis de patrones léxico-sintácticos que funcionan como marcadores de creación léxica, aportando datos que demuestran el interés de estos marcadores lingüísticos para el estudio de las relaciones léxico-semánticas que se establecen entre un neologismo y otras unidades del dominio experto al que se incorpora la nueva palabra. En segundo lugar, describimos el proceso de fijación del significado de cada nueva unidad respecto de las demás unidades con las que mantiene relaciones semánticas de distinto tipo y a diferentes niveles en la jerarquía que describe el sistema conceptual del dominio. Por último, analizamos las implicaciones de los resultados de nuestra investigación para la terminología multilingüe, en particular para el establecimiento de equivalencias de las nuevas unidades de conocimiento a otras lenguas y para la elaboración de un modelo definicional satisfactorio desde la perspectiva de la traducción y la comunicación.

## Palabras clave

Terminología multilingüe, traducción, comunicación, detección de neologismos, corpus especializados, marcadores de creación léxica, patrones léxico-sintácticos

[1] Grupo "*Términos: terminología, documentación y comunicación científica*", Universidad de Jaén, mroldan@ujaen.es

## 1. Introducción

La continua aparición de neologismos es una característica fundamental de las lenguas de especialidad que se corresponde con el avance científico y tecnológico de las sociedades que se sirven de ellas. Como sabemos, la función primordial de los neologismos es denominar nuevos conceptos emergentes, aunque la acuñación de palabras nuevas puede deberse también, entre otras razones, a la necesidad de señalar nuevas especificaciones de conceptos ya conocidos en un determinado ámbito de conocimiento, o ser el resultado de la transferencia de términos procedentes de otros ámbitos que adquieren en el dominio receptor un nuevo sentido especializado.
La detección, análisis y clasificación de neologismos es uno de los objetivos del proyecto "Olivaterm"[2], en el que hemos llevado a cabo un trabajo terminológico sistemático y plurilingüe en el área científica y socioeconómica del olivar y el aceite de oliva. El principal resultado de esta investigación ha sido la elaboración de un diccionario especializado que recoge la terminología actual del sector oleícola en lengua española, junto a las equivalencias en dos lenguas clave en el contexto actual del mercado internacional del aceite de oliva: el inglés y el chino.

En este trabajo, en primer lugar, explicamos los métodos utilizados en el proyecto "Olivaterm" para la detección de neologismos, con especial atención al análisis de patrones léxico-sintácticos que funcionan como marcadores de creación léxica, aportando datos que demuestran el interés de estos marcadores lingüísticos para el estudio de las relaciones semánticas que se establecen entre el neologismo y otras unidades del dominio experto al que se incorpora la nueva palabra. En segundo lugar, analizamos el significado de las nuevas unidades y el tipo de relación semántica que ocasiona su incorporación al sistema. Para terminar, presentamos nuestra propuesta de modelo definicional para nuevas unidades ilustrándola con algunos ejemplos.

## 2. Etapas de preliminares de trabajo

Para la detección de unidades de conocimiento especializado de carácter neológico, como para el conjunto de tareas de un proyecto de investigación en terminología, existen dos etapas de trabajo especialmente determinantes, porque de ellas depende en gran medida el desarrollo del resto del proyecto[3].

---

[2]"Olivaterm" (www.olivaterm.com) es un Proyecto de Investigación de Excelencia (P07-HUM-03041) financiado por la Consejería de Innovación, Ciencia y Empresa de la Junta de Andalucía y el Fondo Europeo de Desarrollo Regional (FEDER) bajo el título "Terminología del aceite de oliva y comercio: China y otros mercados internacionales". La investigación, liderada por la Universidad de Jaén, se inició en enero de 2008 y finalizó en diciembre de 2012.

[3] Sobre metodología para la gestión de la terminología, *vid.* M. Roldán Vendrell (en prensa): «Lingüística y gestión terminológica» en Vicente Marcet *et al.* (eds.), *Pro lingua. Nuevas aportaciones al estudio de la lengua*. Salamanca: Editorial Luso Española.

Nos referimos, por una parte, a la elaboración del sistema conceptual del dominio y, por otra, a la elaboración de un corpus de textos especializados en la materia seleccionada. Cuando se trata de terminología multilingüe, es necesario compilar un corpus para cada una de las lenguas objeto de estudio.

## 2.1. Elaboración del sistema conceptual

La representación de la estructura conceptual de un dominio, sus conceptos relevantes, características y relaciones, permite vincular cada término manejado en la especialidad con el concepto específico al que designa y al nivel de la jerarquía de relaciones léxico-semánticas que le corresponde. Los esquemas conceptuales confieren flexibilidad a la representación conceptual y suponen una base sobre la cual los terminólogos pueden modelar dominios especializados. Su aplicación en terminología multilingüe permite además establecer correspondencias interlingüísticas e intertextuales basadas en la misma representación conceptual subyacente (Fáber 2002: 9-12). En el marco del proyecto "Olivaterm", para obtener una representación del conocimiento especializado sobre el aceite de oliva se desarrollaron las siguientes etapas de trabajo:

1. establecimiento de la estructura conceptual del dominio, integrada por los ámbitos de conocimiento específicos vinculados de forma relevante con el área temática acotada;
2. selección de los subdominios más relevantes dentro de cada ámbito específico;
3. elaboración de un árbol de conceptos en el que se establecen los esquemas categoriales de cada ámbito específico, los parámetros de conocimiento que estructuran cada categoría, y la jerarquía de las relaciones conceptuales (conceptos superordinados y subordinados) de cada área semántica.

Desde el punto de vista estructural, diseñamos un sistema amplio (Roldán Vendrell, 2010: 6), en el que se contemplan todos los ámbitos relevantes para el conocimiento especializado de los aceites de oliva. En este sentido, aunque el proyecto "Olivaterm" nació orientado a la vertiente comercial del tema, para llevar a cabo un trabajo terminológico sistemático, era necesario tener en cuenta el carácter multidimensional de la materia y, consecuentemente, recoger términos vinculados a todos los ámbitos que intervienen en los procesos de elaboración y comercialización de los aceites de oliva. El resultado fue un sistema conceptual multidimensional integrado por los siguientes ámbitos: "Agricultura", que agrupa la terminología relativa a la olivicultura; "Industria", al que están vinculados los términos utilizados en la actual elaiotecnia; "Comercio", que recoge los términos relativos a la distribución y

comercialización de aceites de oliva; "Cultura", bajo el que se agrupa la terminología propia de los ámbitos relacionados con el patrimonio cultural y natural surgido en torno al cultivo del olivo y la elaboración de aceites de oliva; "Salud", en el que incluimos las unidades de conocimiento utilizadas en textos que tratan sobre las propiedades, efectos y aplicaciones del aceite de oliva y los productos con él elaborados; y "Aceites", que constituye el núcleo del sistema conceptual del dominio y recoge el léxico utilizado para clasificar los distintos tipos de aceite y para hablar de su composición y características sensoriales:

*Figura 1. Sistema conceptual del área temática del aceite de oliva (Fuente: Olivaterm)*

Los seis ámbitos que componen el sistema conceptual de "Olivaterm", así concebido, son partes del sistema de representación del conocimiento especializado sobre los aceites de oliva entre las que existen múltiples interrelaciones (como se pretende indicar mediante las flechas incorporadas a la figura 1), de manera que las unidades vinculadas de forma prioritaria a un determinado ámbito del sistema pueden aparecer en textos especializados de los demás campos. De este modo, en textos que hablan sobre estudios de mercado del aceite de oliva es frecuente encontrar léxico de industria, de salud y de cata. Lo mismo puede decirse a propósito de textos que tratan cuestiones sobre cata de aceites, en los que los expertos utilizan continuamente unidades vinculadas primordialmente a los subdominios de la nutrición o la tecnología empleada para la obtención de los distintos tipos de aceite.

## 2.2. Elaboración del corpus textual

El punto de partida para llevar a cabo la descripción actualizada del léxico de un dominio es el estudio de textos especializados en la materia[4]. En concreto, la detección de neologismos en los lenguajes de especialidad se basa fundamentalmente en el análisis de corpus de estudio (corpus de textos recientes) y corpus de referencia (conjunto de textos de fechas anteriores) compilados y seleccionados en las fases iniciales del trabajo terminológico sistemático. En el caso que nos ocupa, para poder establecer cuál es la terminología actual del aceite de oliva, utilizada en situaciones comunicativas diversas, seleccionamos distintos tipos de texto que se pueden clasificar en los tres siguientes grupos:

a. obras de referencia relevantes para el área temática acotada, vinculadas a los distintos ámbitos del sistema conceptual;
b. textos científicos de reciente publicación que versan sobre las líneas temáticas establecidas de manera específica para cada subdominio del sistema conceptual;
c. textos de carácter divulgativo y periodístico de reciente publicación, dirigidos, por tanto, al público general, en los que se aborda el tema del aceite de oliva desde registros no científico-técnicos.

Todos ellos son textos especializados por la temática que tratan, pero su nivel científico-técnico varía en virtud de la situación comunicativa a la que responden (especialista-especialista; especialista-público general; divulgación especializada-público general, etc.)[5]. Por otra parte, hay que señalar que el corpus textual del proyecto "Olivaterm" se compone, a su vez, de tres corpora monolingües, uno para cada lengua objeto de estudio. En las tres lenguas se han seleccionado textos de los distintos ámbitos y registros. Atendiendo a todas estas características, podemos clasificar el corpus de estudio de "Olivaterm"[6] como:

---

[4] En el proyecto "Olivaterm", partimos de una concepción de la terminología ecléctica y de base lingüística, en la línea de la Teoría Comunicativa de la Terminología (Cabré 1999, 2002) y de las propuestas de la Socioterminología (Gaudin 2003) para la elaboración de aplicaciones terminográficas a partir de los datos que facilita el discurso especializado producido en situaciones pragmáticas diversas.

5 Para una propuesta de selección de tipos textuales para el trabajo en terminología, *vid.* Siebel y Jiménez Hurtado: "La pragmática de la terminología: en busca del perfil de usuario", en Faber 2002.

[6] El corpus contiene un total de 4 205 773 palabras (858 textos), distribuidas de entre los tres corpus monolingües de la siguiente manera: español, 3 568 141 (626 textos); inglés, 625992 (155 textos); chino, 10640 (77 textos). Para su análisis e identificación, los textos han sido etiquetados según cuatro parámetros: lengua (ES/EN/CN), tipo de texto (a=científico-técnico/b=divulgativo), ámbito (AG, IN, AA, CC, SA, CU), y referencia de la fuente; por ejemplo: ES-a-AG-DJA-15.

a) especializado
b) multilingüe
c) multidisciplinar
d) multinivel

## 3. Metodología para la detección de neologismos

Una vez que disponemos de un corpus de estudio y de un conjunto de textos de referencia, podemos ayudarnos de diversas herramientas de búsqueda para la detección, análisis y clasificación de nuevos términos. Empezaremos por establecer cuál es, en el momento de la investigación, el conjunto de unidades de conocimiento especializado del área acotada, incluyendo en él tanto unidades nuevas como voces ya conocidas.

### 3.1. Fijación de la terminología del área temática en su conjunto

Una relación fiable de candidatos a término y de posibles neologismos se puede alcanzar siguiendo las siguientes etapas de trabajo:

1º extracción automática y manual de candidatos a término del corpus de trabajo de la lengua objeto de estudio;
2º elaboración de una lista de patrones léxico-sintácticos que puedan funcionar como marcadores de creación léxica;
3º extracción automática de información mediante la aplicación de la lista de marcadores de creación léxica al corpus textual;
4º cotejo, combinación y sistematización de los datos obtenidos en las etapas 1-3;
5º propuesta de glosario en la lengua de partida;
6º revisión de la propuesta de glosario por parte de los especialistas: eliminación de candidatos irrelevantes e incorporación de propuestas adicionales;
7º redacción de las definiciones;
8º selección de ejemplos de uso;
9º incorporación de nuevos términos detectados en las etapas 5- 8;
10° fijación del repertorio de términos del dominio seleccionado.

### 3.2. Identificación de neologismos

Una vez establecido el repertorio de términos de una lengua de especialidad en un momento concreto de su historia, se puede abordar el estudio cuáles de esas unidades son neológicas y establecer de qué tipo de neología se trata en cada caso. En los últimos años, se han diseñado muchos recursos para la búsqueda automática de neologismos, pero, hasta en los más recientes, el funcionamiento de base es todavía casi el mismo que el de las herramientas más antiguas

(Janssen 2009: 68), ya que fundamentalmente sólo existen tres métodos diferentes para la detección automática de palabras nuevas:

a) utilizar una lista independiente de palabras conocidas (lista de exclusión);
b) utilizar patrones lingüísticos que funcionan como marcadores de creación léxica;
c) contar las ocurrencias de las palabras, comparando un corpus de estudio con otro de referencia.

Para optimizar los resultados de la investigación, y siempre que se disponga de los recursos necesarios, lo ideal es combinar los tres métodos arriba mencionados. Cada uno de ellos ofrece ventajas y presenta limitaciones, pero de todos se pueden obtener datos de interés. En el proyecto "Olivaterm", hemos trabajado tanto con listas de exclusión, como con patrones lingüísticos y corpus de referencia, incorporando cada método de detección automática al proceso de trabajo en el momento que estimamos más conveniente. Las herramientas de extracción automática de información[7] aplicadas a corpus de estudio especializados aportan datos importantes sobre el comportamiento y características de los términos seleccionados y facilitan una serie de tareas:

1º detección de posibles neologismos entre las unidades integrantes del glosario fijado para la lengua objeto de estudio;
2º localización y documentación de las primeras acuñaciones de los términos en el corpus de estudio;
3º estudio y seguimiento del uso de los términos en el corpus en cuanto a su frecuencia de aparición;
4º selección de candidatos a neologismo;
5º elaboración de una lista de exclusión de términos considerados no neológicos en el momento de la investigación;
6º cotejo de la lista de candidatos a neologismo con la lista de exclusión;
7º revisión final de los datos obtenidos por parte de los especialistas;
8º fijación del repertorio de neologismos;
9º clasificación las unidades neológicas.

Cuando se trata de terminología multilingüe, el método de trabajo aplicado a la lengua prioritaria debe llevarse a cabo asimismo con resto de lenguas objeto de estudio hasta alcanzar, para cada una de ellas, los cuatro objetivos generales siguientes:

[7] Para la extracción automática de información hemos trabajado con el programa ©WordSmith Tools.

1. establecimiento de equivalencias de los términos de la lengua de partida en las lenguas meta mediante el estudio y análisis de los respectivos corpus textuales;
2. elaboración, para cada lengua meta, de una lista de patrones léxico-sintácticos que puedan funcionar como marcadores de creación léxica;
3. establecimiento de la terminología del dominio seleccionado en cada lengua objeto de estudio;
4. identificación, análisis y clasificación de neologismos.

En lo que sigue, comentamos los datos obtenidos mediante la aplicación de patrones léxico-sintácticos para la detección y estudio de unidades neológicas en el léxico español actual de la olivicultura y la elaiotecnia.

## 4. Aplicación de patrones léxico-sintácticos para la detección de términos y la identificación de neologismos

Como hemos indicado en 3.1., una de las tareas encaminadas a delimitar el repertorio de términos de una lengua de especialidad en un momento dado de su historia es la elaboración de una lista de los patrones léxico-sintácticos a partir de los cuales iniciar una serie de búsquedas. En el proyecto "Olivaterm" establecimos la siguiente lista de marcadores lingüísticos cuya función es introducir unidades terminológicas en el discurso:

| PATRONES LÉXICO-SINTÁCTICOS | |
|---|---|
| 1. *a partir de ahora* | 18. *que podemos llamar* |
| 2. *conocido/-a/-os/-as como* | 19. *que podríamos llamar* |
| 3. *como son* | 20. *que se llama/-an* |
| 4. *de ahora en adelante* | 21. *llamado/-a/-os/as* |
| 5. *de los que existen las/los siguientes clases/tipos* | 22. *mientras que el/la* |
| | 23. *por ejemplo* |
| 6. *denominado/-a/-os/-as* | 24. *por último el/la* |
| 7. *el/primero/segundo/tercero* | 25. *que recibe/-en el nombre de* |
| 8. *en concreto el/la/los/las* | 26. *que se conoce/-en como* |
| 9. *en el caso del/de la/los/las* | 27. *la primera/segunda/tercera* |
| 10. *entre los más importantes* | 28. *se considera/-an* |
| 11. *entre los/las que se encuentra/-an* | 29. *se considerará/-rán* |
| 12. *entre otros* | 30. *seguido/-a del/de la* |
| 13. *hay que tener en cuenta el/las/los/las* | 31. *también denominado/-a/-os/-as* |
| | 32. *también llamado/-a/-os/-as* |
| 14. *por citar alguno/-os* | 33. *tanto el/la....como el/la...* |
| 15. *que hemos dado en llamar* | 34. *un primero/segundo/tercero* |
| 16. *que hemos denominado* | 35. *(...)* |
| 17. *que nosotros denominamos* | |

## 5. Marcadores de creación léxica y relaciones semánticas

El significado mismo de cada uno de estos marcadores discursivos nos da la pauta del tipo de relación semántica que existe entre el término que introducen y otras unidades del sistema ya conocidas (bien porque ya han sido previamente mencionadas en el texto, bien porque son fácilmente recuperables en el contexto de la enunciación). Dicho de otro modo, cada marcador de creación léxica está especializado en un tipo específico de relación semántica. En virtud de este parámetro, se puede hablar de *marcadores de hiperonimia, de sinonimia y de antonimia*. En la siguiente tabla, presentamos los principales marcadores de creación léxica asociados a estos tres tipos de relaciones:

| RELACIONES SEMÁNTICAS | MARCADORES |
|---|---|
| hiperonimia | entre los más importantes<br>entre otros<br>por ejemplo<br>como son<br>entre las que se encuentra/-an<br>entre los que se encuentra/-an<br>por último el<br>seguido del/de la<br>seguida del/de la<br>un primero/segundo/tercero |
| sinonimia | también llamado/-os/-a/-as<br>también denominado/-os/-a/-as<br>también conocido/-a como |
| antonimia (complementariedad) | mientras que el/la<br>tanto el/la… como el/la |

Con estos datos a la vista, la aplicación de patrones léxico-sintácticos puede ser un método útil no solo para la detección de términos y la identificación de neologismos, sino también para el estudio de los procesos de fijación del significado de cada nueva unidad respecto de otras con las que mantiene relaciones léxicas de distinto tipo y a diferentes niveles en la jerarquía que describe el sistema conceptual del dominio.

## 6. Detección, análisis y clasificación de neologismos: nuevos términos, nuevas acepciones, nuevos sinónimos.

En ocasiones, las búsquedas mediante marcadores de creación léxica nos devuelven información sobre nuevas unidades cuyo carácter neológico queda confirmado tras el cotejo con las listas de exclusión. Veamos los datos aportados por una búsqueda a partir del patrón "en tercer lugar":

De los términos detectados en la búsqueda, podemos proponer como candidatos a neologismo los que se relacionan a continuación. De todos ellos, *aceite ocluido* es el único que no había sido detectado en etapas previas de trabajo. Asimismo, en un primer momento desconocíamos a qué hace referencia exactamente el término *aceite normal* en este caso:

| | |
|---|---|
| en tercer lugar | |
| demanda de aceites de oliva | es-a-in-gea-1 |
| sector del aceite de oliva | es-a-in-gea-1 |
| aceite normal | es-a-ag-pao-1 |
| aceite ocluido | es-a-ag-pao-1 |
| gran distribución | es-a-ag-jan-15 |
| calidad | es-a-ag-ex5-90 |

Con la ayuda de herramientas de búsqueda automática, recuperamos el contexto de uso del candidato neologismo:

> Existe un consenso en cuanto a la forma de presentarse el aceite en la pasta. Los diferentes estudios reológicos realizados coinciden en que dentro de la pasta de aceituna molida podemos encontrar de tres formas el aceite. La primera es el llamado **Aceite suelto**, también llamado "**mosto suelto**", que es aquel que se separa de la pasta sin apenas necesidad de presión. Es decir, prácticamente es liberado en la molienda. En segundo lugar podemos encontrar el **Aceite normal** también llamado "**mosto normal**", que es aquel que se separa de la pasta cuando esta se somete a un proceso similar a un filtrado, presión o centrifugación. Este es el que si bien esta en la masa, no está totalmente separado de ella. Es el que presenta más cantidad de aceite. Y en tercer lugar está el **Aceite ocluido** que es la parte del aceite normal que no se separa de la pasta una vez sometida al proceso de presión ó centrifugación. ([ES-a-AG-PAO-1])[8]

[8] Las negritas son nuestras.

La lectura de este texto en el que el término *aceite ocluido* es introducido por el patrón "en tercer lugar" proporciona abundante información. Por una parte, confirma el carácter neológico de esta unidad y permite la detección de otros dos términos con los que *aceite ocluido* mantiene una relación semántica de complementariedad; *aceite suelto* (que podría haber sido detectado asimismo mediante los marcadores "la primera" y "también llamado") y *aceite normal* (introducido por el patrón "en segundo lugar"). Por otra parte, el texto aporta la definición de los tres términos complementarios y hace explícita la variable en virtud de la cual ha sido establecida esta tipología de aceites ("la forma de presentarse el aceite en la pasta"). Asimismo, el texto nos proporciona los sinónimos "mosto suelto" y "mosto normal" para los términos complementarios *aceite suelto* y *aceite normal* respectivamente.

En otras ocasiones, las búsquedas a partir de patrones léxico-semánticos nos revelan datos no sobre nuevos términos, sino sobre nuevas acepciones de términos ya conocidos. Un ejemplo de ello lo encontramos en la aplicación del marcador "mientras que el/la":

La búsqueda nos devuelve los siguientes términos candidatos a neologismo:

```
mientras que el
    aceite de orujo de oliva                          es-a-sa-rma-616-04
    aceite de oliva                                   es-a-sa-rma-603-11
    consumo de aceite de oliva                        es-a-sa-rma-599-11
    precio del aceite                                 es-a-in-int-1
    aceite de oliva puro, aceite de oliva extra virgen es-a-in-int-1
    tiempo de batido                                  es-a-in-ex9-72
    almacenamiento de la pasta                        es-a-in-ex7-35
    polifenoles                                       es-a-in-ex7-30
    almazara, molturan                                es-a-in-ex5-113
    aceite de oliva virgen lampante                   es-a-co-fex-1
    olivo                                             es-a-co-ex5-8
    aceite de semillas                                es-a-ag-upa-1
    sistema por aspersión                             es-a-ag-rma-579-12
    aceituna para almazara, variedades para mesa,
    olivar de regadío, olivar de secano               es-a-ag-fcr-1
```

Todos ellos son unidades ya detectadas con anterioridad en la investigación, a excepción del término *almacenamiento* combinado con el sintagma *de la pasta*. Para analizarlo, recuperamos en el corpus textual el contexto de aparición tras el marcador seleccionado:

> Mediante un análisis de varianza de los datos experimentales se ha observado, utilizando un nivel de confianza del 95%, que los tres factores influyen de forma diferente, siendo la adición de talco el factor que produce un mayor aumento significativo del rendimiento, mientras que el **almacenamiento de la pasta** incide de forma más negativa en la calidad del aceite obtenido al aumentar la acidez y el índice de peróxidos; no obstante, todos los aceites obtenidos se podrían clasificar en la categoría de aceites de oliva virgen extra según la normativa de la Unión Europea. [ES-b-IN-EX7-35]

Asimismo, analizamos otras ocurrencias de esta combinación:

Tras cotejar los datos con la definición elaborada en el marco de la investigación para el término *almacenamiento*, concluimos que se trata de una nueva acepción del término que debe ser añadida a las ya establecidas en referencia al almacenamiento de aceitunas y al almacenamiento de aceite:

> **almacenamiento**
> **1.** s. m. INDUSTRIA. Operación previa a la molienda, durante la cual las aceitunas se guardan en depósitos, trojes o tolvas, a la espera de ser molturadas. *Se obtienen aceites de calidad cuando las aceitunas se trituran lo más rápidamente posible, pues el almacenamiento (atrojado) desencadena procesos fermentativos sobre el fruto*. **Sin.**: atrojado.
> **2.** s. m. INDUSTRIA. Fase posterior a la elaboración que consiste en conservar el aceite, a la espera de su comercialización, en depósitos o trujales y en las condiciones adecuadas para impedir cualquier degradación cualitativa *Los atributos de amargor, astringencia y pungencia pierden intensidad durante el almacenamiento del aceite, ya que disminuye la concentración de los compuestos que los producen*. (Fuente: Olivaterm)

Además de neologismos y nuevas acepciones de términos ya conocidos, la aplicación de patrones léxico-semánticos puede tener como resultado la localización de nuevos sinónimos de términos ya registrados. Veamos el caso del término *morquia*, detectado mediante los marcadores "denominadas" y "denominado":

La búsqueda a partir del marcador de creación léxica "denominadas" dio los siguientes resultados:

```
denominadas
    proteínas                                                es-a-sa-mat-1
    pastas emulsionadas, pastas fluentes, pastas difíciles   es-a-in-gea-1
    plantación intensiva                                     es-a-in-dip-1
    microalmazaras o pequeñas almazaras                      es-a-cc-ex7-28
    grasas animales                                          es-a-ag-upa-1
    alpechín, jamila, morquia                                es-a-ag-pac-1
    nebulización                                             es-a-ag-jan-2
    cultivo compartido o asistido, intensivas y
    superintensivas (refiriéndose a las plantaciones)        es-a-ag-fcr-1
    plagas menores                                           es-a-ag-ex7-83
```

Entre ellos destaca la voz *morquia*, por ser el único término desconocido hasta el momento de esta búsqueda. Lo encontramos también aplicando el patrón "denominado":

```
denominado
    aceite de repaso                                          es-a-sa-rme-616-01
    floral                                                    es-a-sa-jan-19
    aceite de orujo                                           es-a-sa-igr-1
    separadora centrífuga vertical, batidora,
    aceite de oliva lampante                                  es-a-in-rol-108-78
    transesterificación                                       es-a-in-rme-571-17
    orujo seco, orujillo                                      es-a-in-jan-18
    ecológico (refiriéndose al sistema continuo)              es-a-in-jan-11
    sistema continuo                                          es-a-in-gea-1
    separación líquido-líquido, ecológico o de
    dos salidas, alpechín, jamila, morquia,
    extracción parcial, trituradora de discos dentados,
    empiedro, molienda, aceite de oliva, aceite
    de oliva refinado, superintensivo o en seto               es-a-in-gea-1
    envero                                                    es-a-in-ex7-49
    aceite de oliva desautorizado en condiciones de
    baja temperatura denominado aceite desodorizado           es-a-in-ex5-145
    sistema de dos fases                                      es-a-in-coi-2
    stand still                                               es-a-co-par-1
    marqués                                                   es-a-co-ex7-28
    aceite de oliva virgen, aceite de oliva virgen
    extra, aceite de oliva                                    es-a-co-cdo-20
    mercado en destino, mercado en origen                     es-a-co-cdo-20
    dilacerado de la aceituna                                 es-a-ag-pao-1
    orujo                                                     es-a-ag-jan-3
    decánter, centrífuga vertical de sinfín interior          es-a-ag-jan-3
    nueva olivicultura                                        es-a-ag-for-1
    marco dinámico de plantación                              es-a-ag-for-1
    alperujo                                                  es-a-ag-ex7-85
    ordeño                                                    es-a-ac-iag-1
    lampante                                                  es-a-ac-oma-1
    olivar tradicional también denominado menos productivo    es-b-co-dja-2
    termobatido                                               es-b-ac-rme-4
```

Recuperamos dos ocurrencias del término *morquia* en el corpus para analizar su significado en contextos reales de uso y la naturaleza de las relaciones semánticas que mantiene con otras unidades del mismo subámbito en el sistema conceptual:

Esta fase líquida junto con otras aguas de lavado de aceitunas, del aceite, etc. fueron denominadas "alpechín", "jamila" ó "**morquia**" y constituían el efluente total de las almazaras. [ES-b-AG-PAO-1]

En este sistema se produce un efluente líquido denominado «alpechín», «jamila» o «**morquia**», constituido fundamentalmente por partes de las siguientes aguas: de vegetación de las aceitunas, de adición al decánter, de lavado de aceite y de lavado de equipos, este conjunto con el agua de lavado de aceitunas y de instalaciones, constituye el efluente total de la almazara y debido a su constitución orgánica soluble y en suspensión tiene un elevado poder contaminante, por lo que no se puede verter directamente al campo ni a los cauces públicos. [ES-b-IN-GEA-1]

Los textos evidencian que se trata de término sinónimo de *alpechín* y *jamila*, ya registrados en el transcurso de la investigación:

> **alpechín**
> **1.** s. m. INDUSTRIA Residuo líquido, oscuro, fétido y muy contaminante, resultante de la extracción de aceite por los sistemas tradicional y de tres fases, compuesto fundamentalmente de agua de adición y agua de vegetación, además de una fracción mínima de aceite y restos sólidos, y cuya eliminación presenta muchas dificultades. *Dentro de las actividades industriales que representan problemas ambientales merece la pena destacar la generación del alpechín derivado de la producción del aceite de oliva*. **Sin.:** jamila. **2.** s. m. INDUSTRIA. Subproducto líquido resultante de la extracción de aceite por los sistemas tradicional y de tres fases que se utiliza, principalmente, como agua de riego, fuente de energía o fertilizante. *La productividad de los olivos cultivados en suelo tratado con 10 o 30 litros/m2 de alpechín y abonado parcialmente o no abonado, fue satisfactoria tanto en lo referente a las características de las aceitunas como a la calidad del aceite de oliva virgen obtenido.* **3.** s. m. CATA Flavor característico del aceite que ha permanecido en contacto prolongado con las aguas de vegetación. *En función del tiempo de batido puede detectarse como fondo aromático y flavor la sensación a madera, orujo y alpechín.* (Fuente: Olivaterm)

Para terminar, en los dos diagramas siguientes mostramos los datos relativos a *candidatos a término*, *términos ya registrados* y *neologismos* obtenidos mediante de la aplicación al corpus de estudio de los marcadores de creación léxica relacionados en 4:

De acuerdo con estos resultados vemos, por ejemplo, que el marcador de hiperonimia "denomimado" aparece 149 veces en el corpus de estudio, proporcionándonos 41 candidatos a término; de estos, 7 fueron confirmados como unidades ya registradas, mientras que no se identificó ninguna unidad nueva. Por poner otro ejemplo, la ocurrencia del marcador de antonimia "mientras que el" es de 132 apariciones en el corpus, introduciendo 19 candidatos a término; 2 de ellos fueron confirmados como términos y 1 identificado como neologismo.

## 7. La definición de neologismos

La Teoría Comunicativa de la Terminología (Cabré 1999, 2002) defiende que los conceptos no pueden ser unidades de partida, sino que hay que trabajar desde las palabras hacia los conceptos. En este sentido, la definición terminográfica debe establecer la vinculación del término a un concepto determinado, el cual está incardinado en un dominio conceptual específico conformado por una red de relaciones conceptuales (Faber 2002: 19-20). A la hora de ubicar una unidad en el dominio conceptual al que pertenece, el terminógrafo puede recurrir al método de descomposición léxica gradual para llegar desde los lexemas más complejos hasta los más simples y alcanzar así una correcta representación de las relaciones léxico-semánticas que mantienen los términos vinculados a ese dominio. En el proyecto "Olivaterm", hemos tenido en cuenta las propuestas teóricas encaminadas a avanzar en el proceso de codificación del conocimiento especializado en una estructuración definicional coherente con los nuevos enfoques de la terminología (Temmerman 2000;

Faber y Jiménez 2002; García Quesada y Montero Martínez, 2003). De acuerdo con estos presupuestos teóricos, una definición terminográfica debe cumplir dos funciones fundamentales:

a. proporcionar el vínculo entre el concepto y el término, porque mediante la elaboración de la definición se fija la referencia que ese término establece y, al mismo tiempo, se hacen explícitas sus relaciones con otros conceptos dentro del mismo campo;
b. esclarecer el significado de cada unidad respecto del de las demás unidades del sistema conceptual con las que mantiene relaciones semánticas de distinto tipo y a diferentes niveles en la jerarquía que describe el sistema conceptual de un dominio.

Además de estas dos funciones, para elaborar definiciones satisfactorias de los términos vinculados a un ámbito, conviene establecer patrones relacionales recurrentes para todos los términos de una misma categoría, es decir, es necesario partir de los parámetros de conocimiento que definen la categoría conceptual a la que pertenece cada término definido y utilizar la combinación de dichos parámetros como modelo para la definición de todos los términos que pertenecen a dicha categoría. Todo ello nos permitirá discriminar conceptos afines de forma sistemática. Por otra parte, la combinación de los datos descriptivos de cada concepto con los parámetros lingüísticos opositivos establecidos mediante el análisis semántico-estructural de un conjunto de términos presenta la ventaja de facilitar la traducción de definiciones a otras lenguas (Roldán Vendrell y Fernández Domínguez 2012: 18-24; Montoro del Arco y Roldán Vendrell 2013: 83-87). Una vez establecidos de forma inequívoca y sistemática los parámetros de las definiciones, se reduce el riesgo de que en el proceso de traducción se pierda o malinterprete información relevante para una correcta transferencia de conocimiento especializado.

Por lo que respecta al trabajo en el ámbito de la neología, y de acuerdo con todo lo anterior, cuando en un trabajo terminológico sistemático se detecta un neologismo y se desea abordar su definición, ésta debe hacer explícitas las relaciones que vinculan la nueva unidad a su dominio conceptual. Tomemos como ejemplo el neologismo *Shikitita*, detectado mediante la aplicación del patrón "denominada". El término fue creado para denominar una variedad peculiar de olivo y, por extensión, de aceituna y de aceite. Por tanto, debe vincularse preferentemente al ámbito específico de la "agricultura", dentro del sistema conceptual general de "Olivaterm" (figura 1). En él se establecen los niveles jerárquicos de la red de relaciones conceptuales del dominio de la "agricultura", al que cabe incorporar la voz *Sikitita*. Una definición satisfactoria de esta nueva unidad debe hacer explícita su vinculación con el concepto denominado, para fijar la referencia establecida, y, al mismo tiempo, ubicarla en el nivel jerárquico que le corresponde dentro del sistema de conceptos. En este sentido, el primer rasgo definitorio del *Sikitita* es su condición de "variedad";

este rasgo la relaciona con el resto de unidades que conforman la categoría de las "variedades":

*Figura 2. Rama de la categoría "variedades de olivo" del árbol de conceptos de "Agricultura" (Fuente: Olivaterm)*

Para la elaboración de las definiciones de las variedades de olivo/aceituna/aceite, en "Olivaterm" establecimos una serie de parámetros definicionales para esta categoría en los que se contemplan toda la información relevante para los objetivos del trabajo. Por motivos de sistematicidad, coherencia y traductibilidad, las definiciones de las 16 variedades principales de olivo seleccionadas debían contener información relativa a los siguientes parámetros:

| OLIVO | |
|---|---|
| "variedad de olivo…" | |
| **PARAMETROS** | **FORMULAS LINGÜISTICAS** |
| **descripción** | de vigor bajo/ medio/ elevado |
| | de porte erguido/ abierto |
| | de densidad de copa clara/ media/ espesa |
| | de hoja de forma lanceolada/elíptica/elíptico-lanceolada |
| **fortalezas** | resistente a: el frío, la sequía, el repilo, la verticilosis, el desprendimiento, la mosca, la tuberculosis |
| | adaptable/muy adaptable a: suelos calizos, la sequía, terrenos húmedos |
| **debilidades** | altamente susceptible/ susceptible/ tolerante/ resistente/ altamente resistente repilo/ verticilosis/ tuberculosis/ aceituna jabonosa |
| **entrada en producción** | de precocidad media para entrar en producción |
| | de precoz/ tardía/ muy tardía entrada en producción |
| **productividad** | de baja/alta productividad |
| **mecanización** | cuyos frutos presentan una baja/elevada resistencia al desprendimiento, lo que facilita/dificulta su recolección mecanizada |
| **información adicional** | nº x en importancia (superficie x 1000 ha) en España/del mundo |
| | se cultiva/cultivada en las provincias de…/ dominante en la provincia |
| | el término x proviene de/hace referencia a |

| ACEITUNA | |
|---|---|
| "aceituna de la variedad homónima…" | |
| **PARAMETROS** | **FORMULAS LINGÜISTICAS** |
| **descripción** | de peso bajo/medio/elevado/muy elevado |
| | de forma esférica/ alargada/ ovoidal |
| | de ápice redondeado/ apuntado y pezón ausente/ esbozado/ evidente |
| | de color violeta-negro/negro/rojo vinoso en maduración |
| **aptitud** | de doble aptitud |
| | apta para almazara/ aderezo |
| **rendimiento graso** | de rendimiento graso bajo/medio/alto |

| ACEITE | |
|---|---|
| "aceite de oliva elaborado con aceitunas de la variedad homónima…" | |
| **PARAMETROS** | **FORMULAS LINGÜISTICAS** |
| **características organolépticas** | I. dulce y frutado con poca intensidad de amargo y picante<br>II. dulce, frutado de tipo verde con ligero predominio del amargo-picante sobre el dulce<br>III. frutado de tipo verde, amargo y picante |
| | a. bajo/medio/alto en ácido oleico, bajo/medio/alto en linoleico y bajo/medio/alto en polifenoles<br>b. y de estabilidad baja/media/alta/muy alta |
| **información adicional** | DOPs |

Veamos, a modo de ejemplo, la entrada *'Picual'* del *Diccionario de términos del aceite de oliva*[9] elaborada a partir de estos parámetros definicionales:

[9] Mercedes Roldán Vendrell (2013): *Diccionario de términos del aceite de oliva (español-inglés-chino)*. Madrid: Arco Libros.

**'Picual'**
**1.** adj./s. AGRICULTURA. Relativo a la variedad de olivo de vigor medio, porte abierto, densidad de copa espesa y hoja de forma elíptico lanceolada; adaptable a suelos calizos y tolerante a la sequía, la tuberculosis y la aceituna jabonosa, pero susceptible a la mosca y muy susceptible al repilo y la verticilosis; muy apreciada por su precoz entrada en producción y alta productividad; cuyos frutos presentan una baja resistencia al desprendimiento, lo que facilita su recolección mecanizada. *En el olivar tradicional giennense, todas las fincas son de la variedad de olivo picual, de secano, y poseen entre 2 y 4 pies*. **Sin.:** Andaluza, Blanca, Corriente, de Aceite, de Calidad, Fina, Grosal, Jabata, Morcona, Marteño, Nevadillo, Nevadillo blanco, Nevado, Nevado blanco, Picúa, Redondilla, Salgar, Sevillano, Temprana, Lopereño. **Inf. adic.:** la variedad de olivo 'Picual' es la 1ª en importancia (superficie x 1000 ha) del mundo. Representa el 20% del olivar mundial y el 50% del español. Dominante en la provincia de Jaén (con 550.000 hectáreas de olivar que la convierten en el mayor productor mundial de aceite de oliva), y cultivada también en Córdoba y Granada. Su denominación hace referencia a la forma apuntada de sus frutos. **2.** adj./s. AGRICULTURA. Relativo a la aceituna de la variedad homónima de peso medio, forma ovoidal, ápice redondeado, pezón esbozado y color negro en maduración, de rendimiento graso alto y apta para almazara. *De todas las variedades de aceituna, la picual es la más importante del mundo por volumen, ya que supone el 20% del olivar a nivel internacional*. **Sin.:** Marteña, Lopereña. **3.** adj./s. ACEITES. Relativo al aceite de oliva elaborado con aceitunas de la variedad homónima, dulce, frutado y con ligero predominio del amargo y picante, de contenido alto en ácido oleico, bajo en linoleico y alto en polifenoles, y de estabilidad muy alta. *El aceite picual Castillo de Canena Reserva Familiar es impactante en nariz por la intensidad de sus aromas a hierba recién cortada, hojas de tomate y alcachofa.* **Inf. adic.:** DOPs Sierra de Segura, Sierra Mágina y Sierra de Cazorla. IGP Aceites de Jaén (con tutela nacional transitoria al encontrarse en proceso de reconocimiento por parte de la UE).

Paralelamente, aplicando estos mismos parámetros definicionales y tras el correspondiente trabajo de documentación, podemos elaborar el artículo terminográfico correspondiente a la variedad *Sikitita*:

**'Sikitita'**
**1.** adj./s. AGRICULTURA. Relativo a la variedad de olivo que proviene de un cruzamiento entre 'Picual' y 'Arbequina', de vigor muy bajo, porte compacto y llorón y densidad de copa clara, lo que la hace especialmente apta para plantaciones de olivar en seto, de hoja de forma elíptico-lanceolada, resistente al frío, el repilo y la verticilosis, de precoz entrada en producción y alta productividad, cuyos frutos presentan una resistencia baja al desprendimiento y un calibre pequeño, lo que facilita su recolección mecanizada. **Alt**: Chiquitita. **Inf. Adic**.: la variedad 'Sikitita' es la primera variedad de olivo seleccionada para su utilización en plantaciones de olivar en seto. Los obtentores de dicha variedad han sido investigadores de la Universidad de Córdoba y el Instituto Andaluz de Investigación y Formación Agraria (IFAPA) en el marco de un programa de mejora genética de olivo que ambas instituciones llevan a cabo desde 1991. La variedad se ha protegido en la Unión Europea y en Chile, Argentina, México, Estados Unidos (con la denominación 'Chiquitita'), África

del Sur, Australia, Túnez, Marruecos y Turquía. **2.** adj./s. AGRICULTURA. Relativo a la aceituna de la variedad homónima de peso medio, forma ovoidal, ápice redondeado, pezón ausente y color negro en maduración, de rendimiento graso alto y apta para almazara.
**Alt**: Chiquitita. **3.** adj./s. COMERCIO. Relativo al aceite de oliva elaborado con aceitunas de la variedad homónima, dulce, frutado y con amargo y picante equilibrados, de contenido medio en ácido oleico y linoleico y medio en polifenoles, y de estabilidad media. **Alt**: Chiquitita.

## 8. Conclusiones

La aplicación de patrones léxico-sintácticos que funcionan como marcadores de creación léxica es una herramienta útil para la detección, análisis y clasificación de neologismos en corpus especializados. Los datos obtenidos en el proyecto "Olivaterm" demuestran el interés de este recurso tanto para la detección de voces nuevas como para la delimitación de su significado respecto de las demás unidades del sistema. Resulta especialmente interesante la aplicación de *marcadores de hiperonimia, sinonimia y antonimia* para el estudio de las relaciones semánticas que se establecen entre los neologismos y otras unidades ya registradas en el mismo ámbito. Por último, en estas páginas hemos propuesto un modelo de definición de neologismos respetuoso con los principios de sistematicidad y coherencia. Este modelo tiene además ventajas desde la perspectiva de la comunicación y la traducción. Por una parte, es adecuado para la transferencia de conocimiento especializado a un amplio elenco de usuarios tanto expertos como no expertos en la materia. Por otra, facilita el proceso de traducción a otras lenguas y, con ello, la comunicación especializada en contextos multilingües.

## Agradecimientos

Este trabajo ha sido realizado en el marco del Proyecto de Investigación de Excelencia «Terminología del aceite de oliva y comercio: China y otros mercados internacionales» (P07-HUM-03041), financiado por la Consejería de Economía, Innovación, Ciencia y Empleo de la Junta de Andalucía.

## Referencias bibliográficas

CABRÉ M. T. (1999), *La terminología. Representación y comunicación*, Barcelona: IULA. Universitat Pompeu Fabra.
CABRÉ M. T. (2002), "Una nueva teoría de la terminología: de la denominación a la comunicación", En GALLARDO SAN SALVADOR N. (dir.), *Terminología desenvolvimento e identidade nacional*, VI Simposio Ibero-Americano de Terminología. Lisboa: Colibrí, 41-60.

FABER P. (2002), "Investigar en terminología", En FABER P. y JIMÉNEZ C. (eds.), *Investigar en Terminología*, Granada: Comares, 3–24.

FABER P. y JIMÉNEZ HURTADO C. (eds.), (2002), *Investigar en terminología*, Granada: Comares.

GALLARDO SAN SALVADOR N. (dir.), (2003), *Terminología y traducción: Un bosquejo de su evolución*, Granada: Atrio.

GARCÍA DE QUESADA M. y MONTERO MARTÍNEZ S. (2003), "Hacia una gramática de la definición terminográfica", En GALLARDO SAN SALVADOR N. (dir.), *Terminología y traducción: Un bosquejo de su evolución*, Granada: Atrio, 243-254.

GAUDIN F. (2003), *Socioterminologie. Une approche sociolinguistique de la terminologie*, Bruxelles: Editions de Boeck, Duculot.

JANSSEN M. (2009), "Detección de neologismos: una perspectiva computacional" *Debate terminológico*, (Ago. 2009) No. 05, 68-75.

MONTORO DEL ARCO E. T. y ROLDÁN VENDRELL M. (2013), "Terminología, normalización y comunicación: las categorías de aceite de oliva en español, inglés y chino", *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication,* 19:1, 62-92.

ROLDÁN VENDRELL M. (2010), "Lingüística y terminología multilingüe: Olivaterm", En ROLDÁN VENDRELL M. (ed.), *Bases para la terminología multilingüe del aceite de oliva*, Granada: Comares. Colección Interlingua, 1-15.

ROLDÁN VENDRELL M. (ed.), (2010), *Bases para la terminología multilingüe del aceite de oliva*, Granada: Comares. Colección Interlingua.

ROLDÁN VENDRELL M. y FERNÁNDEZ DOMÍNGUEZ J. (2012), "Emergent Neologisms and Lexical Gaps in Specialised Languages", *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication*, 18:1, 9-26.

ROLDÁN VENDRELL M. (en prensa), "Lingüística y gestión terminológica", En MARCET V. *et al.* (eds.), *Pro lingua. Nuevas aportaciones al estudio de la lengua.* Salamanca: Editorial Luso Española.

TEMMERMAN R. (2000), *Towards New Ways of Terminology Description. The Sociocognitive Approach*, Amsterdam/Philadelphia: John Benjamins.

# A view into retronymy as a source of neology

George J. Xydopoulos[1] and Irene Lazana[2]

## Summary

In this paper, we discuss the mechanics of retronymy, its semantics and its contribution to neology. In the first two sections, we define the phenomenon and briefly review the relevant literature. In section 3, we reconsider the main characteristics of retronymy and we view the whole relation in terms of (auto)hyponymy/taxonymy and oppositeness. We also extend our analysis to polarity and to reccurence. Furthermore, we examine the morphological status of retronyms/neonyms and the abbreviating process in retronym/neonym expressions as an indication for transfer from special to general vocabulary. In section 5, we view the mechanism that imports English retronyms/neonyms into Greek and explore a few cases of indigenous retronyms/neonyms in Greek.

## Key words

retronymy, (auto)hyponymy/taxonymy, oppositeness, morphological status of retronyms, abbreviation and neology, imported and indigenous retronymy

## 1. Preliminary remarks

By the term "retronymy" we refer to the relation that holds between the members of pairs as in (1), where the second member is referred to as the "retronym":

1. a. watch – analog watch
   b. milk – whole milk
   c. book – printed book
   d. computer – desktop computer
   e. TV – black & white TV
   f. oven – conventional oven
   g. language – natural language
   h. guitar – acoustic guitar

In all these pairs from English, the first term (henceforth the protonym) becomes more specific with the addition of an adjectival or nominal premodifier (henceforth the retronym) because of the appearance of a new

[1] University of Patras, gjxydo@upatras.gr.
[2] University of Patras, elazana@upatras.gr.

term (henceforth the neonym) that was coined in order to name a new invention or discovery that is conceptually related to the old one:

2. a. digital watch → watch > analog watch
   b. skimmed milk → milk > whole milk
   c. e(lectronic) book → book > printed book
   d. portable/laptop computer → computer > desktop computer
   e. color TV → TV > black & white TV
   f. microwave oven → oven > conventional oven
   g. artificial language → language > natural language
   h. electric guitar → guitar > acoustic guitar

## 2. Review of bibliography and definition

Although retronymy is an easily recognizable and comprehensible relation by the average speaker, there has been very little discussion of it in the literature. Actually, the term "retronymy" was allegedly coined by the American journalist Frank Mankiewicze in the 1980s as recorded in the Oxford English Dictionary (Smith, 2003; Soanes, 2005; Safire, 2007).

The French philosopher Jacques Derrida was the first scholar to discuss the phenomenon in a systematic way. He uses for it the term "paleonymy" (i.e. the Greek- based equivalent of "retronymy") and understands retronymy as the "strategic necessity that requires the occasional maintenance of an old name in order to launch a new concept" (Derrida, 1981: 71).

Mentions and/or very brief discussions of retronymy are found in some introductory linguistics books, specialized dictionaries and terminology-related papers where the relation is defined and illustrated with a few examples mainly from English (see e.g. Ahmad, Collingham, 1996; Hartmann, James, 1998; Olson, 2001; Fromkin *et al.*, 2007; Blöhdorn, 2008; Curzan, Adams, 2008).

The most detailed linguistic discussions of retronymy in the literature that we are aware of are those by Anastassiadi-Symeonidi (2001) and Xydopoulos (2008) for Greek and by Xydopoulos (2009) for English.

Anastasiadi-Symeonidi (2001) refers to the phenomenon using the term "rebaptism" ("rebaptême" in French) and understands retronymy as a special case of lexical change. So, as she puts it, due to advances in technology or science, the conceptual frame of a term is restructured, leading to the reappearance of the old term in a modified form. According to her analysis, retronymy is only obtained if an existing term is renamed through a "rebaptist", and, the existing term is generalized and transformed into a class of objects rather than an object.

Xydopoulos (2008) gives a brief introduction to this sense relation, focusing on data from Greek, and he puts emphasis on the description of the mechanism that generates retronyms. Moreover, Xydopoulos (2009) claims that the old and the new terms are related to each other through hyponymy on the vertical axis and through oppositeness on the horizontal axis (see also Roldán-Vendrell, Fernández-Domínguez, 2012). The relation between the two modified terms is viewed as taxonymy given the fact that they differ in a key distinctive characteristic of their superordinate term. Furthermore, he argues that the item that motivates retronymy is the neonym, the so-called "retronymy instigator". Moreover, the "generalization" of the protonym is understood as its transformation into an autohyponym.

In a nutshell, Xydopoulos (2008, 2009) understands the generation of retronymy as comprising four distinct phases:

3. a. Creation of the neonym, that is of a hyponym to name a new concept on the basis of its basic difference with the old concept: e.g. *digital watch*: "type of display" as the basic difference from *watch*.
   b. Transformation of the protonym into an autohyponym: e.g. *watch* (meaning either "any watch" (generic) or "analog watch" (specific)).
   c. Creation of the retronym, by modifying the protonym with the addition of an opposite (referring to the basic difference) to create a new incompatible co-hyponymic term: e.g. *analog watch*.
   d. Optional abbreviation of well-established terms through nominalization of the premodifier and deletion of the noun head, e.g. *mobile phone* → *mobile*.

## 3. A fresh look into retronymy

In this work we examine retronymy in more details, following the theoretical principles set in Xydopoulos (2008 and mostly 2009). We expand the analysis of oppositeness in retronymy by adding the dimension of polarity. We also reexamine autohyponymy as the chief effect of the recursive character of retronymy and address the ontogenetic and phylogenetic issues involved. Furthermore, we discuss the morphological status of retronyms and neonyms in the context of compounding and we examine the mechanism that transfers retronymic items from special vocabularies into general vocabulary. We also examine the status of retronymy in Greek, stressing the fact that the majority of retronymic vocabulary is imported from English through loan translation.

### 3.1. On the relation between the protonym and the neonym

In almost all accounts, retronymy is treated as an independent sense relation (e.g. Ahmad, Collingham, 1996; Hartmann, James, 1998; Anastasiadi-Symeonidi, 2001; Fromkin *et al.*, 2003; Xydopoulos, 2008 among others). According to Xydopoulos (2009), retronymy should be treated as an epiphenomenon as all the lexical items involved are related to each other in terms of two basic sense relations, that is, of taxonomic hierarchy and, consequently, of oppositeness (see also Roldán-Vendrell, Fernández-Domínguez, 2012 and references therein). In other words, the protonym and its retronym differ in a "key distinctive characteristic" (Cruse, 1986: 88-89, 2004: 150-151). This can be verified by satisfying the following logical entailment criterion (Lyons, 1977: 292-293, 1999: 148-150; Cruse, 1986: 88-89, 2004: 149, 2011: 135-6):

4. *An X is a (different) kind of Y.*

Application of this criterion is illustrated with the pair:

5. a. An *analog watch* is a (different) kind of *watch*.
   b. *A *watch* is a (different) kind of *analog watch*.

The "key distinctive characteristic" in this pair is the way time is displayed, either with a marked disk and hands or with digits. Now, this reads as: *analog watch* is the hyponym (or better a taxonym) of *watch* (the hypernym or superordinate term) and when something qualifies as an *analog watch* it follows that is also a *watch*, and not vice-versa:

6. *watch* [PROTONYM]

[RETRONYM] *analog watch* *digital watch*

This means that by the term *watch* alone we cannot understand whether it is an *analog watch* or a *digital watch*; it can be both.

Hyponymy is understood as a relation of inclusion. From the extensional point of view, the hypernym (*watch*) includes the meaning of the hyponym (*analog watch*) which denotes a subclass of the "watch" class. Equally, from the intensional point of view, the hyponym is semantically richer as it includes the meaning of the hypernym ("watch") plus an extra distinctive characteristic ("analog") (Lyons, 1977: 291; Cruse, 2004: 148, 2011: 134).

### 3.2. The neonym and its relation with the retronymic terms

Let us now turn to the neonym. Why talk about *conventional weapons*, *whole milk*, *desktop computer* etc. if *nuclear weapons*, *skimmed milk*, *laptop computers* etc. had not been devised? So, the neonym, operating as the retronymy-instigator, is the reason for renaming the protonym with the addition of the proper modifier.

In addition, the neonym is said to be semantically incompatible with the retronym. By "incompatibility" we understand non-overlapping properties and mutually exclusive subcategories (Cruse, 2004: 161, 2011: 151; Karaman, 2008: 178; cf. Roldán-Vendrell, Fernández-Domínguez, 2012: 16). This stems from the fact that the neonym and the retronym are co-taxonyms and resides on the nature of the modifiers that constitute the two terms.

7. a. *hardback book* vs. *paperback book*
   b. *wired network* vs. *wireless network*
   c. *tap water* vs. *bottled water*
   d. *conventional oven* vs. *microwave oven*
   e. *heterosexual marriage* vs. *homosexual marriage*

The modifiers in each pair utterly declare that the two entities described constitute two different kinds of the same class. They are modifiers that describe the same "key distinctive characteristic", or else, the same "prototypical characteristic" that distinguishes members of the pair from each other[3]. So the key distinctive characteristic in the pairs is: the "material of the book cover" (7a), the "transmission medium of the signal" (7b), the "source of drinking water" (7c), the "type of thermic energy" (7d) and the "sexual orientation of the couple" (7e).

It is important to note here that, in these pairs, the oppositeness that holds between the members is not inherent (see Cruse, 2004: 162-163, 2011: 153-154). So there is no inherent oppositeness in the *hardback* vs. *paperback*, *tap* vs. *bottled* or *conventional* vs. *microwave* pairs. It is a binary relation that stems from the meaning of the modifiers in a specific context or situation and so it constitutes a case of situational (and accidental) or pragmatic oppositeness due to sociocultural reasons.

#### 3.2.1. Polarity in retronymy

Following the line of thought in Xydopoulos (2009) we would like to further explore oppositeness as related to retronymy from the point of view of polarity as proposed in Cruse (2004: 169-170, 2011: 162-163). Pairs of opposites are

[3] A characteristic can be considered as "prototypical" only when it refers to properties that can distinguish a hyponym from its hypernym and not to general properties such as color or age (Xydopoulos, 2008: 122).

viewed as consisting of a negative and a positive pole. Several, if not all, retronym-neonym pairs, and their defining modifiers, can be viewed as involving polarity. In particular, these pairs can display either morphological polarity or privative polarity.

*3.2.1.1. Morphological polarity*

Some retronym-neonym pairs are based on morphological polarity. In our data (see Appendix), the neonym is systematically the negative member, marked with the appropriate negative morpheme, and the retronym is systematically the positive member without morphological marking, as we see in (8):

8. a. wired network – wireless network
   b. tobacco cigarette – non-tobacco cigarette
   c. caffeinated coffee – decaffeinated coffee

A pair like *wired network* and *wireless network* is a case of morphological polarity with the retronym *wired network* being the positive pole, bearing no marking, and the neonym *wireless network* being the negative pole, bearing the negative morpheme *-less*. Similarly, in the pair *tobacco cigarette – non-tobacco cigarette*, negative polarity is marked with the *non* morpheme, and in the pair *caffeinated coffee – decaffeinated coffee*, the negative member bears the morpheme *de-*.

The absence of the key distinctive characteristic signals the introduction of a new invention or product in technology and this is mapped into language through this morphological process. It is important to note though that morphological polarity is found with a rather small number of retronyms/neonyms probably because terms which are morphologically marked (i.e. denote which characteristic they lack) are less descriptive than terms that denote which characteristic determines them.

*3.2.1.2. Privative polarity*

Unlike morphological polarity, in privative polarity, the positively marked member contains a modifier that refers to the presence of the key distinctive characteristic, while the negatively marked member contains a modifier that refers to the absence of the key distinctive characteristic. In other words, privative polarity is lexically realized and not morphologically marked:

9. a. natural flavour – artificial flavour
   b. silent movie – talking movie
   c. two-parent family – single-parent family
   d. dark chocolate – white chocolate
   e. hardback book – softback book
   f. natural language – artificial language
   g. wild animal – domestic animal

In a pair like *natural flavour – artificial flavour* in (9a), the retronym *natural flavour* is the negative member referring to the "absence of synthetic ingredients", while the neonym *artificial flavour* is the positive member referring to the "presence of chemical ingredients". Similarly, in the pair *silent movie – talking movie* in (9b), the retronym *silent movie* refers to the "absence of sound", being the negative member, and the neonym *talking movie* refers to the "presence of sound", being the positive member. However, in the pair *two-parent family – single-parent family* in (9c), the retronym refers to the "presence of both parents" and the neonym to the "absence of both parents". Other examples of privative polarity include pairs (9d) to (9g) like *dark chocolate – white chocolate* (re: "presence/absence of cocoa solids"), *hardback book – softback book* (re: "presence/absence of sewn spines), *natural language – artificial language* (re: "presence/absence of native speakers") and *wild animal – domestic animal* (re: "presence/absence of wildness") etc. Therefore, contrary to the systematicity in morphological polarity, in privative polarity, negative and positive poles can be lexically attributed either to retronyms or to neonyms.

By examining the data collected from both English and Greek, it appears that there are a few retronym-neonym pairs that could not be viewed as involving any of the two types of polarity we saw earlier:

10. a. analog TV – digital TV
    b. classical music – modern music
    c. alpine ski – water ski
    d. pulse dial – tone dial

Pairs like *analog TV - digital TV*, *classical music – modern music*, *alpine ski – water ski*, *pulse dial – tone dial* do not show the absence of a prototypical characteristic but, instead, an alternative version of some entity. So, oppositeness here, we think, can merely be understood as incompatibility.

*3.2.1.3. Evaluative polarity*

A third type of polarity introduced by Cruse (2004: 170, 2011: 163) is the so-called "evaluative polarity". Here, one member of the pair is considered evaluatively positive or commendatory and the other member is considered negative, on the basis of sociocultural criteria, e.g. *good-bad, kind-cruel, clean-dirty*. However, we cannot say that, systematically, a retronym is evaluatively negative as opposed to a positively evaluated neonym, or vice-versa as claimed by Anastasiadi-Symeonidi (2001: 70-71). Let us examine the following examples:

11. a. natural flavour – artificial flavour
    b. black & white TV – color TV

In the pair *natural flavour – artificial flavour*, the retronym can be positively evaluated as "it does not contain artificial / potentially harmful ingredients". On the contrary, in the pair *black-and-white TV – color TV*, the neonym can be positively evaluated as "it offers a better picture to viewers". So, although some sociocultural evaluation is necessarily involved in oppositeness, evaluative polarity is not a defining property for such pairs, as is the case with morphological and privative polarity. Retronyms and neonyms just denote two different kinds of a class, not necessarily two evaluatively opposed entities. Consequently, in a retronym/neonym pair, when one member is positively evaluated the other is either less positively or neutrally conceived, but never negatively.

### 3.3. The hypernymic status of the protonym

As claimed in Xydopoulos (2009), and references therein, one of the two conditions for retronymy is that the protonym is generalized and transformed into a class rather than an object. This is a crucial phase in the retronymy procedure. The protonym is turned into a hypernym, so its sense becomes general enough to include the specific senses of the retronym and the neonym. However, it is interesting to see that at the same time it can be used both in its general and specific senses, with obvious contextual restrictions as we see in (12):

12. a. *phone* = 1: any type of; 2: landline phone; 3: mobile phone
    b. *computer* = 1: any type of; 2: desktop computer; 3: laptop computer
    c. *water* = 1: any type of; 2: tap water; 3: bottled water

According to Xydopoulos (2009), the fact that a single term can have a default general sense and contextually restricted senses means that the protonym is an autotaxonym, in the sense of Cruse (2004: 108-109, 2011: 115-116). This is supported by the following criterion where the general sense is contrasted with a specific sense:

13. a. Don't call me on my *phone*, use my *mobile*.
    b. I have a *computer* at home but I prefer using my *laptop*.
    c. If you try to defrost food in the *oven* it will take ages; don't you have a *microwave*?

It is important to note here that the protonym is not conceived in its general and its specific sense at the same time. At first, it is understood as the hypernymic term (general sense) and only after the utterance has been completed is it understood as a hyponym (specific sense). The interpretation of a protonym with the one or the other specific sense depends on both linguistic and extra-linguistic factors, given that pragmatic and encyclopaedic knowledge is also crucial here.

### 3.4. Recursion

In general, recursion refers to the property of some linguistic rules and structures to be repeated (see e.g. Crystal, 2008: 405). In retronymy however, recursion is mainly about conceptual and ontological relationships and can have an internal or an external dimension.

Let us first see recursion as an internal dimension. Nowadays, when we talk about *airplanes*, we mostly mean *jet airplanes* and not *propeller airplanes*:

14. a. jet airplane$_j$ → airplane$_j$ > propeller airplane
    b. laptop computer → computer$_k$ > desktop computer$_k$

So, *jet airplanes* and *propeller airplanes* are not equally competitive as most airplanes now have jet engines. Then, the term *airplane* is now the generalization of *jet airplane* and not of *propeller airplane* as it used to be the case before the advent of "jet airplanes". It seems that we are now at a transitory stage where we still expect a new type of "airplane" to appear to instigate retronymy once again. This, we think, is natural as "airplanes with a jet engine" have been around for some fifty years. The same holds for the pair *TV – color TV*.

However, think about the pair *computer – desktop computer*. Here, *computer* still refers to *desktop computers* (unless the linguistic or extra-linguistic information activates the other specific meaning) because "laptops" are still a new invention (round 15 years old). Since *laptop computers* are becoming more and more popular it is easily foreseeable that the term *computer* will refer to "laptops" and will again be renamed by modification when a new type of computer appears in the future, while desktops (as we know them today) will cease to exist.
In the same vein, if we hear the utterance in (15):

15. Mary bought a new TV.

we understand that Mary bought a "color TV" and not a "black and white" one. So the term "TV" is not an autohyponym of "black and white TV" anymore but of "color TV".

Therefore, retronymy is internally recursive because autohyponymy is redetermined and the hypernym takes the specific senses of the other available hyponyms in its hierarchy, as mandated by technological progress. Recursion in retronymy refers to a redetermination of the basic specific meaning, i.e. the specific meaning that speakers obtain when they understand an autohyponym. The autohyponym ceases to denote the specific sense of the retronym (old basic sense) as it now denotes the specific sense of the neonym (new basic sense). Consequently, the appearance (and the establishment) of a new class of objects is again the factor that instigates retronymy, this time as a recursive process. However, in most cases of retronym/neonym pairs, the

redetermination of autohyponymy has not happened yet, as the neonym has not been established to such a degree as to denote the new basic specific sense of an autohyponym.

Recursion in retronymy also has an external dimension. Technological progress can produce new neonyms and so turn existing neonyms into retronyms and so on and so forth. This can be illustrated with the class of "ovens":

16.

In this hyponymic hierarchy, ovens are differentiated according to the prototypical characteristic of "source of thermic energy". Ovens burning wood were invented first, so then the term *oven* meant "wood oven". When the neonym *electric oven* was coined to denote "oven working with electricity" the retronym *wood oven* was coined. Furthermore, when "microwave oven" was invented and named with the corresponding neonym, *electric oven* was renamed as *conventional oven* and the term *oven* became the autohyponym of the term *conventional oven*.

17.

The invention of the microwave as a new kind of oven (i.e. ontogeny) was technologically revolutionary and so activated the creation of new classes of ovens, microwave and non-microwave ones (i.e. phylogeny). Wood ovens and electric ovens were classified under non-microwave ovens with the cover term *conventional ovens*.

This change in ontological relationships between different kinds of oven is the reason why the class of objects that is denoted by the term *electric oven* is renamed. In particular, ontology refers to positioning extra-linguistic objects of reference into a hierarchy (Cabré, 1999: 39, 44; Navigli *et al.*, 2003: 24; Gillam *et al.*, 2005: 2). When this hierarchy is redetermined it is also mapped onto language. Therefore, the way extra-linguistic entities are organised in a conceptual system affects the way these entities are named in language. Again, the appearance of a new class of objects motivates retronymy as a recursive process[4].

[4] Roldán-Vendrell, Fernández-Domínguez (2012: 17) view this type of recursive process as a case of gradable antonymy as there are two opposite poles (e.g. *wood oven* and *microwave oven*) and one element (e.g. electric oven) that displays properties of both poles: an electric oven is used for cooking (like a wood oven) and works with electricity

### 3.5. The morphological status of retronymic expressions

Expanding the analysis in Xydopoulos (2009), let us now consider the morphological status of neonyms and retronyms, an issue that we think needs special attention.
We observe that all neonyms and retronyms, at least in English, have the structure of Noun Phrases, where the head N is modified by a premodifier that is either an Adjective or an embedded Noun Phrase:

18. a. digital TV [[digital]$_A$ [TV]$_N$] $_{NP}$
    b. water ski [[water]$_{NP}$ [ski]$_N$] $_{NP}$

These formations cannot be treated as ordinary syntactic phrases. Arguably, they behave as some type of a unified lexical element. So we put forward the hypothesis that several retronymic expressions examined in this work appear to fulfil specific criteria to qualify as compounds while some others do not.

Two syntactic criteria to distinguish a word from a phrase that have been proposed and generally accepted in the relevant literature are those of "syntactic interdependence of constituents" and of "rigid constituency" (Plag, 2003; Scalise, Guevara, 2005; Spencer, 2005; Lieber, 2005; Giegerich, 2005; Booij, 2005, 2007, 2010)[5]. Let us test whether English pairs like *wet cleaning – dry cleaning*, *granulated sugar – powdered sugar* and *silent movie – talking movie* fulfil these criteria for compoundness.

**(a) Syntactic interdependence of constituents:** We observe that if we attempt to modify either the head noun or the modifier of these expressions, independently, we do obtain marginality, if not ungrammaticality, in all cases examined:

19. a. *[very dry/wet] cleaning
    b. *granulated [sweet sugar]
    c. ??silent [black and white movie]

---

(like a microwave oven). We cannot clearly see what could be the constitution of this continuum, as an antonymy of this type presupposes a common property in both poles.

[5] Other criteria to check the compoundness of an expression, but with limited applicability, include: a) lexicalized meaning, b) the modifier cannot be replaced by the element *one* (e.g. *What kind of lecture did she give? *A morphology one*), c) the first constituent cannot be inflected (but: *children's hour, girl's club*), d) the first constituent has no specific reference (but: *You're a real cat-lover. How many do you have now?*), e) the first constituent cannot be coordinated (but: *cat and dog show, medical and life insurance*), f) the adjective cannot be placed in a predicative position (e.g. **the movie is silent*), g) the adjective can be replaced by a noun (e.g. *dental decay – tooth decay*), h) the argument-predicate structure is lexical (Fabb, 2001; Plag, 2003; Spencer, 2005; Lieber, 2005; Giegerich, 2005; Booij, 2007; Ralli, 2012, to appear).

**(b) Rigid constituency:** If we attempt to conjoin one of the two constituents with an inserted element, thus breaking coherence of the formation, we obtain ill-formedness again:

20. a. *We would like a [[dry and fast] cleaning] of this suit.
    b. *You need [[powdered and white] sugar].
    c. ??I prefer [[talking and American] movies] of the 50's.

These retronym/neonym pairs fulfil both criteria; thus they cannot be considered as ordinary syntactic phrases.

The compoundness of such expressions is supported by the fact that they also display a degree of semantic opacity, as with a standard property related to compounds. If, for example, we take neonyms/retronyms like: *black market* (NB: not a black-colored market), *programming language* (NB: not a language that can program), *plastic surgery* (NB: not a surgery that is made of plastic), *silent movie* (NB: not a movie that does not speak) etc., we find out that they are indeed opaque, as their meaning is not compositional but idiomatic.

Another criterion for distinguishing word from phrase is that of stress. It is widely accepted that in a compound word, primary stress appears word-initially (e.g. '*greenhouse*) while in a syntactic phrase, primary stress appears word-finally (e.g. *green 'house*) (but cf. Giegerich, 2009). Indeed, we have found cases of retronymic expressions bearing one primary stress that are listed as independent lemmas in dictionaries (e.g. Hornby, 2000). So, these expressions have been lexicalized, despite the fact that the distribution of stress is not systematic:

21. a. 'powdered sugar
    b. 'surface mail
    c. 'tap water
    d. dry-'cleaning
    e. granulated 'sugar
    f. hard 'copy

Consequently, we view the criterion as an indication – but not a proof – that an expression is a lexical unit. Of course, the syntactic interdependence of constituents and the rigid constituency criteria must also be fulfilled (see also Plag, 2003).

So, it follows that English retronymic expressions are placed on a continuum: some (e.g. *dry cleaning*, *granulated sugar – powdered sugar*) are lexical units as they satisfy both syntactic criteria (see 19a-b and 20a-b) and have been lexicalised (21a, d-e) while others (e.g. *wet cleaning*, *silent movie – talking movie*) are probably in some process of being lexicalized as they satisfy the syntactic criteria (19a&c; 20a&c) but are not listed yet as independent lemmas.

Let us now turn to retronymy data in Greek. In Greek, retronymic expressions could qualify as phrasal compounds. Ralli (2012, to appear) discusses phrasal compounds in Greek and proposes five distinct criteria for confirming their structural status. Apart from the criteria of syntactic interdependence of the constituents and rigid constituency which we saw earlier for English and which she also adopts for Greek, she proposes three criteria that are specific to Greek, namely: "strict word order", "no specifier doubling" and "no co-reference" (see also Ralli, 2003, 2005, 2007). Greek phrasal compounds have the structure of Noun Phrases, where the head N is modified by an Adjective or an embedded Noun Phrase in the genitive case that follows the head. According to Ralli (2012, to appear), phrasal compounds display syntactic properties as, in [A N] compounds, the adjective agrees with the noun head in gender, number and case while in [N $NP_{gen}$] compounds the NP is assigned the genitive case by the head. The same holds for the majority of Greek retronyms/neonyms:

22. a. fisikjí ghlósa [[fisikjí]$_A$ [ghlósa]$_N$] $_{NP}$
natural language
b. ghlósa proghramatizmú [[ghlósa]$_N$ [proghramatizmú] $_{NPgen}$] $_{NP}$
computer programming language

Let us see if all five criteria that Ralli (2012, to appear) has proposed for Greek phrasal compounds apply indeed in the case of the Greek neonyms/retronyms we have examined.

**(a) Syntactic interdependence of constituents:** This test yields ungrammaticality in all the Greek retronym/neonym cases tested:

23. a. *[sostá emfialoméno] neró
correctly bottled water
b. *[polí mávri] sokoláta
very dark chocolate
c. *[usiastiká violojikós] ghonéas
virtually bilological parent

**(b) Rigid constituency:** This test similarly applies successfully in all cases, as the coherence of the formation cannot be broken:

24. a. *Tha íthela [[emfialoméno kje paghoméno] neró]].
"I would like bottled and chilled water".
b. *'Efagha [[áspri kje ghlikjá] sokoláta]].
"I ate white and sweet chocolate".
c. *'Ine o [[thetós kje ipéfthinos] ghonéas]] tis Marías.
"He is Mary's adoptive and responsible parent".

(24) shows that when a property is added to the prototypical characteristic of a retronym/neonym (e.g. *bottled and* ***chilled*** *water*), the object of reference is confused.

**(c) Strict word order.** This test applies successfully despite the fact that Greek allows for free word order in its NPs:

25. a. ??Sto kjéik vázume sokoláta mávri.
"We put dark chocolate in the cake".
b. *Aftó íne vrísis neró.
"This is tap water".
c. *Ghonéas thetós dhen borí na jíni opjozdhípote.
"Not everyone can become an adoptive parent".

**(d) No article doubling:** This test shows that, although Greek allows doubling of the article in ordinary NPs, this is not permitted in phrasal compounds and consequently in retronymic expressions. Actually, according to Ralli (2007: 228), this criterion is the most basic one for characterizing an [A N] structure as a phrasal compound:

26. a. To emfialoméno (*to) neró íne pjo akrivó.
"Bottled water is more expensive".
b. I mávri (*i) sokoláta periéxji kakáo.
"Dark chocolate contains cocoa solids".
c. Pji íne i violojikí (*i) ghonís su?
"Who are your biological parents?".

**(e) No co-reference:** This test has restricted application to phrasal compounds of the form [N $NP_{gen}$] in Greek. Its application is also successful with retronyms/neonyms of this structure:

27. a. *'Ipja neró [vrísis]$_i$ [i opía]$_i$ íne stin platía.
"I drank tap water from the fountain that is on the square".
b. *Mu arési I sokoláta [ghálaktos]$_i$ [to opío]$_i$ íne apó katsíka.
"I like milk chocolate from goat's milk".
c. *Píra énan ektipotí [krúsis]$_i$ [i opía]$_i$ íne ekofantikjí.
"I bought a matrix printer that makes a terrible noise".

The modifiers in the genitive in (27) cannot be co-referred because constituents of retronyms/neonyms denote a single object of reference (e.g. "milk chocolate" in (27b)) and not two independent entities (e.g. "milk" and "chocolate").

Greek retronyms /neonyms as phrasal compounds are semi-visible to syntax as, although they are created in syntax and display syntactic properties (agreement,

case assignment), they fulfil the above restrictions and thus are subject to the lexical integrity hypothesis (Ralli, 2012, to appear).

From the application of those criteria to English and Greek data, we can conclude that all the retronyms/neonyms tested above qualify as compounds in both languages (but cf. Christofidou, 2010). However, there are cases of retronyms/neonyms in both English and Greek that are not formally compounds. For example, the following English and Greek examples

28. a. natural light – artificial light
fisikó fos – texnitó fos
b. analog TV – digital TV
analojikjí tileórasi – psifiakjí tileórasi
c. printed copy – electronic copy
édipo adíghrafo – ilektronikó adíghrafo

are not classified as compounds as they cannot satisfy all or some of the tests we saw earlier, the bond between the two constituents being very loose. The same can be said for example of the retronym *thalasinó aláti* (sea salt) on the basis of these tests. The item (marginally) satisfies the rigid constituency test (29b) and the strict word-order test (29c) but fails the other two:

29 a. [praghmatiká thalasinó] aláti. **(syntactic interdependence)**
"Really sea salt"
b. ??Mu arési to [thalasinó kje aghnó] aláti. **(rigid constituency)**
"I like sea and pure salt".
c. ??Fére mu [aláti thalasinó] apó to mesolóngi **(strict word order)**
"Bring me sea salt from Mesolongi".
d. [To aláti to thalasinó] íne to pjo kaló. **(article doubling)**
"Sea salt is the best".

It follows that retronyms/neonyms can be classified in at least three morphological groups: (a) compounds; (b) quasi-compounds (*alias* constructs, Ralli, 2007, 2012); or (c) ordinary/regular NPs. This means that several of these items have been lexicalized, others are in the process of being lexicalized, while yet others still maintain their syntactic status. Perhaps the degree of lexicalization of retronymic pairs is dependent on factors like frequency of use or the nature of the constituents. So items like *electronic copy* and its Greek equivalent *ilektronikó adíghrafo* do not qualify as compounds because of the classificatory nature of the adjective "electronic" ("*ilektronikó*"), as opposed to the "printed" one ("*édipo*"), and not because of the structure *per se* (see Ralli, 2003: 109, 2007: 240-241, 2012). We will leave this issue for future research.

### 3.6. Abbreviation through conversion

In many cases in Greek, retronym and neonym expressions can be abbreviated. Abbreviation is a general process in word formation and neology (Ahmad, 2000: 2; Crystal, 2008: 1) and, according to Xydopoulos (2009), it is the last (optional) phase of retronym production.

Greek retronymic expressions can be abbreviated by deleting one of the two constituents of the phrase and converting the remaining constituent. In most if not all of the cases the modifier, as the remaining constituent, is then converted into a noun which maintains the categorial features of the deleted noun (see e.g. Anastasiadi-Symeonidi, 1986: 150):

30. a. [i [psifiakjí]$_{ADJ}$ [~~kámera~~]$_{N}$]$_{NP}$ >[i [psifiakjí]$_{N}$]$_{NP}$
"digital camera" "digital camera"
b. [i [dhoriforikjí]$_{ADJ}$ [~~tilcórasi~~]$_{N}$]$_{NP}$ >[i [dhoriforikjí]$_{N}$]$_{NP}$
"satellite TV" "satellite TV"
c. [o [dekafeiné]$_{ADJ}$ [~~kafés~~]$_{N}$]$_{NP}$ >[o [dekafeiné]$_{N}$]$_{NP}$
"decaffeinated coffee" "decaffeinated coffee"

In phrases with postnominal modifiers in the genitive, abbreviation deletes the head noun and the modifier remains intact and inherits the formal features of the deleted head:

31. a. [o [~~fúrnos~~]$_{N}$ [mikrokjimáton]$_{NPgen}$]$_{NP}$ > [o[mikrokjimáton]$_{NPgen}$]$_{NP}$
"microwave oven" "microwave oven"
b. [i [~~sokoláta~~]$_{N}$ [ghálaktos]$_{NPgen]\ NP}$ > [i[ghálaktos]$_{NPgen}$]$_{NP}$
"milk chocolate" "milk chocolate"

Abbreviation of the expression also applies to English retronyms/neonyms, for example:

32. landline ~~phone~~, LCD/plasma ~~TV~~, decaf ~~coffee~~, cable ~~TV~~, microwave ~~oven~~

However, such abbreviation in English seems to appear in proportionally fewer cases, maybe because, in Greek, the rich inflection in the NP is maintained in the article and this facilitates the semantic recovery of the deleted material.

## 4. From special to general vocabulary

Given that the modifier is always the element that is maintained and the head term the element that is deleted, it follows that it is much more crucial to state the distinctive property of the expression than to state the general object of reference. Furthermore, abbreviation is usually found in cases of lexically established neonyms of high frequency of use while abbreviated forms are

established in general vocabulary without being confused or misunderstood (Cruse, 2004: 108).

We would like to understand this abbreviatory process as a formal indication that the affected term has been transferred from special to general vocabulary. These terms can denote specific properties or methods in a special technological or scientific domain, thanks to the content of the modifier and not of the head noun:

33. [digital camera], [satellite TV], [cesarean section birth]

As we see in (33), the modifiers are technical terms and head nouns are general terms. When such terms (i.e. the modifiers) become widely familiar to non-experts, they are included in the general vocabulary. Furthermore, when they become sufficiently established, they can denote an object of reference on their own without the presence of the respective head noun. In this way, we believe, abbreviation signals transfer from special to general vocabulary. This of course can only happen if the modifier is sufficiently descriptive, that is if it carries all the necessary semantic information and can be used in the appropriate context without causing ambiguity.

However, it is interesting to mention here that we found two cases in our Greek corpus where the modifier is not a technical term but an element of general vocabulary:

34. xjiropíito rúxo ("handmade clothing"), sokoláta ghálaktos ("milk chocolate")

This finding does not contradict our earlier claim, as modifiers in both these expressions originate from Greek scholarly and not popular vocabulary and so their use is stereotypic and highly restricted, as is the case with technical terms, and so they are sufficiently descriptive when the head noun is omitted.

## 5. Imported and indigenous retronymy

The vast majority of retronym and neonym pairs in a language like Greek is imported from other languages. Nowadays, the main source for importing retronyms/neonyms is English, as technological advancement and innovation worldwide is communicated through English. Only a very small number of such pairs is indigenous in the target language (e.g. Greek), reflecting aspects of the culture of the speakers and their homeland.

### 5.1. Imported retronymy

Expectedly, the basic strategy for importing (English) retronymy into Greek vocabulary is mainly through calquing / loan translation. During this process,

the English term is translated into the Greek equivalent term, as closely as possible and on the basis of specific criteria,[6] as we can see in the following examples:

| | | | | |
|---|---|---|---|---|
| 35. | a. | acoustic guitar | > | akustikjí kithára |
| | | electric guitar | > | ilektrikjí kithára |
| | b. | analog watch | > | analojikó rolói |
| | | digital watch | > | psifiakó rolói |
| | c. | black-and-white photography | > | asprómavri fotoghrafía |
| | | color photography | > | éghxromi fotoghrafía |
| | d. | natural flavour | > | fisikjí jéfsi |
| | | artificial flavour | > | texnití jéfsi |
| | e. | desktop computer | > | epitrapézios ipolojistís |
| | | portable computer | > | foritós ipolojistís |

As we can see in the first two cases (35a-b), loan translation consists in the mere transliteration of internationalisms.

However, in some cases, loan translation from English into Greek is selective. To avoid fuzziness or ambiguity, the language chooses only those terms that can be calqued into Greek and rejects all other available terms:

36. a. praghmatikós xóros "*real world*"< *meat world* / *meatspace*
    b. simvatikó taxjidhromí*o* "*conventional mail*" < *snail mail*
    c. nómimi aghorá "*legal market*" < *white market*

These examples are Greek retronyms that constitute broad translation loans from English. In (36a-c), we can observe that Greek prefers to import pairs with modifiers that have literal meanings (e.g. "real world", "conventional mail", "legal market"), in the place of other available expressions with modifiers used metaphorically (i.e. *meat world*, *snail mail*, *white market*). It is so in order to achieve the highest degree of formal and pragmatic equivalence in the Greek equivalent term, given that the way the extra-linguistic world is organised at the ontological level and attributed to at the linguistic level may vary among languages (Cabré, 1999: 39).

---

[6] The majority of the equivalent Greek neonyms/retronyms translated from English satisfy the criteria proposed for calquing in special vocabularies, as verified by Lazana (2012: 65ff). In particular, the Greek equivalents fulfill Babiniotis's (1993, 1994) criteria, namely: *acceptability*, *descriptiveness*, *reversibility* and *translatability*. Lazana (*op. cit.*) tested imported retronyms/neonyms by applying four macro-rules proposed by Xydopoulos (2002, 2004, 2008, 2011): (a) verify the meaning of the proposed equivalent by checking the relevant definitions; (b) verify the degree of standardization for existing equivalents (if any) and adopt or reject them; (c) verify that the proposed equivalent can regularly produce all derivatives needed; and (d) make sure that the proposed equivalent avoids ambiguity as much as possible (see also Anastasiadi-Symeonidi, 1986; Cabré, 1999; Tsakona, 2007a, 2007b, 2008).

The same holds for the following Greek retronyms:

37. a. statheró tiléfono "*stable phone*" < *landline phone*
    b. skji vunú "*mountain skiing*" < *snow skiing*
    c. mávri sokoláta "*black chocolate*" < *dark chocolate*

However, in these cases, loan translation is not selective as the English term is the only one that denotes the respective object of reference. Here, we observe the tendency of Greek modifiers to realize inherent ("*stable* " – "*mobile*", "*black* " – "*white*") and not situational or pragmatic oppositeness ("*landline*" – "*mobile*", "*dark*" – "*white*") between retronyms and neonyms. Even when oppositeness is situational in both languages (as in 37b), the most systematic oppositeness available in Greek ("mountain (ski)" – "sea (ski)") is chosen, instead of the English "snow" – "sea".

There are also kinds of neonyms that are imported into Greek as foreign loans through phonetic adaptation of the forms of the modifiers, while head nouns are represented by the Greek equivalents, as we see in (38):

38. a. LCD TV > tileórasi elsidí
    b. ADSL connection > sýndhesi eidiesél
    c. laser printer > ektipotís léizer

In all these cases, modifiers are English acronyms, and so they are technically untranslatable in their condensed form into Greek. In cases where a (kind of) neonym is denoted by a one-word English term, the whole term is imported into Greek as a foreign loan:

39. a. minivan > miniván
    b. netbook > nétbuk

### 5.2. Indigenous retronymy

We have found just a few retronyms in Greek that are indigenous to the language (see Appendix). This is so because nowadays technological innovation is not disseminated through Greek. We found three examples that refer to Greece's traditional products which, interestingly, display retronymy. The first is with *feta* cheese:

40. féta

At first, there was only *varelisja feta* "feta matured in a barrel", but later *feta* was produced in large quantities and matured in tin containers, and generated the neonym *feta dhoxiu* "feta matured in a container". Similarly, there are two examples that refer to wine, as we see in (41), where retronyms appear as left branches and neonyms as right branches. In (41a), the neonymic form displays morphological polarity, while (41b) is a case of privative polarity referring to the absence of "container/bottle":

41. a.

b.

Indigenous retronymy is also found in Greek toponyms as shown in the following examples:

| | | | | | | |
|---|---|---|---|---|---|---|
| 42. a. | paleó súli | >> | súli | >> | néo súli |
| | Old Suli " | >> | "Suli" | >> | "New Suli" |
| b. | paleó psixjikó | >> | psixjikó | >> | néo psixjikó |
| | "Old Psihiko " | >> | "Psihiko" | >> | "New Psihiko " |
| c. | áno póli pátras | >> | póli pátras | >> | káto póli pátras |
| | "Upper Patras " | >> | "Patras" | >> | "Lower Patras" |
| d. | áno salmeníko | >> | salmeníko | >> | káto salmeníko |
| | "Upper Salmeniko" | >> | "Salmeniko" | >> | "Lower Salmeniko" |
| e. | mikrós podhjás | >> | podhjás | >> | meghálos podhjás |
| | "Small Podias" | >> | "Podias" | >> | "Big Podias" |

As we can see in (42), most Greek toponyms that constitute retronymic/neonymic pairs include the modifiers "old" and "upper" that systematically determine retronyms while "new" and "lower" systematically determine neonyms. This holds even in cases where we would expect a different distribution of modifiers. A representative example is the toponym "Salmeniko" in (42d) where although the new region "Lower Salmeniko" is placed to the north of the old region "Upper Salmeniko" it is modified by "Lower". Apart from locative and temporal modifiers, toponyms that display retronymic/neonymic pairs can also appear with alternative modifiers, like "small' and "big", as in (42e). In all these cases, we have retronymy as toponyms concern areas that are geographically adjacent or close to each other.

However, there are toponyms in both English and Greek that only display neonyms and protonyms in their paradigm, but not retronyms:

| | | | |
|---|---|---|---|
| 43. | a. néa filadhélfia >> | filadhélfia >> | *paleá filadhélfia |
| | "New Philadelphia" | "Philadelphia" | "*Old Philadelphia" |
| | b. néa zmírni >> | zmírni >> | *paleá zmírni |
| | "New Smyrna" " | Smyrna" | "*Old Smyrna" |
| | c. New York >> | York >> | *Old York |
| | d. "New England" | England | "*Old England" etc. |

In these cases, we do not have retronymy, as the toponyms concerned refer to cities or towns (founded by immigrants or refugees) which are geographically remote from each other and so there is no risk of ambiguity or confusion. Therefore, here, the neonymic toponyms do not operate as retronymy instigators.

## 6. Conclusions

In this paper, we discussed the sense relation of retronymy. In sections 1 and 2, we introduced the phenomenon, presented the relevant data and reviewed the relevant literature. Building on the analyses of Anastasiadi-Symeonidi (2001) and, mostly, Xydopoulos (2008, 2009), we focused on the phases of the generation of retronyms. In sections 3.1, 3.2 and 3.3, we reestablished the fact that the old and the new terms are related to each other through hyponymy/taxonymy, on the vertical axis. On the horizontal axis, we claimed that the terms are related through oppositeness, and in particular through morphological and privative polarity. Furthermore, we understood "generalization" of the old basic term as its transformation into an autohyponym/autotaxonym. In section 3.4, we understood the recursive character of retronymy as twofold: (a) as a redetermination of autohyponymy and (b) as a transformation of retronyms into neonyms. In section 3.5, we classified English and Greek retronyms/neonyms in at least three morphological types: compounds (already lexicalised), quasi-compounds (in the process of lexicalisation) and ordinary NPs. In section 3.6, we examined the process of abbreviating retronymic/neonymic expressions by deleting the headword and keeping the modifier. In section 4, we claimed that abbreviation is an indication that the abbreviated retronymic/neonymic expression has been transferred from special to general vocabulary. Finally, in section 5, we flagged loan translation as the basic strategy for importing English retronyms/neonyms into Greek and we examined a few cases of indigenous Greek retronyms/neonyms.

## Acknowledgments

We would like to thank the audiences of ISTAL-19 (2009) in Thessaloniki, Greece, and of Journées du CRTT 2012 in Lyon, France, where this work was presented for their precious comments. We are also grateful to Angela Ralli, Vincent Renner and an anonymous reviewer for useful suggestions. The usual disclaimers apply.

## References

AHMAD K. (2000), Neologisms, Nonces and Word Formation, In HEID U., EVERT S., LEHMANN E., ROHRER C. (eds.), *The 9th EURALEX Int. Congress,* 8-12 August 2000, Munich. Vol II., 711-730, Munich, Universität Stuttgart.

AHMAD K., COLLINGHAM S. (1996), Renewable Terminology, In GELLERSTAM M. *et al.* (eds.), *EURALEX '96: Proc. I-II, Part II - Papers submitted to the Seventh EURALEX International Congress on Lexicography in Göteborg, Sweden*, 759-771, Göteborg, Göteborg University.

ANASTASIADI-SYMEONIDI A. (1986), *Neology in Modern Greek Koine*, Thessaloniki, Epistimoniki Epetirida Filosofikis Scholis. [in Greek]

ANASTASIADI-SYMEONIDI A. (2001), Rebaptism, In *Hellenic Language and Terminology 3rd Conference Papers*, 63-77, Athens, Hellenic Society for Terminology. [in Greek]

BABINIOTIS G. (1993), The Linguistic Side of Scientific terms, *To Vima Newspaper* 20/6. [in Greek]

BABINIOTIS G. (1994), *Language as value: The case of Greek*, Athens, Gutenberg. [in Greek]

BLÖHDORN L. M. (2008), *Von 'Rock Stars' und 'Snail Mail': Retronyme als morphologisches Phänomen in der kulturellen Semantik*, handout of a talk given in Disputationsvortrag on 16 January 2008, University of Kiel.

BOOIJ G. (2005), Construction-dependent morphology, *Lingue e Linguaggio*, Vol. 2, 163-178.

BOOIJ G. (2007), *The Grammar of Words: An Introduction to Morphology* (2nd edition), Oxford, Oxford University Press.

BOOIJ G. (2010), Construction Morphology, *Language and Linguistics Compass*, Vol. 3 (1), 1-13.

CABRÉ M. T. (1999), *Terminology: Theory, Methods, and Applications*, Amsterdam, Benjamins.

CHRISTOFIDOU A. (2010), *Neologism as concept naming. Retronymy and word Boundaries*, paper presented at the Meeting of the Hellenic Onomastics Association. [in Greek]

CRUSE A. (1986), *Lexical Semantics*, Cambridge, Cambridge University Press.

CRUSE A. (2004), *Meaning in language: An introduction to semantics and pragmatics* (2nd edition), Oxford, Oxford University Press.

CRUSE A. (2011), *Meaning in language: An introduction to semantics and pragmatics* (3rd edition), Oxford, Oxford University Press.
CURZAN A., ADAMS M. (2008), *How English works: A linguistic introduction* (2nd Edition), New York, NY: Pearson Education.
CRYSTAL D. (2008), *A Dictionary of Linguistics and Phonetics* (6th edition), Oxford, Blackwell.
DERRIDA J. (1981), *Positions*, Chicago, IL., Chicago University Press.
FABB N. (2001), Compounding, In SPENCER A., ZWICKY A. M. (eds.), *The Handbook of Morphology*, Oxford, Blackwell.
FROMKIN V., RODMAN R., HYAMS N. (2003), *An Introduction to Language* (7th Edition), Boston, MA, Thomson.
GIEGERICH H. J. (2005), Associative adjectives in English and the lexicon-syntax interface, *Journal of Linguistics*, Vol. 41 (3), 571-591.
GIEGERICH H. J. (2009), The English compound stress myth, *Word Structure*, 2.1, 1-17.
GILLAM L., TARIQ M., AHMAD K. (2005), Terminology and the construction of ontology, *Terminology*, Vol. 11 (1), 55-81.
HARTMANN R. R. K., JAMES G. (1998), *Dictionary of Lexicography*, London, Routledge.
HORNBY A. S. (2000), *Oxford Advanced Learner's Dictionary*, Oxford, Oxford University Press.
KARAMAN B. I. (2008), On Contronymy, *International Journal of Lexicography*, Vol. 21 (2), 173-192.
LAZANA E. (2012), *The phenomenon of retronymy in Greek: A lexicological approach*, unpublished MA Thesis, Patras, University of Patras, Department of Philology. [in Greek]
LIEBER R. (2005), English Word-Formation Processes. Observations, Issues, and Thoughts on Future Research, In ŠTEKAUER P., LIEBER R. (eds.) *Handbook of Word-Formation*, 375-427, Berlin, Springer.
LYONS J. (1977), *Semantics* (vol. 1), Cambridge, Cambridge University Press
LYONS J. (1999), *Linguistic Semantics*, Athens, Patakis. [in Greek]
NAVIGLI R., VELARDI P., GANGEMI A. (2003), Ontology Learning and Its Application to Automated Terminology Translation, *Intelligent Systems, IEEE*, Vol. 18 (1), 22-31.
OLSON J. M. (2001), The Effects of GIS and Digital Mapping on Cartographic Vocabulary, *Proceedings of ICC2001*.
PLAG I. (2003), *Word-formation in English*, Cambridge, Cambridge University Press.
RALLI A. (2003), Morphology in Greek Linguistics: The State-of-the-art, *Journal of Greek Linguistics*, Vol. 4, 77-130.
RALLI A. (2005), *Morphology*, Athens, Patakis. [in Greek]
RALLI A. (2007), *Compounding in Modern Greek: A Cross-linguistic Morphological Approach*, Athens, Patakis. [in Greek]
RALLI A. (2012), *Compounding in Modern Greek*, Dordrecht, Springer.
RALLI A. (to appear), Compounding and its locus of realization: Evidence from Greek and Turkish, to appear in *Word Structure*.

ROLDÁN-VENDRELL M., FERNÁNDEZ-DOMÍNGUEZ J. (2012), Emergent neologisms and lexical gaps in specialized languages, *Terminology*, Vol. 18:1, 9-26.

SAFIRE W. (2007), On language: Retronymy, *New York Times Magazine,* 18.11.2007.

SMITH L. (2003), New words for old times, *Wisconsin State Journal,* 23.11.2003.

SOANES C. (2005), World of words: ARBs (acronyms, retronyms, backronyms), *Ask Oxford.com* (http://www.askoxford.com/worldofwords/bubblingunder/archive/bubbling_09/?view=uk). [25 February 2009].

SPENCER A. (2005), Word-formation and syntax, In ŠTEKAUER P., LIEBER R. (eds.), *Handbook of Word-Formation*, 74-97, Berlin, Springer.

TSAKONA V. (ed.) (2007a), *Glossary of sociolinguistics: Bilingualisation of Peter Trudgill's A Glossary of Sociolinguistics*, Athens, National and Capodistrian University of Athens. [in Greek]

TSAKONA V. (2007b), Bilingualisation in practice: Terminological issues in bilingualising a specialised glossary, *International Journal of Lexicography*, Vol. 20 (2), 119-145.

TSAKONA V. (ed.) (2008), *Glossary of semantics and pragmatics: Bilingualisation of* Alan Cruse's *A Glossary of Semantics and Pragmatics*, Athens, National and Capodistrian University of Athens. [in Greek]

XYDOPOULOS G. J. (2002), Problems of translating of linguistic terms from English to Greek, *Greek Language Studies*, Vol. 22, 495-506. [in Greek]

XYDOPOULOS G. J. (2004), Translating a Scientific Paper: Problems and possible solutions, *Proceedings of the International Conference Translating in 21st Century: Trends and Perspectives*, 740-747. [in Greek]

XYDOPOULOS G. J. (2008), *Lexicology: Introduction to the analysis of the word and the lexicon*, Athens, Patakis. [in Greek]

XYDOPOULOS G. J. (2009), *Retronymy or when technology meets language*, paper presented at ISTAL-19, 3-5 April 2009, Thessaloniki, Greece.

XYDOPOULOS G. J. (2011), Terminology of linguistics in Greek: Studying the case of semantics, In XYDOPOULOS G. J., BALTAZANI M., TSAGALIDIS T., GIANNAKIS G. K. (eds.), *12 linguistics papers: Ioannina 2004-2008*, 127-142, Athens, Kodyli. [in Greek]

# Partie 2

## IMPLANTATION, CIRCULATION ET DIFFUSION DES NEOLOGISMES

# Pour une néologie sociale

Jean Quirion[1]

La présente contribution s'intéresse au décalage, parfois perçu, parfois réel, entre les propositions néologiques des aménagistes et le milieu qu'elle vise. Elle a pour objectif de suggérer une façon d'atténuer cette différence. Cette dernière est illustrée de manière amusante dans l'article *néologisme* du dictionnaire d'Antidote (2013), qui se lit comme suit : « [MÉDECINE] Mot incompréhensible créé par un malade mental »… Évidemment, la marque technolectale restreint clairement l'usage du terme, mais la définition amuse néanmoins. Un sourire, parfois moqueur, nait également sur les lèvres de plusieurs locuteurs lorsqu'ils prennent connaissance de créations qui leur semblent plus ou moins appropriées.

Dans le contexte d'aménagement linguistique dans lequel s'expriment nos réflexions, la mauvaise perception de propositions jugées insolites entraine des conséquences négatives à divers égards. Parmi celles-ci, on note tout d'abord une efficacité diminuée tant des politiques que des organismes qui la régissent. À une époque où les investissements publics se font rares, spécialement dans un secteur comme le nôtre, cette situation est déplorable. La perte progressive de la confiance que les locuteurs entretiennent envers leur langue et sa capacité à dénommer pertinemment les nouvelles réalités, générales ou spécialisées, s'inscrit également dans la liste des effets indésirables. Chez le sujet parlant, un terme étranger peut se voir transférer les qualités de clarté et d'adéquation qui garantissent à ses yeux une communication efficace.

Le présent ouvrage est construit autour de trois thèmes : la détection, la circulation et l'implantation des néologismes. Le premier fait l'objet d'une attention soutenue depuis environ deux décennies (Kageura et Umino, 1996 ; Drouin, 2007 ; Halskov et Barrière, 2008 ; Vintar, 2010 ; Gómez-Guinovart, 2012). Des méthodes automatisées ont été mises au point pour le dépistage, champ qui relève de la linguistique informatique. Des organismes nationaux et internationaux veillent par ailleurs au recensement, à la codification et à l'étude des néologismes parmi lesquels NEOROM pour les langues romanes, NEOROC pour les idiomes de la péninsule hispanique, NEOXOC pour le catalan, Antenas Neológicas pour l'espagnol de l'Amérique latine, auxquels s'ajoutent d'autres structures parfois informelles, notamment entre des universités.

Les sujets de l'implantation et, corolairement, de la circulation des néologismes retiendront plutôt notre attention. Notre propos s'insère dans un contexte de

[1] École de traduction et d'interprétation, Université d'Ottawa, jquirion@uottawa.ca

modifications des usages, où des langues entrent en contact sur un même territoire. Implicitement, l'implantation est liée à la circulation, en ce sens qu'une création qui prend de la vigueur voyage nécessairement au sein de divers réseaux.

Notre réflexion s'inscrit dans une perspective d'aménagement linguistique et terminologique, qui peut originer d'un gouvernement, d'une organisation ou d'une communauté. Voyons-en successivement des exemples. Le cas sans doute le mieux connu est celui d'États ou de nations, comme la France, la Catalogne, l'Australie, l'Afrique du Sud ou le Québec. Ils ont choisi d'influencer l'équilibre des langues parlées à l'intérieur de leurs frontières ; leurs politiques linguistiques concrétisent leurs intentions. En ce qui concerne les organisations, voyons l'exemple de l'Université d'Ottawa. Cet établissement, situé dans la province anglophone de l'Ontario, mais limitrophe du Québec, a centré son image de marque sur le bilinguisme. Dans les faits, la proportion de francophones y diminue régulièrement. L'université se soucie donc d'afficher un visage français autant que faire se peut. Il y a environ deux ans, le recteur, ancien politicien fédéral, s'est ouvert un compte Twitter. Évidemment, une telle initiative vise surtout à se rapprocher des étudiants, et les relations publiques de l'université veillent depuis à faire connaitre les interventions du recteur. Or, si l'emprunt de *tweet* s'impose chez certains francophones, on conviendra que le terme n'entretient pas l'image francophile de l'établissement d'enseignement. Les services linguistiques de l'université s'assurent donc d'exprimer chaque fois le concept en français, même si un équivalent ne prévaut toujours pas. Voilà un type d'aménagement émanant d'une institution, bien qu'elle n'adopte pas de politique officielle. Enfin, il existe une troisième et dernière catégorie d'aménagement, plutôt particulière, celle d'une communauté linguistique, tel que relaté par Baxter (2009). Le chercheur cite l'exemple des brittophones qui contribuent à l'encyclopédie en ligne Wikipédia. L'auteur rapporte d'abord que le centre de terminologie bretonne TermBret s'est largement consacré à l'établissement du vocabulaire pour les matières enseignées jusqu'à l'aube de l'université. La terminologie scientifique et technique nécessaire à la rédaction d'articles en breton dans Wikipédia s'avère cependant moins disponible pour des domaines plus spécialisés. Pour y remédier, les internautes bretons ont intégré des néologismes de leur cru au sein de leurs contributions. Du coup, les échanges entre collaborateurs à l'encyclopédie ne concernent plus seulement le contenu des articles, mais également leur forme. Le Web a servi de plateforme pour une gestion spontanée du vocabulaire entre parlants bretons : voilà un exemple remarquable d'aménagement linguistique non coordonné. Nous ne connaissons pas d'autres cas semblables, bien qu'il en existe plausiblement. L'intervention d'une communauté de locuteurs sera discutée ultérieurement comme solution possible au rapprochement entre les néologues et leur public cible. Cependant, de façon générale, l'aménagement par un gouvernement retiendra surtout notre attention.

**Implantation, certes, mais de quoi?**

La nature même de ce qui doit être implanté – et mesuré – se modifie. L'univocité était à l'honneur au moment des premières tentatives d'aménagement et d'implantation. Le courant théorique initié par Wüster était largement reconnu et valorisé ; de nombreux travaux intégraient ses principes. La variation était découragée et l'enracinement d'une seule unité était souvent désirée (Aléong, 1979 ; Aléong, Chrétien, Ostiguy et Martin, 1981 ; Chrétien, 1981; Aléong et Jourdan, 1981 ; Daoust, 1981a ; Maurais, 1984). La poursuite des réflexions, au fil des années, a enrichi les perspectives. Ces dernières s'élargissent, car on souhaite l'installation, non pas tant d'un vocable précis, mais l'extension du recours à une langue sur le plan terminologique – et cela inclut l'oral. La reconnaissance et l'acceptation de la variation tendent, croyons-nous, à assouplir la notion d'implantation. Ainsi, une autorité linguistique gouvernementale visera la généralisation de l'idiome plutôt que de concentrer ses efforts sur des termes particuliers. La promotion de la langue prime : si un mot autochtone est privilégié, peu importe qu'on retienne celui-là ou un autre.

La première section exposera succinctement l'avènement et la nature de la terminométrie, ses méthodes et ses apports en procédant à une brève recension des écrits. La seconde partie dressera un bilan des réponses fournies par la terminométrie à des questions fondamentales sur l'implantation. À la lumière de ces constats, le troisième pan proposera le recours aux réseaux sociaux et à l'externalisation ouverte pour réduire la distance entre les néologues professionnels et les publics spécialisés.

## 1. La terminométrie ou la mesure de l'implantation

Les années 1970 et 1980 voient l'établissement des méthodes de travail en terminologie et en néologie. Bien entendu, ces disciplines existaient auparavant, mais elles n'avaient pas été l'objet d'une attention aussi marquée. À l'époque, les entreprises et les États ont commencé à investir, parfois massivement, dans la langue. Les premières convoitent surtout l'efficacité communicationnelle et les seconds affichent leur volonté aménagiste. Très tôt dans ce processus, les spécialistes sont mis à contribution pour le recensement des appellations et la description de l'idiome de spécialité. En premier lieu, leurs textes ont constitué – et représentent toujours – le pain et le beurre des terminologues. Les experts forment en outre un rouage capital des commissions ministérielles de terminologie instituées dans plusieurs pays. Leur double connaissance des concepts en jeu et de la langue de leur discipline leur donne un rôle central dans ces organes. Cette époque voit également les balbutiements de la terminométrie (Fainberg, 1974 ; Daoust, 1981a ; Daoust, 1981b ; Maurais, 1987 ; Daoust, 1987 ; Benhamida, 1989), signe avancé du souci de la bonne implantation des formes recensées ou inventées.

Pour leur part, les années 1990 ont été témoins de deux avènements majeurs qui ont bouleversé à peu près simultanément le domaine : la terminotique et Internet. L'arrivée du réseau des réseaux a signifié un accès nettement plus large à de la documentation électronique et donc à la création de corpus. L'interrogation de plus en plus aisée de ces ensembles de textes a modifié durablement la façon d'exercer le travail terminologique. À titre d'exemple, citons le cas des recherches ponctuelles, indispensables en néologie. Les linguistes sont passés d'une quête monacale dans des imprimés à des interrogations instantanées. Justement, l'objet textuel est devenu le cœur de maintes applications informatiques pour en retirer la substantifique moelle : concordanciers et extracteurs d'abord unilingues, puis bilingues ; logiciels de terminologie maison ; etc.

La plus grande disponibilité de documents électroniques et la mise au point d'outils pour les traiter créent un terreau fertile à la croissance de la terminométrie. Au début des années 1990, la Délégation générale à la langue française (DGLF ; aujourd'hui Délégation générale à la langue française et aux langues de France [DGLFLF]) émet un appel d'offres original invitant les universités du pays à mesurer l'implantation terminologique des avis officiels. Ces derniers sont rédigés par des commissions de terminologie où œuvrent de concert linguistes et spécialistes. L'idée a porté ses fruits, car six établissements ont manifesté de l'intérêt ; leurs efforts ont jeté les bases d'une réflexion plus importante sur l'implantabilité. Cette initiative aura marqué l'histoire de la terminométrie, lui faisant gagner rapidement ses lettres de noblesse. Des publications ont suivi ; deux numéros de *Terminologies nouvelles* (Delavigne et Gaudin, 1994 ; Rouges-Martinez et Villebrun, 1997) ont été consacrés aux travaux effectués, puis un livre a été lancé en 1997 (Depecker et Mamavi). Les Québécois n'ont pas été en reste et ont fait paraitre coup sur coup trois enquêtes dans le domaine de l'éducation (Martin, 1998), des pêches (Auger, 1999) et du transport (Quirion, 2000 ; Quirion, 2003a). Aujourd'hui, des thèses de doctorat sur la terminométrie ont été menées ou sont en cours dans diverses régions, dont la Turquie (Karabacak, 2009a ; Karabacak, 2009b), la Catalogne (Montané, 2012), le Canada, le Maroc, etc.

Les enquêtes terminométriques ont recours à des méthodes diverses. Les trois principalement utilisées sont l'étude ethnolinguistique, les questionnaires et les entrevues, puis l'analyse de corpus. Chaque approche livre des fragments qui contribuent de fil en aiguille à mettre au jour l'importance relative des variables d'implantation.

L'étude ethnolinguistique par l'observation, participante ou non, renseigne sur l'usage réel par opposition au déclaré (le premier étant plus précis que le second). Ce type de recherche donne de surcroit des informations pertinentes sur la diffusion des néologismes. Par exemple, les attitudes, les comportements, les canaux de communication ainsi que les contextes d'échange officiels et informels peuvent être analysés par l'étude ethnolinguistique. Cette méthode

exige beaucoup de temps, car elle suppose une immersion prolongée dans le milieu dans lequel évoluent les locuteurs. Cette nécessité explique que de rares enquêtes ont eu recours à cette approche à ce jour ; nommons-en deux (Heller, Bartholomot, Lévy et Ostiguy, 1982 ; Vila i Moreno, Nogué Pich et Vila i Moreno, 2007). Produites à 25 ans d'écart, ces recherches ont respectivement eu cours dans une grande entreprise brassicole à Montréal et chez des adeptes de sports plutôt marginaux (escrime, escalade et hockey sur gazon) en Catalogne.

Les questionnaires et les entrevues constituent la deuxième méthode de collecte de données. Ils offrent des informations précieuses sur la réception des néologismes, utiles pour prédire leur implantation : attitudes, perceptions et opinions sont révélées par les sujets parlants. Ces derniers font également état de leurs façons de s'exprimer ; à la différence de l'étude ethnolinguistique, le questionnaire ou l'entrevue ne livre pas l'usage réel, mais bien l'idée que s'en fait le locuteur. Plusieurs travaux ont collecté des données de cette façon au fil des ans (Fainberg, 1974 ; Daoust, 1987 ; Tremblay, 1994; Martin, 1998 ; Vila i Moreno et coll., 2007). Cette dernière publication, déjà citée au paragraphe précédent, décrit également la méthode ethnolinguistique par l'observation. Notons que les deux types d'enquête déploient un protocole exigeant ; pour cette raison, la quantité de termes considérés est ordinairement restreinte à un petit nombre (d'une poignée à quelques dizaines d'unités).

L'analyse de corpus constitue la troisième façon pour mesurer l'implantation de néologismes. À la différence des deux premières, elle porte sur un ensemble allant jusqu'à plusieurs centaines de vocables. Elle présente des données factuelles et objectives sur l'usage. Coutumièrement centrée sur le concept, l'approche permet d'examiner la grappe de termes qui le désignent. Comme elle repose sur des textes, l'exploration s'automatise, du moins partiellement. Une telle recherche offre un avantage unique : celui de pouvoir établir une diachronie à rebours, c'est-à-dire que, s'il est possible de constituer un corpus pour une période écoulée, on peut alors y étudier l'implantation. Encore ici, plusieurs recherches ont eu recours à cette méthode (Auger, 1999 ; Quirion, 2003a ; Quirion, 2003b ; Karabacak, 2009a ; Karabacak, 2009b ; Montané, 2012).

Dans une certaine mesure, l'analyse de corpus peut également éclairer sur les attitudes. Afin d'illustrer notre propos, présentons un exemple d'enquête sur un petit groupe d'unités autour de la notion « clavardage »[2]. Le *grand dictionnaire terminologique* la définit comme suit : « Activité permettant à un internaute d'avoir une conversation écrite, interactive et en temps réel avec d'autres internautes, par clavier interposé ». Un certain nombre de dénominations sont en usage, dont trois plus couramment : *clavardage, tchat* et *chat*. Ont d'abord été

[2] Nous remercions Robin Matte, étudiant en traduction à l'Université d'Ottawa, qui a mené l'étude terminométrique sous notre supervision.

confirmées les fortes utilisations présumées de *clavardage* et de ses dérivés au Canada d'une part et, d'autre part, de *tchat* et *chat* et de leurs dérivés en Europe. Des corpus de périodiques ont été constitués pour l'étude, ce qui a permis de remonter jusqu'en 1995 ; voilà un exemple de l'avantage diachronique tout juste évoqué. De surcroit, afin de révéler les attitudes des journalistes, les marques métalinguistiques entourant le recours à ces formes, révélatrices d'un usage timide, ont été systématiquement notées. Les marques les plus courantes sont les guillemets, l'italique et des formules de distanciation (*comme disent/comme l'appellent [les internautes/certains, etc.]*). Même si l'enquête s'avère modeste, ses résultats ne manquent pas d'intérêt. En premier lieu, pour la dernière période étudiée (2005-2010), *clavardage* (au Canada) et le duo *chat/tchat* (en Europe) affichent tous deux un coefficient d'implantation de 0,96. Cela signifie que ces termes sont retenus 96 fois sur 100 par les journalistes, respectivement canadiens et européens, lorsqu'ils nomment cette activité dans leurs articles[3]. En second lieu, les usages timides apparaissent uniquement dans les textes européens, dans une proportion de 62 %. Cela revient à dire que *chat* et *tchat* sont choisis avec une certaine réserve plus de trois fois sur cinq par les rédacteurs de l'Europe, tandis qu'aucune insécurité linguistique ne semble hanter les auteurs québécois dans leur recours à *clavardage*. Bien que notre exemple ne concerne que l'écrit, ces résultats terminométriques se révèlent utiles pour la néologie comme pour l'aménagement.

Les bénéfices de la terminométrie se trouvent aussi du côté des politiques linguistiques. L'aménagement s'inscrit dans un processus qu'il faut naturellement évaluer. La terminométrie répond justement à ce besoin. Elle offre un regard critique sur le travail néologique effectué, sur les méthodes de création et de diffusion, de même que sur les fruits des dispositions mises en place. Or, que donne l'expérience terminométrique tentée depuis bientôt 20 ans, c'est-à-dire depuis le tournant initié par la DGLF, voire 40 ans si l'on remonte à la première étude connue, soit celle de Fainberg (1974)?

## 2. Quelles connaissances la terminométrie livre-t-elle sur l'implantation de néologismes?

À ce jour, l'investissement dans la mesure de l'implantation a rapporté plusieurs dividendes. Il éclaire sur l'implantabilité en proposant des méthodologies, en fournissant des moyens de validation et en présentant des données qualitatives et quantitatives. Tout d'abord, les premières émanent des études ethnolinguistiques et des questionnaires et entrevues. Leur analyse fait clairement ressortir que le choix lexical représente d'abord et avant tout une décision sociale. La portée interpersonnelle des termes et des mots utilisés, tout comme l'adéquation de ces derniers aux diverses situations de communication, prime chez les locuteurs. Nous reviendrons ultérieurement sur ce constat. Ces

[3] En fait, la variante *chat* présente un coefficient d'implantation de 0,52 et *tchat* de 0,44 ; c'est la somme des usages des deux variantes orthographiques qui s'élève à 0,96.

études soulignent également l'importance d'autres facteurs de variation parmi lesquels les modes de diffusion des innovations et les attitudes jouent un rôle prépondérant. Ensuite, les approches quantitatives regroupent l'analyse de corpus ainsi que les questionnaires et entrevues. Grâce à elles, des portraits d'ensembles vivants, souvent néologiques, sont tirés. En revanche, il s'agit d'images figées ; pour qu'elles se mettent en mouvement, des mesures répétées sont requises. Néanmoins, ces clichés ont révélé jusqu'ici la langue usuelle des innovations pour des domaines et des territoires donnés. Les résultats indiquent que la proximité phonétique, orthographique et syntaxique avec le terme allophone bonifie les chances de succès d'un néologisme. À cela s'ajoutent certaines caractéristiques partagées par les vocables mieux implantés, dont la dérivabilité, le faible nombre de concurrents et la brièveté. Par ailleurs, il semble que les termes qui bénéficient de l'appui d'un avis de normalisation ou de recommandation s'enracinent plus aisément. Mentionnons finalement un apport technique, soit la création d'un prototype informatique pour mesurer l'implantation, capable de désambigüiser semi-automatiquement les unités polysémiques ; dans l'éventualité où cette invention deviendrait accessible à tous, les évaluations quantitatives seraient largement accélérées (Quirion, 2005). Dans l'ensemble, on constate que les portraits dressés par la terminométrie se comparent difficilement à cause de la diversité des méthodes. Malgré les percées, beaucoup de travail reste à accomplir.

Les stratégies d'intervention linguistique, qui ont commencé à voir le jour dans les années 1970, ont aussitôt donné lieu à un certain nombre de questions, dont deux fondamentales :

- Comment améliorer la diffusion et la circulation des travaux terminologiques?
- Comment prédire le succès ou l'échec d'un néologisme?

Esquissons le bilan des réponses apportées à chacune de ces interrogations.

### 2.1. Comment améliorer la diffusion et la circulation des néologies?

À notre connaissance, la diffusion et la circulation des travaux néologiques n'ont fait l'objet que de peu d'études, sauf peut-être sur le résultat – et non sur le processus. La dissémination des unités terminologiques constitue pourtant la condition de base à leur adoption. Pour le confirmer, deux initiatives récentes seront mentionnées en commençant par celle de Vila i Moreno et coll. (2007). Ces chercheurs ont relevé un faible taux d'utilisation des innovations catalanes proposées par Termcat dans trois domaines sportifs ; ils ont tenté d'en découvrir les causes. Une de leurs hypothèses voulait qu'une résistance au changement ou encore une réaction aux politiques linguistiques en vigueur expliquât ce phénomène. Ces suppositions ont été invalidées, tout comme celle concernant un rejet pur et simple des formes suggérées. En réalité, il s'agissait d'une diffusion inefficace : les sportifs ne connaissaient tout simplement pas les

créations de l'organisme catalan. Un autre exemple vient des mécanismes de diffusion néologique de la DGLFLF (Délégation générale à la langue française et aux langues de France, 2010). Dans son rapport annuel de 2010, la Délégation signale que le recours à Internet est couramment pratiqué et donne un échantillon de termes qui sont poussés vers l'internaute au lieu d'être tirés par ce dernier. Ces initiatives s'avèrent louables et nécessaires, bien que ponctuelles et limitées. Peut-on rêver d'une interaction continue avec les spécialistes?

Les réseaux de communication au sein de groupes sociaux ou professionnels deviennent vecteurs d'implantation, car les pairs occupent une place centrale dans la circulation des néologismes. Cette dimension est soulignée par Martin et Loubier (1993 : 88) :

> Les réseaux de communication sont un autre élément dont il faut tenir compte à toutes les étapes d'un processus de changement linguistique planifié et ce, tant au niveau macro-social que micro-social. Au niveau macro-social, notons l'importance de la participation des grands moyens de communication de masse (ex. médias) au processus de changement, mais aussi celle des relations que toute cellule sociale entretient avec son environnement […] La dynamique sociale repose en grande partie sur l'interaction d'un ensemble de réseaux de communication, d'où l'importance que l'on doit attacher à l'étude de ces réseaux.
> À un niveau micro-social, il faut mentionner le rôle capital que peuvent jouer les communications interpersonnelles entre les locuteurs dont il faut étudier les conditions de réalisations dans des situations de communication précises et dans des contextes socioprofessionnels concrets, contextes reliés aux pratiques qui structurent les différentes entreprises et milieux de travail.

Vingt ans plus tard, force est de constater que les projections de Martin et Loubier n'ont pas été exploitées à leur plein potentiel. De fait, le processus de dissémination néologique demeure flou.

### 2.2. Comment prédire le succès ou l'échec d'un néologisme?

De toute évidence, l'ingrédient magique assurant qu'un soufflé lexical va lever n'a pas encore été inventé. Spontanément, les linguistes de tous horizons ont disséqué les néologismes pour le trouver. Ils en ont précisé les caractéristiques formelles, comme sa capacité à s'insérer dans une série paradigmatique ou à générer des dérivés, sa féminisation, etc. Tout bien considéré, si les aptitudes formelles des termes s'avéraient vitales, on pourrait prédire avec une certaine exactitude qu'une proposition s'inscrira ou pas dans l'usage. Or, ce n'est pas le cas. Tout au plus peut-on constater que les facteurs linguistiques contribuent à l'implantation de néologismes, mais qu'ils n'en sont généralement pas la cause profonde. Comme il a été mentionné plus haut, les sujets parlants prennent d'abord en compte l'effet de leurs choix lexicaux sur leurs interlocuteurs.

Des éléments extralinguistiques influencent aussi le sort d'une proposition. L'acceptation néologique évolue dans le temps et est également tributaire des situations de communication. En ce qui concerne le premier point, la terminométrie aura déterminé que le temps requis varie considérablement d'un secteur à l'autre. Prenons deux exemples québécois. D'abord, le domaine du golf se francise posément depuis plus de 40 ans[4]. D'un autre côté, des études par corpus en courte diachronie sur la nanotechnologie montrent une implantation très rapide des nouveaux termes français (Quirion, 2011). On peut émettre l'hypothèse que l'évolution lente pour ce qui est du golf s'explique par la présence préalable de formes anglaises fortement ancrées. À l'opposé, les néologismes en nanotechnologie ont surgi au moment où les concepts apparaissaient. Quant aux situations de communication, des méthodes doivent être établies, comme celle de Leblanc et Bilodeau (2009) pour mesurer les usages à l'oral. Le pluriel est de mise à *usages*, car cohabitent souvent, chez un même sujet parlant, des terminologies qui s'adaptent en fonction du contexte de communication, de l'interlocuteur, de sa position hiérarchique, etc. (Heller et coll., 1982). Montané (2012) signale justement une meilleure implantation à l'écrit des formes suggérées par Termcat lorsqu'elles sont accompagnées de variantes courtes telles que des abréviations ou des sigles. En outre, il faudra vraisemblablement modeler les innovations aux situations de communication à l'oral, largement négligé, car la néologie aménagiste exprime une nette préférence pour l'écrit. Cette attitude affiche néanmoins sa cohérence avec la plupart des politiques linguistiques, qui visent un changement chez les institutions – et non les individus. Tout compte fait, il demeure pour le moins hasardeux de prédire le sort d'un néologisme.

### 2.3. Connaissances insuffisantes de l'élément social

À la lumière de ce qui précède, les questions originelles sur l'implantation et la circulation des nouveaux vocables demeurent d'actualité. Certes, les enquêtes de plus en plus nombreuses menées au fil des ans lèvent progressivement le voile sur les facteurs influençant l'ancrage terminologique. Néanmoins, les recherches doivent impérativement être poursuivies et partagées ; cela souligne à nouveau la pertinence du thème du présent ouvrage. L'aspect formel des matrices néologiques est bien décrit ; on ne peut en dire autant de la diffusion et de la circulation des innovations. La réception des créations lexicales au sein des réseaux de communication et des groupes visés ne l'est pas davantage. Tout bien considéré, le cumul d'études terminométriques indique l'importance cruciale d'approfondir la connaissance des facteurs extralinguistiques (sociaux, culturels, cognitifs, historiques, politiques, etc.) et spécialement des comportements, opinions et attitudes liés aux néologismes.

---

[4] Recherches personnelles de l'auteur, non publiées.

## 3. Une distance à réduire entre néologues et sujets parlants

En vue d'atteindre l'objectif que nous venons d'énoncer, il existe divers moyens de réduire la distance entre les néologues et les sujets parlants. Nous tenterons de démontrer que le succès néologique passe par un meilleur entendement des aspects extralinguistiques, que peuvent apporter un certain nombre d'autres disciplines. Des modes de rapprochement entre néologues et sujets parlants seront ensuite suggérés.

### 3.1. Apports interdisciplinaires

Diverses disciplines peuvent renseigner sur les questions soulevées. De tout temps, les champs de connaissances se sont nourris mutuellement. Leurs interactions sont aujourd'hui encouragées par la facilité d'accès aux productions scientifiques mondiales et la néologie à tout à gagner à multiplier les ponts avec ces autres spécialités. Présentons quelques exemples pour illustrer notre propos, choisis dans des domaines pouvant apporter un éclairage sur l'aspect social qui échappe plus aisément à l'approche linguistique.

Depecker (2013 : 18) enregistre pour sa part un plaidoyer pour l'ethnoterminologie :

> Il est évidemment primordial d'observer les répercussions linguistiques des évènements et de considérer dans cette perspective l'évolution des langues. Mais cela revient à observer les choses de façon principalement externe, sans rendre véritablement compte des motivations des sujets parlants. Et pour cela, il faut entrer dans un mode de description plus détaillée, capable de prendre en compte les dimensions sociologiques, culturelles, psychologiques, bref ethnologiques.

S'inspirant logiquement de l'ethnolinguistique, cette proposition livre un regard pénétrant sur l'humain, au centre des activités d'émission et de réception terminologiques, au sein de groupes significatifs. L'ethnoterminologie, tout comme la théorie de la diffusion des innovations, offre le potentiel de mettre au jour la circulation néologique. La seconde avance justement qu'une nouveauté doit afficher une compatibilité avec les normes du groupe pour être adoptée par un individu (Rogers, 2003) ; ces codes doivent impérativement être connus du néologue s'il doit s'y conformer. Parmi les autres enseignements, Rogers mentionne en outre que toute personne visée par une innovation doit y percevoir un avantage relatif, ce qui renforce l'intérêt d'un dialogue permanent avec les destinataires. Signalons également que cette théorie pourrait éclairer sur les canaux de communication, car elle décrit les apports respectifs de canaux médiatisés et de communications interpersonnelles. Évidemment, s'ajoute la congruence de la sociolinguistique qui fait valoir, depuis longtemps déjà, que la connaissance des réseaux de communication sociale représente un élément clé pour la promotion du changement. Cette contribution est renforcée par Martin et Loubier (1993 : 89) :

> En situation d'aménagement linguistique, retenons qu'il est impossible de saisir un processus de changement linguistique en dehors de la vie sociale d'une communauté, c'est-à-dire en négligeant les caractéristiques du groupe social, ou encore, en ne respectant pas ses normes socioculturelles, d'où toute l'importance de prendre en considération les facteurs sociaux et psychosociaux qui interviennent, à un moment ou l'autre, dans un processus d'implantation du français.

Les sociolinguistes ont également établi que les variantes stigmatisées sont néanmoins conservées à cause du prestige latent qui leur est rattaché, signal d'appartenance à un groupe socioprofessionnel, créant des situations de diglossie terminologique. Tous ces exemples manifestent clairement l'importance de placer le social au cœur des considérations des néologues.

En substance, les politiques n'agissent pas sur la langue, mais sur les locuteurs (Vila i Moreno et coll., 2007). En effet, il a été démontré dans cette section que la terminométrie a permis à la néologie d'effectuer certains progrès, bien qu'insuffisants. Les études quantitatives doivent impérativement être poursuivies. Mais, à tout prendre, les angles qualitatifs révèleraient sans doute plus les mécanismes qui régissent la sélection lexicale. Les néologues ont logiquement fait porter jusqu'ici leurs efforts surtout sur les attributs formels des innovations dans le but d'en améliorer l'implantabilité. Cette approche partielle s'avère nécessaire, mais incomplète, car les questions d'implantation et de circulation dépassent largement la linguistique. Elles relèvent, compte tenu de ce qui précède, de disciplines qui centrent leurs activités sur la compréhension de l'homme dans son environnement. « Si la terminologie a un caractère social parce qu'elle s'inscrit dans un projet d'aménagement linguistique, il n'est pas du tout certain qu'elle soit pour le moment "sociale" de par sa pratique » (Martin et Loubier, 1993:89). Bien que cette affirmation remonte à une vingtaine d'années, le rapprochement souhaité entre néologues et sujets parlants reste largement à opérer.

### 3.2. Apports des sujets parlants

Parallèlement aux travaux de recherche, les néologues de terrain ont tout avantage à cultiver un étroit contact avec les locuteurs. Car, malgré toute sa bonne volonté, le linguiste demeurera presque toujours quelqu'un d'extérieur au domaine ou perçu comme tel. Auger (2001 : 204) rappelle l'importance de l'adaptation aux destinataires et au contexte communicationnel :

> On oublie le plus souvent que l'usager est maître de ses façons de parler ou d'écrire et de ce qu'il est généralement convenu d'appeler son idiome. Aussi le terminologue ne doit jamais se mettre en situation d'imposer des usages qui sont étrangers aux personnes auxquelles il destine son travail terminologique. Son professionnalisme va le conduire intuitivement à hiérarchiser les usages ou les faits de langue en choisissant ceux qui conviennent le mieux à une situation

> donnée. Il saura, par exemple, qu'il n'est nullement utile de formaliser ou de compliquer les communications de travail interpersonnelles d'ouvriers qui travaillent dans un même atelier de l'usine par l'imposition de termes jugés trop complexes. Par contre, il verra à ce que la terminologie figurant dans un mémoire technique destiné à une commission gouvernementale soit irréprochable au plan terminologique et qu'il puisse atteindre le plus grand nombre d'adhérents. Ici la finalité du message va déterminer l'action du terminologue et le contenu qualitatif des termes qu'il va tenter d'implanter.

Normalement, les relations entretenues par les néologues prennent la forme de consultations ponctuelles de locuteurs spécialisés ; il s'agit de contacts personnels, d'informateurs variés, de collaborateurs aux commissions de néologie et de terminologie, etc. Les néologues ont toujours cultivé des liens étroits avec les experts, dépouillant leurs publications, sollicitant leur avis sur des appellations ou des concepts, sur les recherches documentaires, sur l'implantabilité, etc. Cependant, ces spécialistes n'affichent évidemment pas une disponibilité permanente ; de plus, il s'avère impossible de multiplier les initiatives au point d'obtenir une représentation juste des divers groupes qui forment une communauté spécialisée. Si l'on se reporte à la citation d'Auger, force est de constater que le néologue a traditionnellement concentré ses efforts sur le mémoire destiné à une commission au détriment de l'atelier d'usine. C'est pourquoi il importe d'élargir le spectre des situations de communication prises en compte en augmentant les contacts avec les principaux intéressés. On souhaitera une interaction avec un échantillon plus représentatif des locuteurs afin de renforcer la crédibilité des linguistes et de leurs propositions. On se rapprochera ainsi d'une influence plus marquée auprès de la population visée. La section suivante y va d'une proposition à cet égard.

### 3.3. Le recours aux réseaux sociaux et à l'externalisation ouverte

Afin de rapprocher les créateurs des termes de ceux qui les utilisent, le recours aux réseaux sociaux et à l'externalisation ouverte peut être envisagé (Quirion, 2012). Le néologue ne peut se présenter dans tous les milieux où des besoins de dénomination éclosent ; il peut néanmoins être informé par un ensemble de collaborateurs. Les nouvelles technologies, dont le Web 2.0, peuvent l'aider dans cette démarche. De fait, il existe des applications commerciales qui analysent les sentiments exprimés sur les réseaux sociaux ou dans des publications électroniques (MediaMiser Inc., 2013). Elles examinent les énoncés liés à une marque afin de déterminer semi-automatiquement si la mention envers le produit ou l'entreprise affiche un sentiment positif, négatif ou neutre. Cette technologie a d'ailleurs déjà été utilisée de façon exploratoire pour évaluer la réception de néologismes, avec un certain succès. Bien entendu, les linguistes ne peuvent généralement se payer le luxe de services destinés aux grandes sociétés. En revanche, ils ont accès à un vaste noyau virtuel de collaborateurs bénévoles. Voyons de plus près la nature, les moyens et avantages potentiels des réseaux sociaux et de l'externalisation ouverte.

### 3.4. Réseaux sociaux

Les réseaux sociaux représentent le fantasme des néologues : inconnus il y a quelques années, ils se sont répandus à la vitesse de l'éclair et sont désormais utilisés quotidiennement par des centaines de millions de personnes autour du monde. Voici comment le *grand dictionnaire terminologique* (GDT) définit un réseau social : « Communauté d'internautes reliés entre eux par des liens, amicaux ou professionnels, regroupés ou non par secteurs d'activité, qui favorise l'interaction sociale, la création et le partage d'informations » (Office québécois de la langue française, 2013). Nous avons mis en gras certains mots afin de faire ressortir les nombreux éléments d'intérêt qu'offrent les réseaux sociaux. À la lumière des besoins exposés dans la section précédente, tout ce qui « favorise l'interaction sociale, la création et le partage d'informations » sera accueilli favorablement par le monde néologique.

### 3.5. Externalisation ouverte

L'externalisation ouverte constitue un autre phénomène récemment apparu sur le Web. Le GDT la définit comme suit : « Pratique qui consiste pour une organisation à externaliser une activité, par l'entremise d'un site Web, en faisant appel à la créativité, à l'intelligence et au savoir-faire de la communauté des internautes, pour créer du contenu, développer une idée, résoudre un problème ou réaliser un projet innovant, et ce, à moindre coût. Notes → L'externalisation ouverte est associée à l'idée d'une mise en commun de l'intelligence collective. Elle s'appuie sur l'exploitation directe du potentiel d'innovation de la communauté des internautes, qui constitue un immense réservoir d'amateurs éclairés et de spécialistes compétents pour un domaine donné » (Office québécois de la langue française, 2013). Encore ici, nous avons mis en gras les multiples éléments touchant l'aspect social, interactif. On conviendra aisément qu'un « immense réservoir d'amateurs éclairés et de spécialistes compétents pour un domaine donné » porte une valeur inestimable pour une approche interindividuelle de la néologie.

Trois exemples concrétiseront ce qu'apportent ces technologies à la néologie. D'abord, le site Facebook, qui personnifie certainement le réseau social par excellence. L'entreprise derrière le site a eu recours aux utilisateurs du service pour localiser les multiples versions du produit. Les internautes ont ainsi participé directement au processus terminologique et néologique, inscrivant leurs suggestions pour tel menu, telle fonction ou tel message ou en se prononçant sur celles des autres. La terminologie de cette application commerciale est donc le fruit du travail d'une communauté de cybernautes engagés. Note particulièrement profitable pour les néologues, le système de vote qui préside à la sélection finale assure d'emblée une réception majoritairement favorable des propositions. Ensuite, en ce qui concerne l'externalisation ouverte, la France a été le témoin il y a quelques années d'une initiative bénéfique. En effet, les responsables du site http://www.info-

ecigarette.fr/ ont lancé un concours en mars 2008 pour trouver un néologisme pour désigner l'action de fumer une cigarette électronique. Ils ont reçu 66 propositions de leurs membres, lesquelles ont été soumises au vote. Plus de 600 personnes se seront prononcées et le terme *vapoter* a remporté les honneurs. Vérification faite, pendant le mois de mai 2013, la création parait près de 75 fois dans la presse française selon la base de données Eureka.cc ; la majorité de ses occurrences relevait d'un emploi confiant, c'est-à-dire sans marques métalinguistiques. Encore ici, le néologisme mis de l'avant bénéficiait dès l'origine d'un large appui parmi la communauté susceptible de le reprendre. Enfin, en lien direct avec notre propos, le répertoire terminographique collaboratif TermWiki s'érige en « réseau social mondial d'apprentissage ». TermWiki est, à notre connaissance, la première banque de terminologie ouverte. On y présente la discipline comme un foyer d'activités sociales et intellectuelles. Quiconque peut y créer ou modifier des fiches ; chacune peut d'ailleurs être commentée, offrant ainsi un espace de discussions et de négociations terminologiques. Le mode de fonctionnement de TermWiki tombe clairement sous la coupe de l'externalisation ouverte ; pourtant, la banque se décrit comme un réseau social. Cela illustre le chevauchement entre les réseaux sociaux et l'externalisation ouverte, qui partagent plusieurs caractéristiques.

Dans un monde branché où pullulent des communautés virtuelles et où circulent des unités lexicales, les avantages pour la néologie des réseaux sociaux et de l'externalisation ouverte sont nombreux. Considérés passivement, les réseaux sociaux offrent déjà une plateforme pour le recensement, l'observation et la sélection de nouvelles formes. Mais leur plein potentiel se révèle au moment de diffuser des innovations et de prendre le pouls du public visé. Ils peuvent également servir à diriger les internautes vers des sites d'externalisation ouverte. Ceux-ci peuvent être créés pour la cause ou simplement greffés à un site existant, comme dans l'exemple de la cigarette électronique. Les collaborateurs sont invités à accomplir diverses tâches dans cet espace virtuel. Ils peuvent suggérer des néologismes, donner leur opinion sur des propositions, voter pour celle qu'ils préfèrent ou encore signaler l'apparition de nouveaux termes ou concepts.

### 3.6. Les conséquences d'une adoption des réseaux sociaux et de l'externalisation ouverte

Les conséquences d'une ouverture aux réseaux sociaux et à l'externalisation massive sont multiples. Cela explique sans doute partiellement pourquoi les néologues jouent de prudence dans leur adoption. Seront successivement discutés les effets sur le processus néologique, sur le rapprochement avec les locuteurs, sur la cohérence d'une politique linguistique, sur les langues minorisées ou en voie d'extinction et sur le pouvoir accordé aux néologues. Nos réflexions prendront la forme d'une série de questions.

3.6.1. Sur le processus néologique

Comment les néologues peuvent-ils tirer le meilleur profit de l'énergie et de la force du Web 2.0? Il a été préalablement affirmé que la disponibilité accrue et aisée à des corpus électroniques avait modifié considérablement la méthode terminologique. Or, l'accès facilité à un vaste bassin de spécialistes amateurs ou professionnels compétents et volontaires causera vraisemblablement une décentralisation des décisions, entrainant dans la foulée un ajustement du processus néologique. Dans quelles circonstances choisir respectivement les réseaux sociaux et l'externalisation ouverte?

Comment publiciser son initiative? Quelles structures engageraient les locuteurs de façon optimale? Doit-on prévoir des critères de recrutement? Quelles méthodes, quels outils sont requis pour obtenir leur avis sur des termes, pour détecter, faire circuler et implanter des néologismes dans des domaines et des langues en perpétuel changement?

Peut-on former des groupes de spécialistes qui se prononceraient sur l'implantabilité? Les néologues pourraient s'inspirer des sondeurs, qui viennent de revoir leurs propres méthodes. En effet, ces derniers interrogeaient jusqu'à récemment un échantillon aléatoire de la population ; aujourd'hui, tirant parti des nouvelles technologies, ils créent plutôt un bassin d'internautes volontaires. Du coup, les résultats des sondages Internet sont présentés au public sans être accompagnés d'une marge d'erreur, reflet de ce changement de paradigme. Un vaste regroupement de spécialistes d'un domaine, représentatifs de divers milieux, pourrait-il être consulté de la même façon?

3.6.2. Sur le rapprochement avec les locuteurs

Néologues professionnels et néologues amateurs? Il y a nécessité de poursuivre la réflexion sur l'implantation et sur les relations entre les néologues-linguistes et les autres membres de la communauté linguistique auteurs de néologismes. Nous proposons d'appeler les premiers *néologues professionnels* et les seconds *néologues amateurs*, l'adjectif *amateur* étant ici utilisé sans connotation péjorative.

Jusqu'à quel point le rapprochement avec une base plus large de locuteurs peut-il peser sur leurs choix? L'instantanéité des échanges peut-elle constituer un creuset accélérant le traitement et la diffusion de termes accrocheurs?

L'ouverture au Web 2.0 déterminerait-elle les leadeurs d'opinion, ceux vers qui des efforts particuliers devraient être consentis à cause de leur influence au sein des groupes visés? Les meneurs jouent un rôle clé dans l'adoption d'une innovation (Rogers, 2003).

3.6.3. Sur la cohérence de la politique linguistique

L'aménagement dans un État a jusqu'ici toujours été sous la coupe d'une organisation à vocation linguistique. Assistera-t-on à l'émergence d'une néologie aménagiste semi-structurée, amateur, parallèle – voire concurrente – aux efforts gouvernementaux? Il faudra déterminer comment cette initiative sera mise à profit pour éviter qu'elle ne sape le travail des autorités.

Comment s'assurer de la qualité et de la cohérence des néologismes dans un milieu en apparence chaotique? Le changement de paradigme s'avère considérable ; en effet, on passe d'un environnement contrôlé, régulé et policé à une assemblée citoyenne spontanée et joyeusement anarchique.

Doit-on précisément redouter une pléthore synonymique? Une de nos recherches précédentes (Quirion et Lanthier, 2006) a démontré que la multiplicité des concurrents provoquait un effet néfaste sur l'implantation terminologique. En revanche, selon Temmerman et Van Campenhoudt (à paraitre), la réduction de la synonymie et de la polysémie nuit à l'évolution naturelle des langues : « The effort to eliminate polysemy and synonymy in terminology in order to achieve the unrealistic ideal of univocity and unambiguous communication has been shown to interfere with natural language processes ».

La détection, l'implantation et la circulation supposent-elles un territoire restreint? Internet ne connait pas de frontières, alors que ces dernières fondent la mise en œuvre de la politique d'aménagement linguistique d'un État. Malgré des exemples de collaboration continue, les décisions néologiques nationales sont précédées d'une concertation minimale entre les représentants des variétés d'une même langue ou d'idiomes partageant un fonds commun. Quelle touche le recours aux internautes de toutes régions géographiques pourrait-il apporter?

3.6.4. Sur les langues minorisées ou en voie d'extinction

Comment les réseaux sociaux et l'externalisation ouverte peuvent-ils sauvegarder et dynamiser des langues minorisées ou en voie d'extinction? Facebook offre en 2013 une interface en 50 langues. Rares sont les produits commerciaux qui sont traduits pour autant de marchés lorsque leurs promoteurs ramassent la facture… Pour sa part, TermWiki recense des contributions dans plus de 100 idiomes. Oubliés par les entreprises, les locuteurs qui ne parlent pas les langues les plus répandues revitalisent leurs idiomes grâce à Internet. De la même façon, le Web 2.0 peut-il faciliter la veille et le suivi néologique de langues de faible diffusion, permettant leur étude par une équipe réduite de chercheurs?

3.6.5. Sur le pouvoir des néologues

Les néologues sont-ils prêts à partager le quasi-monopole organisationnel sur la création et la diffusion? La décentralisation du processus de décision pourrait transformer progressivement les néologues professionnels en des gestionnaires doublés d'animateurs virtuels. Enfin, ne peut-on pas imaginer que cette ouverture serve de tremplin à une prise en charge accrue de l'oralité en néologie?

Ce ne sont là que quelques-unes des questions qu'il faudra résoudre, sans doute au fil des essais. Somme toute, le changement de paradigme annonce une ère de nombreuses découvertes et réflexions.

## Conclusion et perspectives

La mesure de l'implantation intervient dans trois contextes, soit ceux d'un État, d'une organisation ou d'une communauté linguistique. Un bilan des réponses livrées par la terminométrie au fil des ans sur la diffusion et la circulation des unités terminologiques et sur les pronostics d'acceptation d'un néologisme a ensuite été présenté. Devant une méconnaissance des facteurs sociaux, l'apport potentiel d'autres disciplines, existantes (sociolinguistique, diffusion des innovations, par exemple) ou à inventer (ethnoterminologie), a été souligné. Par la suite, le recours aux internautes, ces acteurs virtuels, a été envisagé par le biais des réseaux sociaux et de l'externalisation ouverte. Dans la foulée ont été soulevées des considérations sur ce changement de paradigme, qui modifierait de façon non négligeable les attributs et méthodes du néologue. Bien que l'aménagement étatique ait été privilégié, plusieurs des propositions énoncées s'appliquent également, *mutatis mutandis*, aux organisations et, évidemment, aux communautés. Une politique qui met l'accent sur l'implantation de la langue plutôt que sur celle de termes, combinée à la participation accrue des locuteurs au processus néologique, combattra à la source l'aliénation dont les néologues deviennent parfois victimes.

La terminométrie ne se révèlera vraiment utile pour l'implantation que dans la mesure où elle guidera les actions futures, ce qu'elle n'a réussi jusqu'ici que partiellement. Il est indiqué de poursuivre les travaux terminométriques dans divers domaines, régions et époques pour effectuer des comparaisons. Il faut également cultiver une approche multidimensionnelle et non simplement linguistique. Une main peut être tendue aux disciplines qui éclairent le processus décisionnel du locuteur, en tenant compte de ses interactions sociales. Les réseaux sociaux et l'externalisation ouverte gagneront à être exploités dans un contexte où, depuis une décennie environ, le citoyen se fie davantage à ses pairs qu'aux organismes reconnus. Il y a possibilité de tabler sur un changement social, catalyseur d'une transformation importante de la façon d'envisager la néologie.

Alors, existe-t-il un traitement pour le « mot incompréhensible créé par un malade mental » que représente parfois le néologisme? Peut-être, si l'on ajoute à notre corps de linguistes la contribution d'une équipe multidisciplinaire de chercheurs en sciences humaines et sociales, appuyée par une armée de bénévoles.

## Références

ALEONG S. (1979). *Le vocabulaire du hockey sur glace au Québec : formation, évolution et utilisation. Compte-rendu d'un projet de recherche subventionnée par l'Office de la langue française du Québec.* [Inédit.]

ALEONG S., CHRETIEN M., OSTIGUY L. et MARTIN A. (1981). « De la mise en œuvre d'une terminologie de la langue française de l'automobile dans trois écoles d'enseignement technique au Québec : un premier bilan », *La Banque des mots,* 21, p. 45-66.

ALEONG S. et JOURDAN C. (1981). « De vente à solde ou les dangers de l'hypercorrection », *Meta,* 26 (2), p. 148-158.

AUGER P. (1999). *L'implantation des officialismes halieutiques au Québec : essai de terminométrie*, coll. *Langues et sociétés,* 37, [Montréal], Office de la langue française.

AUGER P. (2001). « Essai d'élaboration d'un modèle terminologique/terminographique variationniste », *TradTerm,* 7, p. 183-224.

BAXTER R. N. (2009). « New Technologies and Terminological Pressure in Lesser-Used Languages: The Breton Wikipedia, from Terminology Consumer to Potential Terminology Provider », *Language Problems and Language Planning,* 33 (1), p. 60-80.

BENHAMIDA L. (1989). *Translators and interpreters as adopters and agents of diffusion of planned lexical innovations: The francophone case.* (Ph. D., University of Illinois at Urbana-Champaign). [Inédit]

CHRETIEN M. (1981). « Reconnaissance, connaissance et utilisation de la terminologie française, québécoise et anglo-américaine de la haute-fidélité chez le public montréalais », In J. M. Klinkenberg, D. Racelle-Latin et G. Connolly (dir.), *Langages et collectivités : le cas du Québec. Actes du Colloque de Liège (mars 1980),* Montréal, Leméac, p. 145-160.

DAOUST D. (1981a). *Diffusion et utilisation de la terminologie technique de langue française dans douze entreprises québécoises. Étude réalisée pour l'Office de la langue française par SORECOM.* [Inédit.]

DAOUST D. (1981b). *Facteurs organisationnels et sociolinguistiques qui sous-tendent la diffusion et l'utilisation des terminologies techniques de langue française dans l'entreprise*. [Inédit.]

DAOUST D. (1987). « L'implantation terminologique : comportements, perceptions et attitudes », *Bulletin de l'ACLA,* 9 (2), p. 17-33.

DELAVIGNE V. et GAUDIN F. (dir.). (1994). *Terminologies nouvelles. Implantation des termes officiels (actes du séminaire de Rouen),* 12, Bruxelles : Rifal.

Délégation générale à la langue française et aux langues de France (2010). *Rapport annuel de la Commission générale de terminologie et de néologie*, 30 mai 2013, http://www.dglf.culture.gouv.fr/cogeter/Rapport_Cogeter_2010.pdf

DEPECKER L. (2013). « Pour une ethnoterminologie ». In J. Quirion, L. Depecker et L. Rousseau (dir.), *Dans tous les sens du terme*. Ottawa : Presses de l'Université d'Ottawa/Office québécois de la langue française, p. 13-29.

DEPECKER L. et MAMAVI G. (dir.) (1997). *La mesure des mots. Cinq études d'implantation terminologique*. Rouen : Publications de l'université de Rouen.

DROUIN P. (2007). « Identification automatique du lexique scientifique transdisciplinaire ». *Revue française de linguistique appliquée,* 12 (2), 45-64.

Druide informatique inc. 2013. *Antidote 8 [Logiciel]*.

FAINBERG Y. A. (1974). « Official Hebrew Terms for Parts of the Car: A Study of Knowledge, Usage and Attitudes », *International Journal of the Sociology of Language,* 1, p. 67-94.

GOMEZ-GUINOVART X. (2012). « A Hybrid Corpus-Based Approach to Bilingual Terminology Extraction ». In I. Moskowich et B. Crespo (dir.), *Encoding the Past, Decoding the Future: Corpora in the 21st Century*. Newcastle upon Tyne, England: Cambridge Scholars, p. 147-175.

HALSKOV J. et BARRIERE C. (2008). « Web-Based Extraction of Semantic Relation Instances for Terminology Work ». *Terminology: International Journal of Theoretical and Applied Issues in Specialized Communication,* 14 (1), p. 20-44.

HELLER M., BARTHOLOMOT J., LEVY L. et OSTIGUY L. (1982). *Le processus de francisation dans une entreprise montréalaise : une analyse sociolinguistique,* coll. Langues et sociétés, Montréal, Office de la langue française.

KAGEURA K. et UMINO B. (1996). « Methods of Automatic Term Recognition: A Review ». *Terminology: International Journal of Theoretical and Applied Issues in Specialized Communication,* 3 (2), p. 259-289.

KARABACAK E. (2009a). « Acceptance of terminology sanctioned by the Turkish Language Society: A study of the use of economic terms in Turkish newspapers », *Terminology: International Journal of Theoretical and Applied Issues in Specialized Communication,* 15 (2), p. 145-178.

KARABACAK E. (2009b). *Unplanned terminology development: A synchronic and diachronic study on economic terms in Turkish newspapers.* (Ph.D., Northern Arizona University). [Inédit]

LEBLANC B. et BILODEAU C. (2009). « Discours épiterminologique : indice du degré de réussite d'implantation des terminologies françaises », *Neologica. Revue internationale de néologie,* 3, p. 167-182.

MARTIN A. (1998). *Les mots et leurs doubles : étude d'implantation de la terminologie officialisée dans le domaine de l'éducation au Québec*, coll. Langues et sociétés*,* 36, [Montréal], Office de la langue française.

MARTIN A. et LOUBIER C. (1993). *L'implantation du français : actualisation d'un changement linguistique planifié*, coll. Langues et sociétés*,* Québec, Office de la langue française.

MAURAIS J. (1984). *La langue de la publicité des chaînes d'alimentation : étude sur la qualité de la langue et sur l'implantation terminologique,* coll. Dossiers du Conseil de la langue française, Études et recherches, 18, Québec, Conseil de la langue française.

MAURAIS J. (1987). « État des recherches sur l'implantation terminologique ». *Unesco ALSED-LSP Newsletter,* 10 (2), p. 25-33.

MediaMiser Inc. (2013). *Page d'accueil.*, 31 mai 2013, http://www.mediamiser.com/.

MONTANE A. (2012). *Terminologia i implantació : anàlisi d'alguns factors que influencien l'ús dels termes normalitzats de la informàtica i les TIC en llengua catalana.* (Ph. D., Universitat Pompeu Fabra). [Inédit]

Office québécois de la langue française (2013). *Grand dictionnaire terminologique*, 17 mai 2013, www.granddictionnaire.com

QUIRION J. (2000). *Aspects évaluatifs de l'implantation terminologique.* (Ph. D., Université de Montréal). [Inédit]

QUIRION J. (2003a). *La mesure de l'implantation terminologique : proposition d'un protocole. Étude terminométrique du domaine des transports au Québec*, coll. Langues et sociétés, 40, Montréal, Office québécois de la langue française.

QUIRION J. (2003b). « Methodology for the Design of a Standard Research Protocol for Measuring Terminology Usage », *Terminology: International Journal of Theoretical and Applied Issues in Specialized Communication,* 9 (1), p. 29-49.

QUIRION J. (2005). « L'automatisation de la terminométrie : premiers résultats », In Association Hellénique de Terminologie. Actes du 5e Congrès international : Langue et terminologie helléniques, Nicosie (Chypre), du 13 au 15 octobre 2005, Athènes, Techniko Epimelitirio Elladas, p. 61-70.

QUIRION J. (2011). « Dynamique terminologique et terminométrie : une complémentarité nécessaire pour le suivi de l'évolution de l'usage des termes et des connaissances », *Terminology : International Journal of Theoretical and Applied Issues in Specialized Communication,* 17 (1), p. 113-133.

QUIRION J. (2012). « Néologie traductive, néologie aménagiste et néologie collaborative massive : l'unité dans la disparité », *Neologica*, 6, p. 129-142.

QUIRION J. et LANTHIER J. (2006). « Intrinsic Qualities Favouring Term Implantation: Verifying the Axioms », In L. Bowker (dir.), *Lexicography, Terminology, and Translation.* Text-based Studies in Honour of Ingrid Meyer, Ottawa, PUO, p. 107-118.

ROGERS E. M. (2003). *Diffusion of innovations* (5e éd.), New York, Free Press.

ROUGES-MARTINEZ J. et VILLEBRUN I. (dir.). (1997). *Terminologies nouvelles*, 16, Bruxelles : Agence de coopération culturelle et technique et Communauté française de Belgique.

TEMMERMAN R. et VAN CAMPENHOUDT M. (dir.) (à paraitre). *The dynamics of culture-bound terminology in monolingual and multilingual communication*. Amsterdam : John Benjamins.

TREMBLAY L. (1994). *Convergence et divergence dans l'emploi de termes communs recommandés par l'Office de la langue française.* (M. A., Université Laval). [Inédit]

VILA I MORENO X., NOGUE PICH M. et VILA I MORENO I. (2007). *Estudis d'implantació terminològica. Una aproximació en l'àmbit dels esports,* coll. En Primer Terme, 3, Vic/Barcelona : EUMO Editorial/TERMCAT Centre de Terminologia.

VINTAR Š. (2010). « Bilingual Term Recognition Revisited: The Bag-of-Equivalents Term Alignment Approach and Its Evaluation ». *Terminology: International Journal of Theoretical and Applied Issues in Specialized Communication,* 16 (2), p. 141-158.

# Analyser la néologie dans le domaine de l'ingénierie nucléaire

Danielle Candel[1] et Marie Calberg-Challot[2]

## Résumé

Cet article présente une réflexion sur l'évolution du vocabulaire dans le domaine de l'ingénierie nucléaire, au cœur du secteur plus général du nucléaire. Le nucléaire est une science jeune, d'un peu plus d'un siècle seulement – son vocabulaire est jeune lui aussi. C'est, de ce fait, un secteur et une activité dont les experts, qui nous sont contemporains, peuvent offrir des témoignages sur la création ou l'histoire d'un terme, élément très précieux dans le type d'étude que nous avons expérimenté, où évolution diachronique et variation synchronique se mêlent et se complètent. Il montre que la néologie relève de plusieurs besoins tels la nécessité de dénommer de nouvelles réalités, la lutte contre les anglicismes et un besoin d'adaptation des termes face aux destinataires du message. Nous présentons un double corpus d'étude. Il couvre d'une part une période de près de quinze ans, au cœur des travaux de terminologie officielle. Il est d'autre part relatif à un ensemble d'enquêtes et d'entretiens réalisés en vue d'un dictionnaire de l'ingénierie nucléaire. Nous résumerons finalement quelques caractéristiques et tendances de la création terminologique ainsi observées. Les connaissances de ce domaine, qui reste en pleine évolution technologique, représentent un excellent terrain d'analyse et en font un secteur particulièrement propice à l'étude de la néologie : nous montrons que la mise en avant des termes nouveaux ne peut se faire sans la prise en compte des réactions du public, sans son approbation.

## Mots clés

néologie en pratique, ingénierie nucléaire, terminologie officielle, terminologie en immersion industrielle

## Introduction

Nous proposons dans cet article une réflexion sur les néologismes dans le domaine bien délimité de l'ingénierie nucléaire, au cœur du secteur plus général du nucléaire, dans la lignée et en complément de travaux tels que ceux de

[1] CNRS HTL (Histoire des Théories Linguistiques), Université Paris Diderot, danielle.candel@univ-paris-diderot.fr

[2] CNRS HTL (Histoire des Théories Linguistiques), Université Paris Diderot, challot@orange.fr

Delavigne (2001), Candel et Calberg-Challot (2007) ou Candel et Tombeux (2008)[3].

Deux particularités sont à souligner. D'un côté, le nucléaire est une science d'un peu plus d'un siècle seulement, et l'industrialisation et la mise en application des recherches dans ce domaine datent de quelques décennies : le nucléaire est une science jeune, et son vocabulaire l'est aussi. D'un autre côté, le vocabulaire du nucléaire se distingue de celui des autres domaines par le fait qu'il a moins emprunté à l'anglais : il est souvent d'origine française. Ces deux raisons font sans doute qu'on s'est assez peu penché sur ce vocabulaire jusque récemment. A cela s'ajoute que les termes sont le plus souvent des néologismes de sens, dont l'aspect retient moins l'attention que des néologismes de forme ; et lorsque ce sont des néologismes de forme, il s'agit de syntagmes techniques construits à partir d'unités simples dont la forme nous est familière par ailleurs.

Mais le secteur du nucléaire dans son ensemble fait parler de lui de plus en plus. Cela est dû, surtout, à trois accidents : les accidents de Three Mile Island en 1979, de Tchernobyl en 1986 et de Fukushima en 2011. Ils ont changé la donne. L'ingénierie nucléaire et le domaine qu'elle représente, cette science et cette technique modernes issues des recherches fondamentales menées en chimie, en sciences physiques puis plus particulièrement en physique nucléaire, peuvent être étudiées dans des limites historiques précises. Surtout, la jeunesse de ce domaine met en lumière que les connaissances ne sont pas figées et que ce domaine, qui reste en pleine évolution technologique, est propice à l'étude de la néologie. C'est aussi une activité dont les experts, qui nous sont contemporains, peuvent offrir des témoignages, comme ceux portant sur la création ou l'histoire d'un terme (Calberg-Challot, 2012), où évolution diachronique et variation synchronique se mêlent et se complètent. C'est un excellent terrain d'analyse, qui montre que la néologie relève de plusieurs besoins, tels la nécessité de dénommer de nouvelles réalités, une solution de remplacement aux anglicismes, et un besoin d'adaptation des termes face aux destinataires du message. Nous vérifions que « la néologie n'est pas seulement l'affaire des normalisateurs, mais qu'elle permet une compréhension des processus de la découverte scientifique » (Humbley, 2003 : 263). C'est aussi ce que nous expérimentons quotidiennement en France dans le travail auprès des commissions de terminologie et de néologie (Candel, 2010).

Nous présenterons d'abord notre double corpus d'étude : le premier terrain d'étude couvre une période de près de quinze ans, au travers des travaux des Commissions de terminologie ; le second est relatif à un ensemble d'enquêtes et d'entretiens réalisés en vue d'un dictionnaire de l'ingénierie nucléaire au sein de la société Areva NP[4] (1). L'étude de ce corpus fournit des indications sur les modes de repérage des termes, ainsi que sur leur implantation, leur circulation ou leur remplacement, notamment à travers des descriptions successives qui

[3] Citons aussi Djambian, 2010.

[4] Areva Nuclear Power.

leur ont été consacrées (2). Nous résumerons finalement quelques caractéristiques et tendances de la création terminologique que nous avons ainsi observées (3).

## 1. Deux terrains d'analyse de la néologie en pratique : terminologie officielle et industrie

### 1.1. Néologie terminologique institutionnelle

1.1.1. Une collaboration multidisciplinaire

Le cadre du processus officiel d'enrichissement de la langue française permet une collaboration entre experts d'un vaste ensemble de domaines et généralistes. L'expertise des participants est de nature différente, et ils s'appuient volontiers sur des corpus ouverts, des documents variés. Nous y participons nous-même activement, puisque la Délégation générale à la langue française et aux langues de France et le laboratoire Histoire des théories linguistiques du CNRS sont liés par une convention annuelle, par laquelle HTL s'engage à apporter aux commissions de terminologie et de néologie, outre un ensemble de données documentaires, une expertise linguistique[5].

Dix-huit commissions spécialisées de terminologie et de néologie œuvrent pour la néologie et la terminologie du français, dans les domaines respectifs suivants : Affaires étrangères et européennes, Agriculture, Automobile, Chimie, Culture et communication, Défense, Économie et finances, Éducation, Environnement, Équipement et transports, Informatique et composants électroniques, Ingénierie nucléaire, Jeunesse et sports, Pétrole, Santé, Sciences et techniques Spatiales, Sports, Télécommunications et affaires postales.

Le domaine du nucléaire touche les trois ensembles des sous-secteurs suivants : Physique, Technologie et exploitation des réacteurs, Fusion thermonucléaire ; Sécurité nucléaire, Radioprotection et qualité ; Cycle du combustible et Déconstruction.

La Commission générale de terminologie et de néologie réunit quant à elle des représentants des Académies (Académie des sciences, Académie française et service du dictionnaire de l'Académie), du ministère de la Culture, de la diplomatie, des secteurs juridique, littéraire, ou de l'enseignement, des linguistes, des lexicographes, des terminologues, sans oublier l'Afnor et le CNRS[6]. Les consultations de correspondants francophones (l'Office québécois de la langue française, le Bureau de la traduction à Ottawa) viennent enrichir cet ensemble[7].

---

[5] La responsabilité scientifique de ce contrat de recherche incombe à D. Candel (depuis 1999).

[6] Notre laboratoire (v. note 2).

[7] Les résultats des travaux doivent être validés par l'Académie française.

1.1.2. Une production terminologique et néologique marquée par des consignes officielles

Il importe, comme nous l'avons rappelé ailleurs, d'indiquer les consignes données aux commissions de terminologie et de néologie dans le *Décret du 3 juillet 1996 relatif à l'enrichissement de la langue française (appelé le « décret Juppé »)*, indispensables pour comprendre le système engendré par le processus d'enrichissement de la langue française tel que le gère la Délégation à la langue française et aux langues de France. Ces consignes sont : « établir l'inventaire des cas dans lesquels il est souhaitable de compléter le vocabulaire français », « recueillir, analyser et proposer les termes et expressions nécessaires, notamment ceux équivalant à des termes et expressions nouveaux apparaissant dans les langues étrangères, accompagnés de leur définition » et telles qu'elles apparaissent, plus précisément, à travers les critères de choix à observer pour la proposition de nouveaux termes : « La nécessité : la création d'un nouveau terme est-elle indispensable pour désigner la notion ? » ; « La transparence : le terme est-il immédiatement associé à la réalité ou à la notion qu'il désigne ? » ; « La bonne formation : le terme respecte-t-il le système morphologique et syntaxique du français ? ».

Ces consignes soulignent à elles seules le rôle de la néologie. On comprend à travers chacune d'elles que toutes les fiches traitées sont censées relever de la néologie. La seule déclaration officielle de ces fiches leur confère ce statut. Ces nouvelles dénominations servent (1) à « compléter le vocabulaire français » et (2) à « proposer des termes et expressions nécessaires ». Surtout, on perçoit aisément le rôle du sentiment néologique (Sablayrolles, 2003). Mais une chose est la néologie en commission officielle, une autre est sans doute la néologie sur le terrain du nucléaire lui-même.

## 1.2. Néologie terminologique en immersion industrielle

1.2.1. Une collaboration entre l'industrie et l'université

Non sans lien avec le premier secteur décrit, l'accueil chez Areva d'une étudiante en traitement automatique des langues et plus généralement en sciences du langage[8], en stage de DESS puis en bourse CIFRE, a permis une collaboration entre des spécialistes de l'industrie nucléaire et une spécialiste linguiste. Cette collaboration est d'autant plus précieuse que nous avons très peu de témoignages de linguistes sur le terrain industriel. Ce manque est dû aussi au fait que le domaine est « sensible » et difficilement accessible.

[8] Marie Calberg-Challot, accueillie chez Areva de 2003 à 2007.

1.2.2. Une production terminologique testée auprès des chercheurs industriels

Cette collaboration s'est concrétisée par des enquêtes, celles de l'analyste extérieur à l'entreprise, qui s'implante directement dans l'entreprise et interroge les experts à même leur terrain professionnel de recherche.

Le résultat prévu de ces travaux étant une participation à l'élaboration d'un nouveau dictionnaire, les échanges, aussi bien sur des termes que sur leurs définitions ou sur leurs conditions d'emploi, ne pouvaient se dérouler qu'en toute confiance[9]. D'une certaine façon, la collaboration s'est effectuée comme à égalité d'expertise. C'est-à-dire que ces travaux se sont souvent faits avec une mise à plat des expériences pour un échange fructueux, sur les usages des termes et sur les impressions sur ces termes et usages.

C'est ce qui s'est fait notamment pour le « corpus du sous-domaine du Combustible » (Calberg-Challot *et al.*, 2008). C'est un corpus de référence, écrit par les ingénieurs et pour les ingénieurs de ce même sous-domaine. Il a été choisi par les experts comme étant représentatif de leur activité et de leur pratique. Ce corpus étant par ailleurs récent, il est important pour l'observation de la néologie contemporaine. Composé de vingt textes, il comporte quasiment 200 000 occurrences-mots. Cet ensemble a d'abord été traité en vue d'une compatibilité avec le logiciel Syntex (Bourigault *et al.*, 2005). De cet ensemble, plus de 18 000 syntagmes nominaux ont été extraits, et plus de 5 000 noms. Seuls 922 termes ont été retenus à l'issue d'une première sélection réalisée par un expert. D'autres travaux ont été menés à la suite de cette étude et de nouvelles sélections plus fines ont été réalisées en site industriel (Calberg-Challot *et al*, 2007). Cette démarche de validation par divers experts est importante pour les ingénieurs et spécialistes du sous-domaine du Combustible comme pour les chercheurs en linguistique.

Dans bien des cas, l'action même d'isoler et de caractériser des termes comme relevant du domaine de l'industrie nucléaire en fait des néologismes terminologiques, qu'il s'agisse des termes étudiés en commissions de terminologie et de néologie ou des termes sur lesquels on s'arrête dans les études de terrain, avec les spécialistes du domaine.

---

[9] Il convient ici de remercier Xavier Dumont sans qui l'expérience « Areva » à laquelle on se réfère dans ces pages n'aurait pu avoir lieu.

## 2. Une terminologie dans sa dynamique : sélection, présentation et analyse des données

### 2.1. Les données du terrain officiel

2.1.1. Le nucléaire en progression

Afin d'évaluer le domaine du nucléaire par rapport à l'ensemble du vocabulaire publié au *Journal Officiel*, la répartition des termes par domaines est intéressante. La situation au 30 juin 2012 est donnée à la *figure 1*.

*Fig. 1. Le nucléaire par rapport aux autres domaines, d'après la base terminologique et néologique FranceTerme : 5,6% des fiches en 2012 (contre 4,5% en 2004).*

Ce relevé a été réalisé à partir de l'ensemble des 5399 fiches de la base FranceTerme disponibles le 30 juin 2012[10]. Pour mémoire, des relevés réalisés de manière semblable[11] le 12 octobre 2004 permettaient de répertorier un ensemble de 171 termes du domaine du nucléaire, pour un total de 3788 fiches disponibles dans la base, soit une proportion de 4,5% de fiches de ce domaine contre 5,6% aujourd'hui. La prise en compte du domaine de l'ingénierie nucléaire est donc en augmentation dans le processus d'enrichissement de la langue française.

[10] Précisons que les fiches du domaine « Biologie » y sont dues à des listes émanant de l'Académie des sciences, qui complètent celles des domaines plus généraux de la commission de la « Santé ». Les fiches du domaine de l'Audiovisuel sont généralement issues des fiches de la commission « Culture et communication », celles des domaines de l'économie, de la gestion et des finances, de la Commission de l' « Economie et des finances ». Ce total de 5399 fiches comporte aussi les fiches des Affaires étrangères, de l'Agriculture, de l'Automobile, de l'Education, de l'Environnement et du Sport.

[11] Egalement par D. Candel.

Pour une meilleure présentation de ce corpus, nous en avons classé les termes en fonction de la partie du discours qui leur est propre, et du schéma morphologique qui est le leur.

Les termes du nucléaire revêtent le plus souvent la forme de noms (N) ou de groupes nominaux (surtout des formes NA ou NdeN), avec quelques adjectifs. La *figure 2* représente environ 80% des termes de ce domaine, répartis en fonction de leur modèle morphologique.

*Fig. 2. Les termes officiels du domaine du nucléaire :*
*Nom + Adjectif, Nom, NomdeNom, NomdeNom + Adjectif, Adjectif (30 juin 2012).*

Il convient d'ajouter à cet ensemble environ soixante autres formes nominales, qui revêtent des formes variées puisqu'on en compte environ trente modèles différents.

2.1.2. La fluctuation des données officialisées

Les résultats obtenus par les commissions de terminologie et de néologie, qui donnent pourtant lieu à des fiches publiées qui elles-mêmes ont valeur de recommandations officielles, sont fluctuants. Sans rien révéler des travaux tels qu'ils se déroulent et tels que nous pouvons les observer au sein même des commissions de terminologie et de néologie[12], nous pouvons néanmoins

[12] Nous sommes en effet tenues à une obligation de réserve, les travaux internes aux commissions n'ayant pas vocation à être diffusés (Danielle Candel participe au circuit des commissions de terminologie et de néologie depuis juin 1997, d'abord au titre d'expert « es-qualité » de commissions, puis au titre d'expert linguiste dans le cadre d'une convention annuelle avec la Délégation générale à la langue française et aux langues de France; Marie Calberg a participé à des groupes de travail de la commission de terminologie et de néologie de l'Ingénierie nucléaire pendant une partie du séjour qu'elle a effectué chez Areva).

attester de changements intervenus chez les auteurs des fiches, c'est-à-dire parmi les membres des commissions. Pour ce faire, nous analysons des fiches en fonction d'indices de variations dans le temps.

Deux principaux types d'indications sur l'instabilité des néologismes sont perceptibles à travers les fiches publiées[13].

*2.1.2.1. Ces indices d'instabilité des termes ou des concepts nous viennent des reformulations des termes ou de leurs définitions.*

**(a) Une variation historique explicite.** Elle accompagne la définition. Ainsi, dans le champ proposant « Attention : Cette publication annule et remplace... » comme ci-dessous :

> ***Journal officiel* du 22/02/2009**
> **activité,** n.f.
> Domaine : Ingénierie nucléaire/Radioprotection
> Définition : Nombre de transitions nucléaires spontanées qui se produisent dans une quantité donnée de radionucléides par unité de temps. (…)
> Équivalent étranger : activity (en)
> **Attention :** Cette publication annule et remplace celle du terme **« activité nucléaire »** au *Journal officiel* du 22 septembre 2000.
>
> ***Journal officiel* du 20/02/2011**
> **combustion massique**
> Domaine : Ingénierie nucléaire/Cycle du combustible-Physique des réacteurs
> Définition : Énergie libérée par un combustible nucléaire au sein d'un réacteur en fonctionnement, rapportée à sa masse initiale en métal lourd tel l'uranium, le plutonium ou le thorium. (…)
> Équivalent étranger : burn-up (en), specific burn-up (en)
> **Attention :** Cette publication annule et remplace celle du terme « **épuisement spécifique** » au *Journal officiel* du 22 septembre 2000.

où nous avons une indication sur la préférence à donner à deux termes en concurrence. Dans la première des deux fiches, on note une simplification du terme.

**(b) Le recours à la reformulation.** Celle qui est présentée dans les extraits qui suivent diffère de celle rencontrée par exemple dans les abréviations que sont les siglaisons. Nous analysons ici les champs « Abréviation » et « Forme abrégée », dans lesquels l'ellipse porte sur l'élément « deN » :

> « cascade d'enrichissement » reformulé en « cascade »
> « commande d'un réacteur » reformulé en « commande »
> « gaine de combustible » reformulé en « gaine »
> « ralentissement des neutrons » reformulé en « ralentissement »,

[13] Il s'agit toujours du corpus arrêté le 30 juin 2012.

l'élément « deN » sur lequel porte l'ellipse pouvant aussi être enchâssé dans le terme initial, comme dans :

« circuit de refroidissement primaire » reformulé en « circuit primaire »
« réaction de fission en chaîne » reformulé en « réaction en chaîne ».

L'ellipse peut aussi concerner simplement un adjectif :

« contamination radioactive » reformulé en « contamination »
« fusion thermonucléaire » reformulé en « fusion »
« retombées radioactives » reformulé en « retombées »
« colis de déchets radioactifs » reformulé en « colis de déchets »

ou bien un ensemble « deNA », comme dans :

« colis de matières radioactives » reformulé en « colis »

ou encore, plus curieusement :

« combustible porteur d'actinides mineurs » reformulé en « combustible porteur ».

Comme on le remarque dans plusieurs de ces derniers exemples, l'ellipse a souvent trait à l'adjectif « radioactive ». Ce peut être là le signe d'une gêne, du repérage d'un élément « tabou », d'une nécessaire prudence, la radioactivité étant quelque chose qui effraie.

Un dernier exemple est très différent des précédents : l'ellipse y porte juste sur la préposition « de », et l'on passe ainsi facilement d'un « NdeN » à un « NN » :

« zonage de déchets » est reformulé en « zonage déchets ».

Ce que nous voudrions souligner à ce propos – et peut-être cette remarque nous est-elle inspirée par une expérience de plus de douze ans auprès des commissions de terminologie –, c'est qu'il est tout à fait plausible que dans la série des exemples cités, la forme abrégée, celle qui est donc proposée en second ressort, ait été la proposition initiale des experts du domaine. C'est donc celle qui correspondrait à leur usage effectif, voire à leur choix. Dans les faits, l'expression « zonage de déchets » a-t-elle donné lieu à un nouveau néologisme, « zonage déchets » ? N'est-ce pas plutôt l'inverse, qui s'est produit ? Si c'est le cas, « zonage de déchets » serait un néologisme que l'on pourrait qualifier d' « artificiel », et qui aurait supplanté le néologisme « zonage déchets », que l'on peut appeler « naturel », ou « spontané ».

**(c) Un champ « Synonyme ».** La présence d'un champ « synonyme » est classique dans les fiches de terminologie et de néologie. Différentes catégories de formations synonymiques ont été rencontrées pour le présent vocabulaire de l'ingénierie nucléaire.

Soit le premier élément est maintenu et la suite du syntagme est modifiée : « cellule chaude » a pour synonyme « cellule de haute activité » et « rapport de flux thermique » a pour synonyme « rapport d'ébullition critique ». Soit le premier élément et au contraire modifié : « conduite d'un réacteur » a pour

synonyme « pilotage d'un réacteur » ; « couche de demi-atténuation » a pour synonyme « épaisseur de demi-atténuation » ; « facteur de conversion » a pour synonyme « rapport de conversion, taux de conversion » ; « teneur isotopique » a pour synonyme « richesse isotopique », et « teneur isotopique pondérale » a pour synonyme « richesse isotopique pondérale ».

Pour compléter l'exposition du champ de la synonymie dans les fiches consacrées à des termes du domaine de l'ingénierie nucléaire, il faut citer le cas où la marque « synonyme » peut être doublée d'une autre marque, à mi-chemin entre une marque diastratique et une marque de technicité. C'est l'expression « langage professionnel » (Candel, 2010) : « déchet ordinaire » a pour synonyme « déchet classique *langage professionnel* » ; « dose prévisible » a pour synonyme « dose prévisionnelle *langage professionnel* ».

Soulignons que deux cas retiennent tout particulièrement notre attention. C'est d'abord celui de « radioactivation », dont il est précisé qu'il a pour synonyme « activation ». Cet exemple nous conforte dans la conclusion précédemment énoncée, concernant le « tabou » lié à la notion de radioactivité, et aux termes qui l'expriment. C'est ensuite celui, dû à un phénomène plus rare, qui est le changement complet du terme. C'est un cas particulièrement marquant et on aura l'occasion de revenir sur ce sujet à propos de « récupérateur de corium » qui a pour synonyme « cendrier ».

Pour terminer, il nous faut remarquer qu'il est possible, ici comme pour les « reformulations » formelles en « abréviations » ou « formes abrégées » évoquées plus haut, que les spécialistes soutiennent spontanément les termes qui ne sont proposés ici qu'en synonymes, et non ceux proposés en entrées.

La synonymie peut donc elle-même faire appel à cette autre forme d'expression qu'est la marque « Langage professionnel», par exemple lorsqu'elle cherche à gommer la notion de « radioactivité».

**(d) Un champ « On trouve aussi ».** Cette catégorie est relativement peu représentée dans le corpus étudié (v. la *figure 3* ci-dessous), et nous ne nous y attarderons pas davantage.

*2.1.2.2. D'autres indications sont repérables à travers des données que nous qualifierons d'éparses.*

Il s'agit des marques « Langage professionnel » et « Note ».

**(a) Marque « Langage professionnel ».** Nous avons cité les termes de « déchet classique » et « dose prévisionnelle », proposés comme alternatives des formes officiellement données en entrée. Nous avons aussi rappelé plus haut le terme de « zonage de déchets » et de son synonyme « zonage déchet », ce dernier terme étant immédiatement précisé, lui aussi, par la marque « langage

professionnel ». S'ajoutent à ces exemples les cas de « îlotage », « inétanche » et « mise à l'arrêt définitif (MAD) ».

Plus explicitement, sous « neutrons instantanés », on lit : « On trouve aussi, dans le langage professionnel, l'expression 'neutrons prompts', qui n'est pas recommandée » ; et, sous « criticité instantanée » : « On trouve aussi, dans le langage professionnel, l'expression 'criticité prompte', qui n'est pas recommandée ». A la précision de « langage professionnel », qui ne semble pas avoir de connotation négative en soi dans le contexte du processus officiel de néologie et de terminologie, s'ajoute donc une formule de restriction très claire. C'est une mise en garde contre l'utilisation effective du terme « professionnel » de « criticité prompte ». Finalement, la présence de l'indication « langage professionnel » qui qualifie l'entrée sert donc de marque prescriptive, dissuasive, destinée à lutter contre la diffusion de l'emploi professionnel. Mais ce dernier est sans doute un emploi déjà effectif : n'est-il pas précisément là, le véritable néologisme terminologique à signaler ?

**(b) Il s'agit d'autre part des informations données dans le Champ « Note ».** Les auteurs de fiches terminologiques et néologiques expriment à travers les notes différents types de précisions qui témoignent d'une intéressante variété.

Ainsi, lorsque pour la fiche « réacteur à neutrons thermiques » publiée en 2000, on lit en note « Le terme 'réacteur thermique' est déconseillé », c'est la concurrence entre deux termes qui est décriée, une forme longue, admise, et une forme brève, refusée.

Pour la fiche consacrée en 2004 au terme « dose », la définition « Quantité d'énergie communiquée à un milieu par un rayonnement ionisant » est suivie d'une longue note : « Ce terme général prend un sens particulier lorsqu'il est employé en association avec un ou plusieurs adjectifs qualificatifs tels que « absorbée », « collective », « efficace », « équivalente », « individuelle ».

Cette information est déroutante car on assiste ici à une « déterminologisation » des termes. Les ensembles « dose absorbée », « dose collective », « dose efficace », « dose équivalente » et « dose individuelle », qui sont effectivement des « termes », ne devraient pas être re-décomposés en un substantif, « dose », suivi d'un adjectif, ou plutôt d'un des quatre adjectifs attestés dans cet article. D'ailleurs, la base FranceTerme atteste bien en 2004 les fiches « dose absorbée », « dose collective », « dose efficace » et « dose équivalente ». La note a juste donné l'occasion à ses auteurs de signaler un inconfort terminologique et néologique : il y a là un certain malaise entre un terme simple, à valeur par ailleurs très générale, et un terme composé de ce même nom et d'un adjectif, qui est davantage ressenti comme un néologisme fautif, comme si cette formation si classique en « NA » devait être évitée. Il y a donc incontestablement une instabilité d'usage, de sentiment linguistique, et les

néologismes signalés comme tels sont en même temps, pour certains, mal acceptés.

L'instabilité des néologismes dans les fiches de terminologie est montrée à la *figure 3*, qui se compose de deux parties.

*Fig. 3. Repérage des indications sur l'instabilité des néologismes dans les fiches de terminologie et de néologie (la partie supérieure de la fiche traite des termes autres que ceux du nucléaire ; la partie inférieure concerne les termes du nucléaire).*

La partie supérieure de cette figure résume une étude des fiches de FranceTerme, à l'exception de celles du domaine du nucléaire, soit 5092 fiches. La partie inférieure traite exclusivement des 307 fiches du nucléaire.

Force est de constater que ces indications sont bien plus nombreuses pour les termes du nucléaire que pour les termes de autres domaines : on y note deux fois plus d' « annule et remplace », une fois et demie plus d'abréviations et deux fois plus souvent la marque « langage professionnel », cette dernière s'accompagnant classiquement d'une forme considérée comme préférable à celle ainsi marquée. Quant aux notes, elles sont elles aussi deux fois plus nombreuses pour le nucléaire, ce qui montre sans doute que le domaine du nucléaire appelle davantage de commentaires que les autres domaines.

### 2.2. Les données du terrain de la recherche industrielle

Nous nous pencherons sur deux termes particulièrement éclairants : « traitement » et « retraitement ».

2.2.1. Les termes « traitement » et « retraitement »

Ces termes ont donné lieu à de nombreux commentaires et réflexions. On relève de nombreuses nouvelles versions du terme, de nouveaux usages, comme le montre la *figure 4*. Mais ces termes et ces usages sont parfois abandonnés, puis repris plus tard, acquérant alors une nouvelle valeur néologique.

*Fig. 4. Usages successifs de « traitement », « retraitement », « gestion ».*

Il se trouve que certains d'entre eux ont déjà retenu l'attention du processus officiel de terminologie et de néologie. Ainsi, on constate que dans les recommandations de terminologie et de néologie officielles, « retraitement », consacré par l'usage, supplante « traitement » :

> **retraitement,** n.m. (*J.O.* du 22/09/2000)
> *Domaine* : Ingénierie nucléaire/Cycle du combustible
> *Définition* : Traitement des combustibles usés pour en extraire les matières fissiles et fertiles de façon à permettre leur réutilisation, et pour conditionner les différents déchets sous une forme apte au stockage. (…)
> *Équivalent étranger* : reprocessing (en)

Et en octobre 2012, le terme « retraitement » est le seul, dans le groupe relevé, à figurer dans FranceTerme, outre « uranium de retraitement », qui d'ailleurs y renvoie :

> **uranium de retraitement** (*J.O.* du 20/02/2011)
> *Abréviation* : URT
> *Domaine* : Ingénierie nucléaire/Cycle du combustible
> *Définition* : Uranium issu du traitement de combustibles usés.
> *Voir aussi* : combustible d'oxyde d'uranium, retraitement
> *Équivalent étranger* : reprocessed uranium (en), RepU (en)

Mais c'est en s'entretenant avec les professionnels de l'industrie du nucléaire, à même leur terrain professionnel, que l'on arrive à mieux comprendre certains éléments de l'évolution du vocabulaire et de ses fluctuations. On constate par exemple qu'à l'origine de la réflexion, des hésitations étaient déjà bien présentes entre « traitement » et « retraitement ». Aujourd'hui et après avoir longuement utilisé le terme « retraitement », on souligne l'usage du terme « traitement » et on

explique les raisons sous-jacentes à cet usage. Ainsi, il nous a été rapporté[14] que le terme « traitement » a été introduit par Areva Nc[15] afin de disposer d'une dénomination jugée meilleure puisque le combustible usé n'a pas encore été traité et que c'est son premier « traitement » que l'on cherche à dénommer, et non un « retraitement ». Par ailleurs, comme nous l'avons compris en nous entretenant avec des spécialistes du secteur industriel, il est vrai que l'usage de « traitement » intrigue et dérange davantage que le terme « retraitement » : « traitement » exprime un processus plus direct que « retraitement » et met davantage l'accent sur une nécessaire transformation du combustible usé, c'est-à-dire irradié, alors que l'usage de « retraitement » rassure davantage, dans la mesure où il s'agit d'une deuxième étape de processus, donc de quelque chose de quasi-routinier, ce qui introduit une certaine distance par rapport au combustible usé, au combustible irradié lui-même.

Enfin, lors d'une campagne de publicité plus récente du groupe Areva, on ne parle même plus de « traitement » ni de « retraitement », nous a-t-on expliqué, mais de « gestion des déchets ». C'est que le terme « gestion » a une connotation plus positive que le terme « déchet ». Ainsi, « gestion des déchets » ou « gestion des combustibles nucléaires usés » supplanterait « traitement » et « retraitement ». On trouve également, dans des notes rédigées par les industriels du nucléaire, la proposition du néologisme « traitements des déchets radioactifs », candidat à définition en 1993 ; mais en 1999, il est proposé, dans les instances de définition officielles, de rejeter ce terme, jugé « trop général » ; on a avancé le terme « traitements des effluents radioactifs »[16], qui sera, à son tour, rejeté avant de laisser la place à « gestion de déchets radioactifs ». Les liens étant souvent étroits entre les industriels que nous avons pu consulter sur leur terrain d'étude et la commission officielle de terminologie, on retrouve ce terme dans le *Journal officiel* du 3 août 2000 avec la définition suivante : « Ensemble des dispositions et opérations règlementaires et techniques relatives aux déchets radioactifs depuis leur production jusqu'à leur stockage inclusivement, destinées à assurer la protection des personnes et de l'environnement ». Le terme de « gestion des déchets » est certainement plus neutre que « traitement des déchets ».

2.2.2. Les termes « cendrier » et « récupérateur de corium »

Avec l'exemple de « cendrier » et « récupérateur de corium », nous pouvons suivre à la trace la naissance d'un néologisme.

En 1980 et 1988, l'Office de la langue française (Québec)[17] fournit respectivement, comme équivalent à l'anglais « core catcher », « récupérateur du

---

[14] Explications de Xavier Dumont, Areva, 30 juin 2006.
[15] Areva Nuclear Cycle.
[16] Mais « effluents » renvoie à des gaz et à des liquides, et « déchets » renvoie plutôt à des solides.
[17] Consulté le 18 septembre 2004.

cœur » puis « récupérateur », avec, les deux fois, le « quasi-synonyme » « cendrier ». Des emplois divers sont relevés dans plusieurs textes portant sur le nucléaire, mais le terme « cendrier » est absent. Initialement utilisé au sujet des réacteurs à neutrons rapides, « cendrier » est né après les accidents de Three Mile Island et de Tchernobyl. Lors de l'accident de Three Mile Island[18], le réacteur n°2 de type REP, d'une puissance de 900 mégawatts électriques, est le siège d'une fusion partielle du cœur et d'une contamination importante à l'intérieur de l'enceinte de confinement. C'est le fond de l'enceinte de confinement qui, jouant le rôle de réceptacle et retenant le cœur en fusion, donne naissance au nouveau concept et à la dénomination de « récupérateur de corium ». C'est après cet accident que les termes « cendrier » et « récupérateur de corium » sont entrés en concurrence.

La Société française d'énergie nucléaire (SFEN) a publié en 2004 la définition suivante pour le terme « corium » : « Amas de combustibles et d'éléments de structure du cœur d'un réacteur nucléaire, fondus et mélangés, pouvant se former en cas d'accident grave ». Le terme de « cendrier » a fait l'objet de quelques échanges parmi les experts, dont voici la définition publiée par la SFEN (en 2004) : « Dispositif se trouvant sous le cœur d'un réacteur nucléaire, destiné en cas d'accident à récupérer le corium et à en faciliter le refroidissement », l'équivalent anglais étant « core catcher ». Mais il se pose, parmi les experts, la question d'introduire le terme « récupérateur de corium » qui semble mieux adapté.

Le « Programme international majeur de recherche en sûreté nucléaire : Phébus PF (Produits de Fission) », mené à Cadarache par l'Institut de Radioprotection et de Sûreté Nucléaire (IRSN) en juin 2003, est un important programme de recherche en sûreté des centrales électronucléaires dans le monde et qui a fait l'objet d'un dossier de presse. Ce programme fait état d'une vaste collaboration internationale réunissant la plupart des pays utilisant l'énergie nucléaire pour produire l'électricité. Or il n'y est fait aucune mention de « cendrier » ni de « récupérateur de corium », qui sont remplacés par des périphrases et des reformulations désignant soit le contenu, soit le contenant, avec « bain de corium », « bain de quelques kilogrammes de combustible », « lit de débris » ou « un tiers s'écoule vers le fond de cuve ».

Dans la brochure Areva de mars 2003 intitulée *EPR, un choix stratégique*, on trouve le terme « zone protégée dans un revêtement réfractaire » : « A l'intérieur de l'enceinte est aménagé une zone protégée par un revêtement réfractaire. C'est là qu'en cas de fusion du cœur, la partie du cœur fondu qui pourrait s'échapper de la cuve serait recueillie et refroidie ». Dans une deuxième édition de la même brochure près d'un an plus tard, en février 2004, on trouve cette fois le terme « compartiment spécifique » : « toute portion du cœur fondu

[18] Source : DGEMP (Direction générale de l'énergie et des matières premières), « L'énergie nucléaire en 110 questions ».

s'échappant de la cuve serait recueillie et contenue (dispositions passives), puis refroidie, dans un compartiment spécifique situé à l'intérieur de l'enceinte de confinement ». Enfin, dans les cédéroms de présentation de l'EPR, deux termes désignent la même notion : « réceptacle réfractaire de corium », en légende d'un schéma, et, dans l'extrait suivant, « compartiment spécifique » : « La protection accrue contre les accidents graves est l'un des points forts de l'EPR. Ainsi, en cas de fusion du cœur, de probabilité extrêmement faible, le bâtiment réacteur est équipé d'un compartiment spécifique destiné à recueillir et à refroidir la partie du cœur fondu qui pourrait s'échapper de la cuve ». Dans la version anglaise de cette même présentation, le terme utilisé est « corium retention area ».

Le terme « cendrier », initialement utilisé pour une autre technologie que celle des réacteurs à eau sous pression, ne se maintiendra pas dans les usages. Il a été remplacé par « récupérateur de corium ». On constate d'ailleurs que si dans l'édition 2003 du *Vocabulaire de l'ingénierie nucléaire* publié par la SFEN, l'article « récupérateur de corium » renvoie vers l'article « cendrier » :

> récupérateur de corium
> ***Domaine*** : Ingénierie nucléaire/Technologie des réacteurs.
> ***Voir*** : cendrier.

la tendance s'inverse complètement dans l'édition 2007 de ce même *Vocabulaire* :

> récupérateur de corium
> ***Domaine*** : Ingénierie nucléaire/Technologie des réacteurs.
> *Synonyme* : cendrier, n. m.
> *Définition* : Dispositif se trouvant sous le cœur d'un réacteur nucléaire et destiné, en cas d'accident, à récupérer le corium et à en faciliter le refroidissement.
>
> cendrier
> ***Domaine*** : Ingénierie nucléaire/Technologie des réacteurs.
> ***Voir*** : récupérateur de corium.

Devant ce besoin de nommer un nouveau concept, nous sommes donc face à un grand nombre de variantes terminologiques et néologiques : « cendrier », « récupérateur », « récupérateur du cœur », « récupérateur de corium », mais aussi « réservoir interne à l'enceinte », « zone protégée par un revêtement réfractaire », « compartiment spécifique » et « réceptacle réfractaire de corium ». Mais le dernier mot sera laissé à l'un de nos experts industriels[19], qui recommande impérativement, en 2012, le terme de « récupérateur », présent dans la légende du schéma de la Coupe verticale du bloc réacteur Superphénix, de chez Framatome, en 1985. L'ensemble de ces choix constitue un bon indicateur de la variation dans la construction d'un néologisme : les termes les

[19] De chez Areva.

plus positifs sont naturellement privilégiés par les experts du métier. D'une part « récupérateur » évoque quelque chose de plus positif que « cendrier », et d'autre part, la mention du si redoutable corium est évitée.

## 3. Spécificités du secteur du nucléaire dans les propositions néologiques

### 3.1. Après l'expérience du chantier « institutionnel »

Ce qui prévaut dans le secteur du chantier institutionnel, c'est, on l'a vu, un effort de « transparence » dans le choix des termes, ainsi que dans la bonne formulation des définitions qui les accompagnent. Les termes et les définitions doivent être clairs et compréhensibles par le plus grand nombre, et, en principe, en lien avec le suivi de l'actualité. Mais il faut aussi souligner la prudence et la réserve dans l'expression de ce domaine « sensible » qu'est le nucléaire.

Il est par ailleurs difficile de savoir exactement quels sont les termes dont la création est « nécessaire », pour reprendre encore les termes des consignes du décret du 3 juillet 1996 cité plus haut. Par exemple, y a-t-il bien un consensus sur cette nécessité de rendre « transparent » pour le plus grand nombre de lecteurs, pour un public élargi, un vocabulaire qui traite de sujets particulièrement délicats comme l'est précisément celui du nucléaire, et qui parfois divisent la société ? C'est une question qui se pose de manière récurrente[20]. D'ailleurs nous avons souligné plus haut des sortes de concepts et de « termes tabous », phénomènes qui peuvent sans doute mener à laisser de côté certains concepts ou certains termes, ou qui peuvent encourager des « néologismes artificiels » au détriment de formations néologistiques « naturelles » ou « spontanées ».

Ce recours à la « transparence » risque par ailleurs d'être trompeur. En effet, dans le domaine à l'étude, nous l'avons dit, le vocabulaire provient en général du fonds de la langue française générale. Les membres des commissions appartiennent eux-mêmes à des spécialités diverses et ne sont donc pas toujours eux-mêmes familiers des concepts dénommés. Ils passent parfois par une étape qui les fait réagir, à propos des termes qui leur sont proposés, en fonction de leurs propres connaissances et expériences. Cela est tout à fait naturel mais il y a là un danger : des termes qui ont aussi un usage en langue générale, qui ont une valeur générale et qui revêtent une forme déjà familière au grand public peuvent en eux-mêmes être trompeurs et orienter l'appréciation du terme dans une mauvaise direction. Et c'est cette étape qui risque d'engendrer, momentanément, des erreurs de jugement et de réaction envers un néologisme terminologique. Une fausse transparence peut influencer de manière erronée

[20] Et que nous avons pu aborder lors du colloque « Les chantiers du nucléaire. Quelles approches du nucléaire par les sciences humaines et sociales ? » (Lyon, ENS, 27-29 juin 2012).

l'approche du néologisme créé ou rencontré. Ce sont plusieurs aspects du sentiment néologique qui interviennent ici.

### 3.2. Après l'expérience du chantier « d'immersion industrielle »

Même si les deux terrains d'analyse que nous avons présentés ont des points en commun, des différences entre les deux expériences sont perceptibles. Dans le cadre industriel, la sensibilité à l'usage, ou plutôt aux usages, l'emporte sur la normalisation. On relève, plus encore qu'en terminologie officielle, une sensibilité aux connotations et à l'aspect social de la « correction politique », avec une éventuelle retenue dans l'expression, et le recours à l'euphémisme et au tabou. Cette prudence s'observe surtout dans les efforts d'ouverture aux médias.

Mais la conscience terminologique du niveau de spécialisation d'un terme néologique ne pose pas vraiment de problème aux spécialistes, et c'est à l'analyste linguiste que revient la charge de faire la part des choses dans les informations récoltées.

La langue de spécialité mobilise les connaissances de la langue générale et il existe naturellement un continuum entre la langue générale et la langue de spécialité. Ce fait ne représente pas de danger dans l'appréciation des néologismes candidats dès lors que l'on est entre spécialistes. Malgré tout il subsiste une zone d'ombre pour toute personne extérieure à la communauté d'experts utilisant cette langue de spécialité, qu'elle soit terminologue ou linguiste. Cette zone d'ombre, c'est la connaissance et le savoir. La terminologie du nucléaire est marquée par une forte culture d'entreprise, et en même temps, les experts éprouvent le besoin de définir leur terminologie. Même s'ils parlent la même langue que les non-spécialistes, les connaissances que véhiculent leurs discours n'en sont pas pour autant plus accessibles. Mais on a tout intérêt à s'incliner devant eux en respectant et en enregistrant leurs néologismes – en les aidant cependant à rédiger leurs définitions.

## Conclusion

Le nucléaire est une science jeune, dont le vocabulaire est souvent d'origine française, et où les néologismes sont le plus souvent des néologismes de sens. Les termes « activité », « cendrier », « colis » ou « gaine » en sont de bons exemples. Plus encore que pour d'autres domaines, il est donc important, pour qui veut repérer ces néologismes et en relever des spécificités liées à leur mode d'implantation et de circulation, de travailler en relation étroite avec les spécialistes du domaine. C'est ce que nous avons fait, dans deux cadres différents. Il s'agit d'une part du processus de la terminologie et de la néologie officielle, et d'autre part d'une expérience de travail en milieu industriel. Les deux expériences, éloignées l'une de l'autre, présentent pourtant des points communs, certains acteurs étant d'ailleurs partiellement présents sur chacun des

deux terrains. En effet, si elles diffèrent par leurs objectifs, ce qui se ressent dans les résultats affichés, c'est-à-dire les néologismes relevés, elles présentent finalement des points communs.

Des néologismes apparaissent, pour être ensuite supplantés par d'autres, mais pour réapparaître ultérieurement. On relève ainsi un usage récurrent de « traitement du combustible irradié », ou encore de « retraitement » qui apparaît et qui repart successivement pendant un demi-siècle. Un mouvement cyclique affecte ces dénominations qui sont typiques de néologismes problématiques rencontrés. Un autre aspect marquant des corpus étudiés a trait aux phénomènes de spécialisation. Ainsi en va-t-il, pour rester dans un cas de figure proche de ce qui vient d'être dit, des dénominations évoquant le traitement du combustible irradié : elles vont de « traitement », tout simplement, à « retraitement des combustibles irradiés ». Si cette série de termes a été discutée de manière récurrente, c'est aussi parce que des formes usuelles peuvent acquérir des nuances nouvelles, et devenir de nouveaux néologismes. Comme nous le voyons dans ces exemples, il arrive qu'au fil des années et des évènements, les termes qui, en principe, dénotent la même réalité fluctuent.

Des fluctuations, de différentes natures, sont donc perceptibles. Et, malgré des volontés clairement affichées d'interventionnisme, ce sont les usages qui, finalement, « fixent » les termes, les lexicalisent, pour, dans une première étape, les ériger en néologismes terminologiques. Voilà pourquoi « retraitement » perdure, s'est implanté ou, plutôt, se réimplante régulièrement. Exprimer le progrès scientifique ne peut se faire sans l'approbation du public, sans une réaction favorable du public quant à la bonne réception des termes nouveaux.

## Références

BOURIGAULT D., FABRE C., FRÉROT C., JACQUES M.-P., OZDOWSKA S. (2005), in « Syntex, analyseur syntaxique de corpus », *Actes des 12èmes journées sur le Traitement Automatique des Langues Naturelles*, atelier EASY : campagne d'évaluation des analyseurs syntaxiques, Dourdan, France, Vol. II, 17-20.

CALBERG-CHALLOT M. (2012), *Dynamique de la langue et de la terminologie dans le domaine du nucléaire*, Thèse de doctorat, Université Paris Diderot.

CALBERG-CHALLOT, M., CANDEL D., BOURIGAULT D., DUMONT X., HUMBLEY J., JOSEPH J. (2008), « Une analyse méthodique pour l'extraction terminologique dans le domaine du nucléaire », *Terminology*, Vol. 14/2, 183-203.

CALBERG-CHALLOT, M., CANDEL D., ROCHE C. (2007), « De la variation des usages au consensus terminologique : vers un dictionnaire de l'ingénierie nucléaire », *Actes TOTh 2007*, 119-141.

CANDEL D. (2010), « Pour une évaluation de la pratique néologique dans les commissions de terminologie et néologie en France », In CABRÉ I

CASTELLVI M. T. *et al.* (dirs), Cineo *Actes del I Congrés Internacional de Neologia de les Llengues Romaniques*, Barcelona, Institut Universitari de Linguistica Aplicada, 433-445, Barcelone, Université Pompeu Fabra.

CANDEL D., CALBERG-CHALLOT M. (2007), « Néologie en marche, motivations et réactions. Comment la terminologie du nucléaire évolue », *Actes du XXIXe Colloque international de Linguistique Fonctionnelle*, 45-49, Université d'Helsinki.

CANDEL D., TOMBEUX V. (2008), « Aspects de la néologie de spécialité en lexicographie générale : à propos du nucléaire dans *le Trésor de la langue française* », In SABLAYROLLES J.-F. (dir.), *Néologie et terminologie dans les dictionnaires*, 101-125, Paris, Champion.

DELAVIGNE V. (2001), *Les mots du nucléaire. Contribution socioterminologique à une analyse des discours de vulgarisation*, Thèse de doctorat, Université de Rouen.

DJAMBIAN C. (2010), *Valorisation d'un patrimoine documentaire industriel et évolution vers un système de gestion des connaissances orienté métiers,* Thèse de doctorat, Université Jean Moulin Lyon 3.

FranceTerme, Base de données des termes publiés au *Journal officiel* par la Commission générale de terminologie et de néologie : *franceterme.culture.fr*

HUMBLEY J. (2003), « La néologie en terminologie », In SABLAYROLLES J.-F. (dir.), *L'innovation lexicale*, 261-278, Paris, Champion.

SABLAYROLLES J.-F. (2003), « Le sentiment néologique », In SABLAYROLLES J.-F. (dir.), *L'innovation lexicale*, 279-295, Paris, Champion.

# Linguistic and sociolinguistic factors that influence the detection, implantation and circulation of natural terminology in academic uses of Basque

Igone Zabala[1], Itziar San Martin[2] & Mikel Lersundi[3]

## Abstract

Nowadays the Basque language is used quite extensively within university courses. 46.72 % of students study their degree completely or partially in Basque. The use of Basque as a vehicular language within university courses will determine to a large extent the future specialists' knowledge and the adequate use of terminology, and it is definitely an option that favors and helps the use of the language in professional domains. Consequently, the natural development and self-regulation of the terminology carried out in university classrooms cannot be disregarded. This article describes the factors that hinder the description, circulation and fixation of terminology used in university teaching. It describes the motivation for the Weaving Terminology Networks (TSE) project, whose aim is precisely to compensate for those hindering factors: the lack of fluid networks among experts and the inaccessibility of the texts used in academic communication. The corpus we have created in the TSE program, as well as the opinions of the participant experts that have been collected in forums and seminars have enabled us to detect other difficulties related to the language policy of the university system that one might consider as prescriptive and interventionist.

## Keywords

Spontaneous neonymy, natural self-regulation of terminology, Basque academic registers

## 1. Introduction

The aim of this paper is to evaluate the language policy of the University of the Basque Country (UPV/EHU). It describes the achievements with respect to the status and corpus planning, and concludes that there are some deficiencies in the corpus planning that may obstruct the circulation and fixation of natural terminology in university level academic registers in Basque. Some of these factors are the result of the current sociolinguistic situation of Basque, but

[1] University of the Basque Country (UPV/EHU), igone.zabala@ehu.es
[2] University of the Basque Country (UPV/EHU), itziar.sanmartin@ehu.es
[3] University of the Basque Country (UPV/EHU), mikel.lersundi@ehu.es

some others are the consequences of intervention by language standardizing and correcting agents.

Basque is a minority language and its standardization process started in the late 1960s. Its official status in 1982 brought about the extension of the language to formal contexts such as education, administration and mass media. The process of revitalization of Basque has been quite successful: the sociolinguistic inquiries since 1991 have shown a constant growth in the number of bilingual speakers. The extension of Basque to academic contexts has been decisive in the revitalization of the language. The fact that at university level most majors are offered in Basque shows that a real use of the language exists in all academic specialized fields. University teachers pass on specialized knowledge of the area to the students that intend to become part of the experts' community. The existence of academic communication between teachers and students suggests that there is a natural development of the specialized registers in the different areas. However, since specialized knowledge is developed and fixed in international discourse communities that employ English or other major world languages, communication in Basque among teachers in an area is not as fluid as among teachers and students. This makes the circulation and fixation of terminology more difficult.

The intervention carried out by language standardizing and correcting agents does not involve planned systematic terminology work: interventions are mostly non-systematic and the natural lexical updating that occurs in university classrooms is ignored. Moreover, the interventions depend excessively on official terminology and on the general normative dictionary (*Hiztegi Batua*). On the other hand, planned official terminology and general and specialized dictionaries blindly extrapolate the decisions of the normative general dictionary of Euskaltzaindia (Academy of the Basque Language) to specialized entries, which sometimes involves a setback in the lexicalization of terms that are well-established in usage. In our view, these attitudes are related to the fact that the process of elaboration of specialized registers in Basque occurs simultaneously with the standardization of the general language. Thus, the linguistic control over academic texts and dictionaries sometimes mixes the aims and criteria of the standardization of the general language and those of the functional elaboration of specialized registers (Elordui & Zabala, 2009): the former implies the reduction of dialectal variation whereas the latter should seek the promotion of functional variation.

The project *Terminologia Sareak Ehunduz* project (Weaving Terminology Networks, hereafter TSE) was specifically designed to overcome some of the problems described above. It started in the University of the Basque Country (UPV/EHU) in 2008 and its main aim is to make the texts and the terminology employed by teachers visible for other experts and for linguists (Zabala et al., 2011). It allows us to access the experts' teaching materials and the terminology contained in them. Besides, the seminars and forums in which the teachers

have participated enable linguists to get first-hand information about the experts' opinions and attitudes about terminology. All these elements (texts and opinions) are essential considering the theoretical assumptions we are making in our work about how semantic categories are conceived by the experts, how they activate the specialized value of lexical units and how terms circulate across discourse communities. The basic idea we are assuming, which is shared by all new approaches to terminology, is that real terminology can only be studied in discourse. Thus, real texts and the opinion of the experts themselves are central to any study about real terminology. More specifically, we assume that the specialized value of lexical units is activated by experts in specialized uses of language (Cabré, 1999, 2001). We also assume that understanding the circulation of terms among specialists, their reactions with respect to neologisms, and their resistance to official terminology requires a socioterminological approach, which focuses in the implantation of terms in discourse rather than in their standardization (Gambier, 1987: 320).[4] Hence, real terminology can only be studied in discourse: «Texts provide data on how particular authors understand elements of the world, how they understand the existing lexical items which serve to communicate about these elements of the world and how they may be brought to the creation of new lexical elements» (Temmerman, 2000 : 40).

The new approaches to terminology, whose assumptions we have adopted in this work, have shown that terminology serves two different purposes: a strictly representational function of specialized knowledge, and the function of communication, development and transfer of knowledge (Cabré, 1999 : 40). In the representational role, terminology is basically symbolic, and it is mainly univocal. In contrast, in the communicational function, terminology participates in the variation that is intrinsic to natural languages. This differentiation in the functions of terminology leads us to a distinction between descriptive and prescriptive activity. Any study related to the communicative function of terminology must necessarily have a descriptive starting point (Cabré, 2001 : 27). The so-called General Theory of Terminology (TGT) (Wüster, 1979) was conceived as a contribution to communicational univocity and it is effective in prescriptive and strongly structured contexts in which univocal communication is prioritized, such as in international standardization, documentation, and in artificial intelligence tasks. In contrast, in situations involving natural and social communication tasks, such as the development of minorized languages, a communicational approach is required (Cabré, 1998 :12, 2001: 27). The new approaches to terminology distinguish between standardization and normalization of terminology. Whereas the former turns a specific reference form into a norm, normalization refers to making a form normal, usual or

[4] This approach by Gambier was among the perspectives that led to the development of the so called Socioterminology. Some of the works that can be placed within this framework are: Corbeil (1988), Boulanger (1991), Guespin (1993), Gaudin (1993), Auger (1999), Quirion (2003).

habitual, and it may either refer to the intervention of an organization in order to establish a preference of certain forms over others or to the fixation of certain variants through self-regulation (Cabré, 2003). Thus, in contexts of natural communication, the forms that are normalized by institutions (official terminology) constitute just one (but not the only) of the variants that appear in specialized discourse: the adequacy of forms or denominations has a more central role than the correction or normalization of forms (Cabré, 2002).

Thus, these theoretical assumptions carry some methodological implications concerning languages in the process of development and elaboration, which we have adopted in the TSE project:

- Any normalizing initiative of terminology should be based on a previous description of real usage by specialists.
- Any normalizing initiative whose aim is to contribute to the development and elaboration of a language that is in the process of normalization must take into account the communicative function of terminology and its intrinsic functional variation.
- In a language in the process of revitalization and elaboration, disregarding the experts' real use of terminology may hinder its natural development and self-regulation carried out in the experts' discourse communities.
- The terminology used by experts in a language in the process of development and elaboration must be described and analyzed from a dynamic and diachronic perspective, so as to identify the trends that best respond to the development.

This paper has 4 sections. Section 2 describes the process of revitalization of Basque from the beginning of the standardization in 1968 and the creation of the Basque Autonomous Government in 1978. Section 3 is concerned with the general expansion of Basque to academic fields and particularly to the university. It describes the achievements and deficiencies of the language policy of the public University of the Basque Country (UPV/EHU), and it shows how the Weaving Terminology Networks (TSE) project tries to compensate for some deficiencies derived from sociolinguistic factors. In section 4 we describe the factors that obstruct the circulation and fixation of natural terminology in academic Basque and we show how intervention sometimes blocks the natural development and self-regulation of terminology carried out within the discourse communities of the experts. The last section provides some concluding remarks.

## 2. The process of revitalization of Basque

Basque is a minority language that co-exists in a linguistic community where the dominant languages are two major world languages: French in the northern part of the Basque Country and Spanish in the southern part. French and Spanish

are the official state languages in France and in Spain, respectively, and this status is linked to the right and the duty to know them. The status of Basque varies in the different regions of the Basque Country (Euskal Herria). In the Spanish side, the law on Basque or Euskara (1982) made Basque co-official together with Spanish in the so-called Autonomous Community of the Basque Country, which contains the territories of Bizkaia, Gipuzkoa and Araba. However, in the territory of Nafarroa, Basque has only the status of an official language in the "Basque-speaking zones". In fact, the Navarrese law on Basque (1986) determines the areas of the territory in which Basque is official. Finally, in the French Basque Country, the legal status of Basque is radically different, since this region lacks political autonomy and French is the only official language of the republic.[5] We must point out that in this work we are taking into account data and facts concerning basically the Autonomous Community of the Basque Country, where language policies and initiatives aimed at revitalizing the Basque language are more active.

Since the creation of the Autonomous Government (1978) for three of the provinces in the Basque Country (Bizkaia, Gipuzkoa and Araba) and the recognition of Basque as an official language in this area, Basque has been introduced in administration, public education and mass media. Thanks to the change in the status, the process of revitalization has been quite successful. Sociolinguistic inquiries since 1991 have shown a constant growth in the number of bilingual speakers, especially among the 16-24 year old age range. This growth has been accompanied by more transmission of the language. In fact, the 5th sociolinguistic inquiry made in 2011 showed the existence of 600,050 bilingual speakers, 181,000 more speakers than in 1991.

However, the revitalization of the language is not restricted to the increase in the number of speakers: the development of the language itself is also necessary. In the first half of the 20th Century Basque was restricted to private use. It was losing speakers, dialects were increasingly more different from each other and therefore understanding was hindered. The creation of a standard was necessary for several reasons: to avoid further differentiation of dialects, to extend language to formal domains and to gain speakers. The Royal Academy of the Basque Language started the process of standardization of Basque in 1968. It is important to note that the codification of the general standard overlaps with the elaboration of specialized registers, as the latter was already being carried out since 1970. This overlap has, as a consequence, a great instability in the development of specialized registers, which are based on the standard code and are continuously being adapted to the evolution of this code.

[5] About the legal status of Basque see López Basaguren (2012).

Table 1 provides a chronogram of the most relevant landmarks in the codification of the general standard Basque and in the elaboration of specialized registers.[6]

The Royal Academy of the Basque language (Euskaltzaindia) started the process of standardization in 1968 and, for this task, a compositional model based on the central dialects was followed. In the 1960s the orthographic and morphosyntactic rules for the standard variety were codified; during the 1980s the rules for adapting loanwords were set down and the rules for word formation were thoroughly described. A better knowledge of the word creation rules turned out to be of paramount importance to drive the development and analysis of specialized vocabulary that was necessary for the modernization of the language. However, the codification of the normative dictionary (*Hiztegi Batua*) did not start until the 1990s, and its first edition was not published until 2000.

The extension of the use of the language into public domains that require formal registers (teaching, administration and mass media) happened faster than the codification of the standard variety. By 1970 a small group of scientists and engineers started to elaborate on the language for specialized uses, mainly by creating teaching materials and popular science articles (Irazabalbeitia, 2002; Etxebarria, 2002). In 1972 the cultural association Elhuyar was created, whose aim was to promote science in Basque. In 1974 the popular science journal in Basque *Elhuyar* was created. It was at this time that the Basque Summer University (Udako Euskal Unibertsitatea) was created with the aim of promoting the creation of a Basque University. We may say that in those first few years the elaboration, diffusion and implantation of terminology occurred in a parallel fashion, since the discourse community was very small and terminological work was being developed by general consent.

Along the same lines and in coordination with the dynamics described above, in 1977 UZEI (The Basque Institute for the University) was created, with an aim to produce and spread terminology in order to facilitate the use of Basque in the university. According to Irazabalbeitia (2002), by 1980 the foundations for making science in Basque were set: a basic lexicon, basic nomenclatures and basic phraseology. In 1986 UZEI created the terminological database *Euskalterm* and in the same year the Basic Law of Normalization of the Use of Basque came into effect. In 1987 UZEI came under the tutelage of the Basque Government. In 2001 *Euskalterm* became the Basque Public Term bank and in 2002 the Commission for the Normalization of Terminology (*Terminologia Batzordea*) was created.

[6] For the steps of standardization we use the terminology in Haugen (1983).

| | CODIFICATION OF THE GENERAL STANDARD BASQUE | ELABORATION OF SPECIALIZED REGISTERS |
|---|---|---|
| 1968 | Selection (Standard) | |
| 1970 | Codification (Graphization & Grammatization) | The creation of specialized glossaries. Usage of Basque in scientific texts by scientists and engineers |
| 1971 | | *Elhuyar* group is created for promoting the use of Basque in science |
| 1974 | | *Elhuyar* journal and the Basque Summer University (UEU) are created |
| 1977 | | UZEI institute is founded for creating and diffusing terminology |
| 1980 | Norms for adapting loanwords / word creation rules | |
| 1982 | | EIMA committee is created for the linguistic control of primary and secondary school teaching materials |
| 1986 | | *Euskalterm* (UZEI) (Terminology database for the diffusion of terminology) is set up |
| 1987 | | UZEI becomes an entity supervised by the Basque Government |
| 1990 | Lexical codification of the normative dictionary starts | |
| 2000 | First publication of the normative dictionary | The Basque Language Service at the university is created for the creation and linguistic control of materials for university teaching |
| 2001 | Continuous feeding of the normative dictionary | *Euskalterm* becomes the Basque Public Term Bank |
| 2002 | | The Official Commission for Terminology is established |

*Table 1: Chronogram of the codification of the standard Basque and the elaboration of specialized registers.*

As the chronogram in Table 1 shows, there is an overlap between the most important steps of the codification process of the general standard and those of the institutionalized initiatives aimed at the elaboration of the specialized registers. Often, the same agents participate in both processes, and as we will show in section 4, this makes it difficult to distinguish clearly the aims of each process. On the other hand, we must consider the fact that the use of Basque in academic contexts has grown a great deal in the last four decades and hence, the community of teachers/experts and students that use specialized registers is different nowadays compared to the first group that used to decide on issues about Basque specialized registers by consent. Nowadays, it is unthinkable to carry out terminology planning without taking into account the users themselves and their real use of terminology. Thus, the official language and terminology planning for Basque requires all the processes described in Auger

(1986): systematic collection and study of notions and terms that are being used, normalization, diffusion, implementation, evaluation and control, and updating.

## 3. Basque in academic contexts

The extension of Basque to public education (primary and secondary) and to the university system has been of key importance for achieving a growth in the number of speakers, for the implementation of the general standard code, and for the development of the academic registers and terminology. Most parents choose the educational model of full immersion in Basque (model D). Thus, 70.36 % of the students registered in primary school in the academic year 2011-2012 chose model D. 61 % of the students aiming to enter university took their entrance exams in Basque. 46.72 % of the university students study their degree completely or partially in Basque; in the year 2011-2012, 78 % of the compulsory credits were offered in Basque and 45.10 % of the teachers were bilingual. The use of Basque as a vehicular language in the university studies will determine to a large extent the future specialists' knowledge and adequate use of the terminology and it is definitely an option that favors and helps the use of the language in professional domains. However, due to the fact that Basque is still in the normalization process, the input that the students receive from their teachers and teaching materials is not considered as sufficient in order to fully develop their linguistic competences in specialized registers. In order to compensate for these deficiencies, all graduate studies offer two optional subjects that deal with linguistic issues and provide the students with opportunities to improve upon their communicative academic and professional skills in Basque. Other refresher courses are also offered to the teachers, in which help is provided to adapt to the processes of standardization and elaboration of the norms of usage of the general language.

From the data provided above, we may deduce the importance of academic communication in Basque. Subjects from all areas are taught in Basque within university, and it is obvious that no teaching can be carried out without terminology. In order to satisfy the demand that exists to study in Basque, two elements are necessary: teachers/experts of the areas capable of teaching in Basque, and handbooks and textbooks that complement the teaching. Sometimes, teaching handbooks and other types of publications are created collaboratively by individual teachers themselves or by groups of teachers. On the other hand, the Vice-chancellorship of Basque and Multilingualism of the University has made great efforts in translating many reference manuals of various disciplines with the aim of covering the deficiencies that are still evident in University level Basque teaching materials. This work is typically conducted by translators, and afterwards, some expert teacher checks the translation. In any case, all publications that aim to be published by the University of the Basque Country must necessarily be reviewed and accepted by the Language Services (*Euskara Zerbitzua*). These activities aim at compensating for the lack

of published materials in Basque. However, it is necessary to point out that there exist other reasons that are related to linguistic control. In fact, it is often believed that the texts produced by translators are more correct and genuine than the ones produced directly by experts or teachers.[7] On the other hand, based on the traditional idealized view of terminology expressed by Wüster (1979), any terminological variation is often viewed as the sign of a lack of normalization, and it is often assumed that linguistic control of terminology must include the reduction of variation that it is intended to eradicate.[8] Thus, it is assumed that the variants that have been accepted and prioritized in the texts that have been checked and published by the Language Services as well as the new terms proposed in such texts establish a precedent that needs to be taken into account when processing future texts. Thus, controlled texts are implicitly assigned a semi-prescriptive function. There are no precise studies about the implantation of the terminological proposals contained in the controlled texts of the Language Services of the university and it is difficult to evaluate the success rate of such a policy. However, as will be shown in section 4, there is evidence that interventionist and prescriptive policies can block the development and natural self-regulation of real terminology that is being used.

### 3.1. Achievements and deficiencies of the language policy at the University of the Basque Country (UPV/EHU)

Undoubtedly, the language policy of the University of the Basque Country (UPV/EHU) has obtained notable achievements regarding the status and corpus planning of academic Basque. Nevertheless, we have detected some deficiencies that might suggest that this language policy may be hindering the natural development of academic registers rather than promoting it.

As for the achievements, there has been a considerable increase in the number of bilingual teachers: the rate of bilingual teachers grew from 13 % to 45.2 % between the years 1988 and 2011. This increase in the number of users of academic registers in Basque is significant. On the other hand, there has been a considerable growth in publications of translated manuals and of teaching materials created by teachers online: 30 manuals and 143 online teaching materials were published between 2007 and 2012. Besides, we must point out

---

[7] This seems to be a belief that has no real basis. In Zabala et al. (2012), we have shown that in the case of terminology in human anatomy, linguists and professors of this area have collaboratively worked and have managed to elaborate and agree upon terminology that is more correct and patrimonial than the terminology proposed initially by translators.

[8] In the languages in the process of normalization, asystematic variation is typically greater than in normalized languages, but, in contrast, because of the lack of functional development in such languages, functional variation is typically underdeveloped. A study of variation in texts of genomics of different levels of specialization (Elordui & Zabala, 2005) has proven that Basque terminology shows signs of some development of functional variation.

that there has been important progress in terms of the quality of published texts, as well as regarding the grammaticality, naturalness and the implementation of the standard code. Finally, it is also important to note that there has been a growth in the production of end-of-degree projects, of master theses and doctoral dissertations written and defended in Basque.

Nevertheless, an analysis of such linguistic policies from the new linguistic-communicative and sociocognitive approaches to terminology reveals two types of deficiencies: on the one hand, we have detected deficiencies that are strictly due to sociolinguistic reasons which have not been taken into account when designing the language policy of the institution. Others are due to the prescriptive and interventionist policy that is being applied. This policy is based on assumptions and practices that have been proven to be insufficient and ineffective in contexts of natural and social communication in which the development of a minorized language is being promoted.

As for the former type of deficiencies, it cannot be ignored that there is a lack of fluid communication networks among teachers. The fact that university teachers pass on specialized knowledge of the area to the students that intend to become specialists suggests that there is a fluid and well established communication between teachers and students and between students in general. However, the communication between many of the experts that use Basque in teaching is very limited, since specialized knowledge is developed and fixed in international discourse communities that employ English or other major world languages. Thus, there is a great asymmetry between the use of Basque among future experts in teaching and the use among experts, which is mostly related to research activities.

On the other hand, since real communication in Basque occurs for the most part in university classrooms, this type of communication mostly consists of oral and written spontaneous texts that teachers elaborate upon and check in the everyday teaching-learning process. As a consequence, although the methodology derived from the new approaches to terminology requires access to the real terminology used in the discourse, it is really difficult to access the texts created in real communication contexts.

Other deficiencies are related to the intervention over academic registers that is being made by checking texts for publication, through the refresher courses that are offered to teachers and through the recommendations that are often issued by mailing lists and by the online reference and consultation websites. We have detected three types of deficiencies: on the one hand, it seems that the aims and criteria of the functional elaboration of specialized registers are mixed with those of the codification and implementation of the general language. In fact, we have observed that there is a uniform application of the stylistic recommendations and norms of the general language that do not acknowledge the specificities of specialized communication in the activities carried out by the

Basque Language Services (checking of texts, courses for teachers, online recommendations). On the other hand, non-systematic intervention in terminology is being carried out without systematic terminology planning. This intervention is based on the premise that terms must be univocal and that any kind of variation must be reduced. The intervention is also mainly based on purely linguistic considerations, leaving aside sociolinguistic and terminological criteria. Lastly, the interventions of the Language Services blindly prioritize the variants collected in general and specialized dictionaries as well as in the *Euskalterm* terminological database over other variants used by experts, without taking into account the fact that the variants collected in the former are often lexicographic or terminological proposals whose real implantation in discourse has not yet been studied and evaluated.

These deficiencies in the approach, aims and practices of the language policies of the university suggest that the linguistic control and intervention that is being made over academic texts and over the linguistic competence of the users that directly participate in specialized communication may hinder the natural development of specialized registers, and especially of terminology.

### 3.2. The Weaving Terminology Networks (TSE) Project

The Weaving Terminology Networks (TSE) project (Zabala et al., 2011) was designed to help overcome some of the difficulties and deficiencies described in section 3.1. It started in 2008, and the main goals were to compensate for the lack of fluid communication networks among teachers and to make real texts and terminology visible for consultation. The teachers that participate in the TSE program upload their teaching material to a documentary database. Then, the texts are linguistically processed and published in a consultation interface (*Garaterm* corpus) (Zabala et al., in progress). On the other hand, the semiautomatic terminology extractor *Erauzterm* (Alegria et al., 2004) extracts the term candidates contained in the texts and the teachers validate the terminology of their subject. Next, the teachers create a plurilingual glossary and assign the Basque terms the equivalents in other languages. Finally, the glossaries obtained by this methodology are dumped in a tool that allows online consultation of terms: *Terminologia Zerbitzurako Online Sistema* (TZOS) (Arregi et. al., 2010). Thanks to the processes and tools that we are using in the TSE program, we manage to give access to the texts and terminology employed by the teachers in the teaching-learning activity of their subject. Moreover, the TSE program is a training program too, and we discuss the terminology and texts of the teachers in forums and seminars.

The TSE project allows us to access teachers' teaching materials and terminology contained in them. Since 2008, around 200 teachers have participated in the project and we have created a corpus of about 6 million words, as well as about 70 plurilingual glossaries. Besides, the seminars and forums enable linguists to get first-hand information about the experts'

opinions and attitudes about official terminology and about the linguistic control that correctors exert over their texts. The experts' reflections have also provided us with evidence that, although neonymy in Basque is overwhelmingly secondary (aimed at providing the equivalents for the terms in other languages), the conceptualization that the experts have carried out through the texts they have processed and/or produced in other languages is crucial for understanding the linguistic motivation of the Basque equivalents. Besides, the experts' reflections have provided us with evidence that, in some cases, the intervention that is made on the teachers' texts (either through corrections of their texts or by the application of the recommendations that consultation websites offer) may interfere in the development of specialized terminology in Basque, or more specifically, in the creation, circulation, functional adaptation and natural fixation of terminology. Moreover, as we have pointed out in section 3.1., the type of intervention that is typically carried out is non-systematic, and depends largely on the "official" terminology collected in terminological databases and in dictionaries of different degrees of specialization. However, the promotion or imposition of specific variants may interfere with the natural dynamics of the development: the creation, circulation, and adaptation of terminology in real use involve a very dynamic process in a language like Basque, which is in the process of revitalization and elaboration.

In our view, no intervention over terminology and over specialized registers should be made without a previous description and monitoring of the development and self-regulation of real terminology. Such monitoring is also necessary to evaluate the implantation of official terminology in discourse. It is this conviction that is the basis of the TSE project.

## 4. Factors that hinder the circulation and fixation of natural terminology in academic Basque

The TSE program has allowed us to access the teaching texts from the many teachers that use Basque in their classes. Besides, through seminars and forums we have had the chance to exchange with the experts about their linguistic practices. This interaction has allowed us to realize that they are open to changing linguistic practices when linguists suggest more correct, more genuine or more efficient options from a communication perspective. Moreover, we have obtained first hand information about the problems and reactions that experts express when they are faced with official terminology proposals that are intended to be implanted through linguistic control over academic texts. These are some of the problems we detected:

- Many variants that are used by experts are not collected in the Basque Public Term Bank.
- The decisions of the general normative dictionary are extrapolated blindly to the Basque Public Term Bank and to the linguistic control of academic registers.

- Lexical units of the general language that have activated specialized values are sometimes corrected, without taking into account pragmatic-discursive criteria, which blocks the necessary innovation for the development of terminology.
- The normative dictionary is an ongoing project and sometimes includes non-correct variants without collecting existing correct variants that are used by experts.

### 4.1. Variants used by experts that are not collected in the Basque Public Term Bank *Euskalterm*

Currently, term databases and terminological dictionaries still contain few entries. Besides, many of those entries are proposals which date from the 1980s, a time when Basque was being introduced into university studies. Besides, there is no research about the implantation of the terminology that *Euskalterm* offers, nor of the real terminology used by experts. As a consequence, the official terminology that is typically consulted by users is not updated nor adapted to the development and natural self-regulation of real usages. Many terms and variants that university teachers use are not collected in *Euskalterm*, and instead, it contains terms and variants that have not managed to be implanted in the discourse of the experts or variants that are being replaced by others that are more suitable for the pragmatic-discursive needs of a given specialized field. We must keep in mind that, when one variant of a term is used by experts but not collected in the official terminological database *Euskalterm*, linguistic intervention will tend to correct it and to replace it with a codified variant.

One example of the situation described above is the term *muscle*. In the *Euskalterm* database, we find the variant *muskulu* in 164 compounds and syntagmatic terms, but the patrimonial variant *gihar* does not appear. The latter was not collected in the database in the past, because this variant was not considered as adequate for the anatomical term. We have compared the occurrences of the variants *muskulu* and *gihar* in two general and two specialized corpora. As for the former, we have made searches in the general corpus *Ereduzko Prosa Gaur* (EPG) (25.1 millions of words), which contains literature and press texts collected between the years 2000 and 2006 and selected according to linguistic quality criteria. On the other hand, we have used the *XX. Mendeko Euskararen Corpus Estatistikoa* 'The Basque Statistical Corpus of the 20th Century' (XX.c.e.) (4.6 millions of words ), which is a reference corpus, statistically balanced and based on the collection of all texts produced in Basque between 1900 and 1999. As for the specialized corpora, we have employed *Zientzia eta Teknologiaren Corpusa* (ZT corpus) (8.5 millions of words), which is composed of science and technology texts of various degrees of specialization published between 1990 and 2002. Lastly, we have used the corpus created in the TSE project between 2009 and 2011. This corpus currently contains 6.2

million words from the teaching materials of the university teachers that participate in the program.

The analysis and comparison of general and specialized corpora reveals that the loanword *muskulu* is not the only variant that we find in the use of specialists, as Table 2 shows. On the one hand, the patrimonial variant *gihar* is attested more often in general corpora (EPG and XX.c.e) than in specialized corpora (ZT and TSE). However, it is noteworthy that the ZT corpus, which is composed of linguistically controlled texts, shows that *muskulu* has been used much more often than *gihar* (1,768 vs 140 occurrences), whereas the TSE corpus, which exclusively contains spontaneous texts used in university classrooms shows that *gihar* has up to half of the occurrences compared with *muskulu*. Most of the occurrences of *gihar* in TSE are attested in texts of anatomy in the faculties of Medicine and Dentistry, and in the faculty of Pharmacy. This shows that the experts in the area have activated the specialized value in the patrimonial variant *gihar*. The usage rate of these terms with respect to the total number of words of each corpus reveals that, compared to the other corpora, the TSE corpus contains a considerably greater number of texts of anatomy.

| | General corpora | | Specialized corpora | |
|---|---|---|---|---|
| | EPG<br><br>25.1 M of words | XX.c.e.<br><br>4.6 M of words | ZT corpus<br><br>8.5 M of words | TSE corpus<br><br>6.2 M of words |
| *gihar* | 498 | 118 | 140 | 1,241 |
| *muskulu* | 246 | 64 | 1,768 | 2,798 |
| Occurrence rate of the concept (p. M of words) | 29.6 | 39.6 | 224.5 | 651.5 |

*Table 2: Comparison of the occurrences of the variants gihar and muskulu in general and specialized corpora that contain linguistically controlled texts or spontaneous texts.*

Considering all this, a socioterminological approach that takes into account the evolution of real terminology in discourse should acknowledge the use of the patrimonial variant and at least collect it as a synonym. If not, the productive use of this term in anatomy will be blocked either in texts which are controlled and corrected or when the *Euskalterm* database is consulted by translators and teachers. We must keep in mind that Basque is in the process of lexical-discursive development and that the frequency in the use of variants should be considered from a dynamic and diachronic perspective. What is typically expected from such processes is an evolution from variants that are dependent on the dominant languages to patrimonial variants. In fact, in the seminars and forums with the experts, we have observed that they tend to choose the more genuine forms as reference terms or denominations.

### 4.2. Blind extrapolation of the decisions of the normative dictionary to terminology

The decisions of the normative dictionary of the Basque Language Academy are often applied and adopted blindly in the Public Term Bank. The aim of the normative dictionary is to codify the general standard lexicon and the methodology employed does not include the elaboration and analysis of specialized corpora. The general normative dictionary is also employed as a reference source in correcting academic texts. In fact, a study of the implantation of the entries of possible terminological value that are included in the normative dictionary revealed an implantation rate of 95 % in academic texts (Loinaz, 2007). This shows that there exists a strict linguistic control over such texts and that there is a tendency to prioritize the entries collected in the normative dictionary, disregarding all pragmatic-discursive criteria.

We find a clear example of such an intervention in the term *liseriketa* 'digestion', a term that has been widely and exclusively used for the last two decades in teaching in all areas in Basque. However, in the edition from the year 2000 of the normative dictionary this term was proscribed and the loanword *digestio* was prioritized. The reason for this decision was that the verb *liseritu* 'to digest', from which the term *liseriketa* was derived, was phonologically irregular. Although the normative dictionary does not specify any sign of field of specialization for the entry *digestio*, the term *Euskalterm* bank and the *Elhuyar* dictionary have applied this norm to all the terms that contain either the verb *liseritu* or the noun *liseriketa*. Thus, terms such as *liseri-aparatu* 'digestive system' are proposed as *digestio-aparatu*.

The change just described was strictly based on linguistic criteria and did not take into account sociopragmatic or terminological considerations. The consequence is a setback in the elaboration of the derivative and compositional paradigm in which the verb *liseritu* 'to digest' used to participate. In fact, if we compare the different editions of the *Elhuyar* dictionary, we realize that all the terms derived from the verb *liseritu* 'to digest' that used to appear in the 1996 edition, previous to the decision taken by the normative dictionary, have disappeared in the 2000 edition: *liserigailu* 'digestor', *liserigaitz* 'indigestible', *liserigarri* 'digestible', *liserigarritasun* 'digestibility', *liseriketa* 'digestion'. Moreover, only two of the derivatives have been replaced by adapted loanwords: *digestio* 'digestion' and *digestore* 'digester'. Note that the derivatives of the verb *digeritu* formed by the suffixes *-garri*, *-gaitz* and *-gailu* would be phonologically inadequate, since they would contain two velar sounds. The application of the rule of the general normative dictionary to terminology has caused a setback in the development of terminology through genuine Basque suffixation, since it has contributed to the disappearance of derived lexical units that were well-established in use.

The application of the rule that we are describing in this section has also had consequences in the system of compound terms that refer to organs of animals. These terms had traditionally been formed by compounds of the type [V+N]: *liseri-aparatu* 'digestion system' (lit. 'digest system'), *ugal-aparatu* 'reproductive system' (lit. 'reproduce system'), *arnas-aparatu* 'respiratory system' (lit. 'breathe system'). However, such terms have turned into compounds of the type [N+N] in the official terminology, which results in a much more asystematic paradigm: *digestio-aparatu* (lit. 'digestion system'), *ugaltze-aparatu* (lit. 'reproducing system'), *arnas-aparatu*[9] (lit. 'breath system').

### 4.3. Correction of lexical units of the general language that have activated specialized values

The teachers that participate in the seminars and forums of the TSE program often show resistance to the corrections that are made on characteristic terminology within their field which appears in their texts. Such corrections contradict the teachers' linguistic experience, and the reasons that are provided for such corrections are often not convincing for them. We will next describe two cases. On the one hand, we will deal with terms specific to the fields of molecular genetics (*adierazi* 'to express', *adierazpen* 'expression', referred to genes), and next, we will describe facts about a group of lexical items in the field of statistics (*estimatu* 'to estimate', *estimazio* 'estimation', *estimatzaile* 'estimator'). Unlike the former, the examples of statistics that we will deal with include terms that appear in almost any discipline, precisely because statistics contains the tools that are used in almost every knowledge area.

As for the expressions of the fields of molecular genetics such as *geneak adierazten dira* 'genes are expressed', *gene-adierazpen* 'gene expression', the experts often complain about the fact that they are obliged to change the patrimonial forms *adierazi* and *adierazpen* to the loanwords *espresatu* and *espresio*. This surprising correction is based on the fact that the general dictionaries collect the verb *adierazi* as a transitive dyadic or triadic verb (*hori adierazi dute* 'they have expressed that', *hori adierazi diote* 'they expressed that to him/her'). From the information gathered from general dictionaries, the language correcting agents deduce that the intransitive use of the verb is a syntactic *calque* that is not compatible with the Basque system. This is a restricted interpretation of the Basque linguistic system and it ignores the fact that it is the semantic features of the subject that determine the syntax of the verbs within the system of the language. Thus, when the subject is human, the sequence **Jon ondo adierazten da* 'Jon expresses himself well' may be considered as a syntactic *calque* that may even be considered as ungrammatical within the system of the language. However, as Zabala (2004) explains, *geneak* 'genes' are the molecular support of genetic information which is necessary for protein synthesis, and when it is the

[9] In the case of *arnas-aparatu,* the first element *arnas* 'breath' may either be analyzed as the root of the verb *arnas(tu)* 'to breathe' or as the noun *arnasa*, whose final *–a* has been deleted for being the first element of the compound.

subject of the sentence, its thematic role is *internal cause*. When the subject receives this role, Basque invariably employs the intransitive auxiliary *izan* 'to be': *zuhaitza loratu da* 'the tree has blossomed', *txoria lumatu da* 'the bird has plucked'. From these facts we may conclude that the justification for the intervention or correction described above is not right, since it is based on too strict and too rigid an interpretation of the lexical-syntactic system of Basque.

Temmerman's (2000) sociocognitive approach sheds more light on the development of the terminology of molecular genetics and helps us better understand the way in which the term at hand was created. From analyzing terminology within the discourse area, and from interviews carrie out with experts, she reaches the conclusion that in the life sciences there is a growth of understanding and knowledge through metaphorical reasoning, whose results are metaphorical lexicalizations for many (new) categories in the discipline. Metaphorical reasoning amounts to the understanding of a new fact, situation or process based on the imagined analogy with something one already understands. The result of analogical reasoning is the metaphorical naming of new categories with existing lexemes (Temmerman, 2000 : 69-71). Temmerman postulates that the domain metaphor underlying the understanding of molecular genetics is that heredity is based on information stored in our genes (DNA). Several sub-domains are expanded from this general domain and one of these is that «DNA is a language. Genes are messages written in a language», thus they are encoded in a sequence of letters, and they are transcribed into messenger RNA, which is then translated into protein. Protein synthesis is in this metaphorical reasoning the result of gene expression.

The Basque experts that have conceptualized the knowledge of their field by using the aforementioned metaphoric strategies express the concepts by using the corresponding Basque lexical terms in accordance with such a conceptualization: *kode* 'code', *kodetu* 'codify', *adierazpen* 'expression', *transkripzio* 'transcription', *itzulpen* 'translation'. Thus, a semantic-pragmatic analysis of the term *gene-adierazpen* clearly shows us that the Basque experts in molecular genetics have activated the specialized value of the lexical units by following the logic of the semantic-pragmatic conditions in which the discourse of the specialized area is developed. The term *gene-adierazpen* is collected in *Euskalterm* but, nevertheless, the Language Services of the university keep on correcting the term inertially in their publications. Moreover, they even recommend the term *gene-espresio* over the term *gene-adierazpen*.

As for the second example that we intend to develop in this section (*estimatu*, '*estimate*' *estimazio* 'estimation', *estimatzaile* 'estimator'), some participants in the TSE project show resistance to the corrections and suggestions to replace them with the neologisms *zenbatetsi, zenbatespen* and *zenbatesle*. The reason for this is that such corrections contradict their linguistic experience from the last decades. The justifications for the corrections are based on the idea that in Basque, *estimatu* and *estimazio* may only mean 'to appreciate; to be grateful' and

'esteem, regard; affection', respectively. However, the Basque General Dictionary (*Orotariko Euskal Hiztegia*), which collects the lexicon of literary tradition in Basque, contains the term with the meaning corresponding to statistics in *estimatu*, *estimazio* and *estimatzaile*. Moreover, *Orotariko Euskal Hiztegia* also contains the entry *zenbatetsi* with the following note: «"Tasar calcular" Azkue cita este neologismo, que no aprueba totalmente.» ("To rate, calculate" Azkue mentions this neologism, which he does not approve entirely). As for the normative general dictionary (*Hiztegi Batua*), although it accepts this entry, it also shows some doubts about its composition: «Forma onartzeko eragozpenak aipatu ditu lantaldeak. Berria da eta ez guztiz erregularra osaeraz; baina hedatua, eta onarpen eske aurkeztua.» (The work team has mentioned several problems in accepting the entry. It is a new form and it is not completely regular in its composition. Nevertheless, it is a unit which is widely used, and its inclusion in the dictionary has been requested). In fact, the verbs that are formed with *etsi* 'to have an opinion about' (*onetsi* 'to accept, to approve', *gaitzetsi* 'to disapprove', *balioetsi* 'to assess, to evaluate, to estimate; to value',…) include a nominal predicate that provides the main meaning: *ontzat hartu* 'to accept, to approve', *baliozkotzat hartu* 'to assess, to evaluate'… Nevertheless, the neologism *zenbatetsi* combines the verb *etsi* with a question word *zenbat* 'how much'), and its irregular nature makes the whole unit hard to interpret and to employ in syntactic compositions. From a discursive viewpoint, *zenbatetsi* does not seem to be a suitable solution for all the contexts in which *estimatu* may appear.[10]

On the other hand, if we compare the occurrences of the different meanings in general and specialized corpora, we see that in general corpora such as EPG, *estimatu* and *estimazio* mostly appears associated with the meaning 'to appreciate; to be grateful' (976 and 132 occurrences, respectively) and that it only contains the statistical meaning in a few examples (17 and 25, respectively). On the contrary, in specialized corpora (ZT and TSE), we only find the statistical meaning. As for the distribution of the variants *estimatu*, *estimazio*, *estimatzaile* and *zenbatetsi*, *zenbatespen*, *zenbatesle* with the statistical meaning, we see that the former are attested widely in all types of corpora:

[10] The database *Euskalterm* collects *zenbatespen* as a synonym of *estimazio* in many entries.

| | General corpora | | Specialized corpora | |
|---|---|---|---|---|
| | EPG<br><br>25.1 M of words | XX.c.e.<br><br>4.6 M of words | ZT corpus<br><br>8.5 M of words | TSE corpus<br><br>6.2 M of words |
| *estimatu* | 17 | 24 | 69 | 138 |
| *estimazio* | 25 | 24 | 32 | 172 |
| *estimatzaile* | 0 | 2 | 19 | 100 |
| Occurrence rate (p. M of words) | 1.7 | 10.9 | 14.1 | 66.1 |
| *zenbatetsi* | 6 | 2 | 4 | 24 |
| *zenbatespen* | 3 | 0 | 6 | 60 |
| *zenbatesle* | 0 | 0 | 0 | 1 |
| Occurrence rate (p. M of words) | 0.4 | 0.4 | 1.2 | 13.7 |

*Table 3: Number of occurrences of the entries* estimatu, estimazio, estimatzaile *and* zenbatetsi, zenbatespen, zenbatesle *in their statistical sense in general and specialized corpora.*

In conclusion, it seems that the intervention over the group of terms related to statistics is not backed by or grounded in solid linguistic-discursive and lexicographic justifications. Moreover, it intends to change the actual usage of variants that are widely used.

### 4.4. Inclusion of non-correct variants in the Normative Dictionary without collecting existing correct variants

The normative dictionary of the Basque Language Academy (*Hiztegi Batua*) is an ongoing project that is being fed with new entries and meanings as the committee that is in charge of its elaboration analyzes and discusses new word lists. In some cases, we find that some variants that are clearly incorrect have been included in the dictionary and, in contrast, other variants that are attested in real use by experts have not been analyzed or included. We find a paradigmatic example of this in the entry *konjokatu* 'conjugate'. This is an analogical form that is not acceptable semantically or structurally: it is composed of the loaned prefix *kon-*, which does not exist in Basque, and the verb *jokatu* 'play'. The word comes from the Latin form *conjugare* 'to yoke together'. Thus, there would be two correct options: the adequately adapted loanword *konjugatu* or the Basque compound form *uztartu*, which has been

formed by the noun *uztarri* 'yoke' and the verbal suffix *-tu*, which is sometimes used by experts and would be the patrimonial equivalent of the Latin parasynthetic derived form.

Collecting incorrect variants in the normative dictionary and not the correct and genuine variants blocks the circulation and implantation of the uses that experts make of the correct forms, which should be the objective of terminology planning. In correcting texts, it is often ignored or forgotten that the general and normative dictionary is an unfinished product and, as a consequence, many forms that are not collected in this dictionary are often rejected for the sake of forms collected in this dictionary.

## 5. Conclusions

In this work we have analyzed the language policy of the University of the Basque Country (UPV/EHU) and its contribution to the development of academic registers. The analysis has been done from the perspective of the new linguistic-communicative and socioterminological approaches to terminology. We have described the achievements of the policy, such as the increase in the number of students that study partially or completely in Basque, the growth in the number of bilingual teachers and in the number of credits that are offered in Basque, and the production of teaching materials and academic works (end-of-degree projects, master's theses, doctoral dissertations). We consider that these achievements are meaningful, since the use of Basque as a vehicular language in university classrooms is crucial for the development of specialized registers and terminology from all areas, and also for its diffusion in society through experts that are trained in such areas. The improvement in the linguistic quality of the texts controlled and published by the Language Services of the university can also be considered as an achievement in the grammaticality, the naturalness and the contribution they make to the implementation of the standard code. However, we have also detected some difficulties and deficiencies in this language policy:

- The process of elaboration of specialized registers in Basque occurs simultaneously with the process of standardization of the general language. As a consequence, the linguistic control of texts is often considered as a uniformizing activity which mixes the opposing goals and criteria of both processes. The goal of the standardization of the general language is the reduction of dialectal variation by establishing a standard variety, whereas the lexical-discursive development requires the promotion of functional variation.

- The use of Basque as a vehicular language in university classrooms is crucial for the development of specialized registers and terminology in all fields, and also for its diffusion in society through experts that are trained in such areas.

- The wide use of Basque as a vehicular language at university level implies the existence of the use of terminology in different areas. However, whereas a solid basis exists in communication between teachers and students, there are no fluid communication networks among experts within kwnoledge areas. This factor hinders the circulation and fixation of terminology.

- The description of the terminology used in academic communication is crucial for understanding the way it is created and self-regulated. Nevertheless, the fact that most of the communication is oral and carried out through spontaneous unpublished written texts makes it difficult for linguists and other experts to access such texts. This factor obstructs both the description of the terminology as well as the circulation and fixation of the terms created and used by the experts.

- The Weaving Terminology Networks (TSE) program was created with two aims: to describe the real terminology used by university teachers and to compensate for the lack of fluid communication networks among experts, helping the circulation and fixation of natural terminology.

- Based on our study of the texts and the opinions we have gathered from the experts that participate in the TSE program, we have detected three factors that we believe may be hindering and slowing down the natural development of academic terminology:
  - Planned official terminology and general and specialized dictionaries blindly apply the decisions taken for the general normative dictionary.
  - There is a lack of terminology planning in the linguistic services of the university (UPV/EHU), which is reflected in non-systematic intervention on terminology that sometimes lacks robust and plausible criteria.
  - Interventions made in translating and correcting academic texts excessively depend on official terminology and do not take into account the natural continuous updating of lexical elements that occurs in university classrooms.

## Acknowledgements

This research was supported by the University of the Basque Country and the Elhuyar Foundation (Garaterm-2 project [US10/01]), and by the Basque Government (WIKITZUL project [S-PE12UN082]).

## References

ALEGRIA I., GURRUTXAGA A., LIZASO P., SARALEGI X., UGARTETXEA S., URIZAR R. (2004), An Xml-Based Term Extraction Tool for Basque, In *4th International Conference on Language Resources and Evaluations (LREC)*. Lisbon.

ARREGI X., ARRUARTE A., ARTOLA X., LERSUNDI M., SANTANDER G., UMBELLINA J. (2010), TZOS: Terminologia Zerbitzurako On-line sistema, In X. ALBERDI & P. SALABURU (eds.), *Ugarteburu Terminologia Jardunaldiak,* 136-153, Bilbao, The Publication Service of the University of the Basque Country.

AUGER P. (1986), Francisation et terminologie: l'aménagement terminologique, In G. RONDEAU & J.C. SAGER (eds.), *Termina 84: terminologie et coopération internationale : la terminologie, outil indispensable au transfert des technologies*. Colloque international de terminologie, Luxembourg, 47-55, Québec, Birsterm.

AUGER, P. (1999), L'implantation des officialismes halieutiques au Québec: essai de terminométrie, Coll. «Langues et sociétés», Québec, Gouvernement du Québec, 221.

BOULANGER J.C. (1991), Une lecture socioculturelle de la terminologie, *Cahiers de linguistique sociale*, Vol. 18, 13-30.

CABRÉ M.T. (1998), Prefacio: Importancia y valided de la teoría de Wúster, In M.T. CABRÉ (ed.) *Introducción a la teoría general de la terminología y a la lexicografía terminológic,* 11-13, Barcelona, Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra.

CABRÉ M.T. (1999), *La terminología. Representación y comunicación*, Barcelona, Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra.

CABRÉ M.T. (2001), Sumario de principios que Configuran la nueva propuesta teórica, In M. T. CABRÉ & J. FELUI (eds.), *La terminología científico-técnica*, 17-26, Barcelona, Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra.

CABRÉ M.T. (2002), Terminología y lenguas minoritarias: necesidades, universalidad y especificidad, In *VIII Conferencia internacional de linguas minoritarias. Políticas Lingüísticas e Educativas na Europa Comunitaria*, 89-102, Santiago de Compostela, Xunta de Galicia. Consellerias de Educación e Ordenación Universitaria. Dirección Xeral de Política Lingüística.

CABRÉ M.T. (2003), Terminología y normalización lingüística, In X. ALBERDI & P. SALABURU (eds.), *Ugarteburu Terminologia Jardunaldiak,* 11-25, Bilbao, The Publication Service of the University of the Basque Country.

CORBEIL J.C. (1988), Les terminologies devant Babel, *Actes du colloque Terminologie et technologies nouvelles*. París.

ELORDUI A., ZABALA I. (2005), Terminological Variation in Basque: Analysis of Texts of Different Degrees of Specialization, *Sky Journal of Lingusitics*, Vol. 18,71-92.

ELORDUI A., ZABALA I. (2009), Euskara Batuaren garapen lexiko-diskurtsiboa: batasunetik aniztasun funtzionalera, In R. ETXEPARE, R.

GÓMEZ, J. A. LAKARRA (eds.), *A Festschrift for Bernard Oyharçabal*, Anuario del Seminario de Filología Vasca Julio de Urquijo (ASJU), XLIII: 1-2, 231-246.
ETXEBARRIA J.R. (2002), Euskal prosa zientifikoaren historiaren inguruan, *Hegats*, Vol.30, 63-75.
GAMBIER Y. (1987), Problèmes terminologiques des pluies acides: pour une socio-terminologie, In *La fertilisation dans les langues romanes*, *Meta*, Vol. 32-3, 314-320.
GAUDIN F. (1993), *Socioterminologie: des problèmes sémantiques aux pratiques institutionnelles*. Rouen, Université de Rouen, Publications de l´Úniversité de Rouen, 182.
GUESPIN L. (1993), Normaliser ou standardiser?, *Le Langage et l'homme*. Vol. XXVIII, 4, De Boeck Université, 213-222.
HAUGEN E. (1983), The Implementation of Corpus Planning: Theory and Practice, In J. COBARRUBIAS & J.A. FISHMAN (eds.), *Progress in Language Planning International perspective,* 269-290, Berlin, Mouton.
IRAZABALBEITIA I. (2002), Zientzia komunikazioa Euskal Herrian: hurbilketa modukoa, *Hegats*, Vol. 30, 47-75.
LOINAZ M. (2007), Estudio de un sistema de medición y diagnóstico de la implantación de la terminología en euskera (proyecto de investigación TEIS), *Espais Terminològics 2007. Neología terminológica: el tractament dels malleus*. Barcelona, Eumo Editorial.
LÓPEZ BASAGUREN A. (2012), The Legal System of a Bilingual Society, In P. SALABURU & X. ALBERDI (eds.), *The Challenge of a Bilingual Society in the Basque Country*, 33-50, Reno, Center for Basque Studies Current Research Series 9.
QUIRION J. (2003), *La mesure de l'implantation terminologique: proposition d'un protocole. Étude terminométrique du domaine des transports au Québec*, coll. Langues et sociétés, Vol. 40, Montréal, Office Québécois de la Langue Française.
TEMMERMAN R. (2000), *Towards New Ways of Terminology Description. The sociocognitive approach*. Amsterdam/Philadelphia, John Benjamins.
WÜSTER E. (1998[1979]), *Introducción a la teoría general de la terminología y a la lexicografía terminológica*. M.T. CABRÉ (ed.), Barcelona, Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra. [translation of Wüster E, (1979) *Einführung in die Allgemeine Terminologielehre und Terminologische Lexikographie*. 2 Vol. Viena].
ZABALA I. (2004), Causation and Semantic Control. Diagnosis of Incorrect Uses in Minorized Languages, In B. OYHARÇABAL (ed.), *Inquiries into the Lexicon-Syntax Relations in Basque,* 255-284, Bilbao, ASJU.
ZABALA I., SAN MARTIN I., LERSUNDI M., AZKUE J.J., MENDIZABAL J.L. (2012), The Elaboration of Human Anatomy Terminology for the Basque Language: the Contribution of Translators, Linguists and Experts, *Terminalia,* Vol. 6, 15-25.
ZABALA I., SAN MARTIN I., LERSUNDI M., ELORDUI A. (2011), Graduate Teaching of Specialized Registers in a Language in the Normalization Process: Towards a Comprehensive and Interdisciplinary Treatment of Academic Basque, In S. MARUENDA-BATALLER & B.

CLAVEL-ARROITA (eds.), *Multiple Voices in Academic and professional Discourse*, 208-218, Cambridge, Cambridge Scholars.
ZABALA I., LERSUNDI M., LETURIA I., MANTEROLA I., SANTANDER G. (in progress), GARATERM: euskararen erregistro akademikoen garapenaren ikerketarako lan-ingurunea, In X. ALBERDI & P. SALABURU (eds.), *Ugarteburu Terminologia Jardunaldiak V: Terminologia naturala eta terminologia planifikatua euskararen normalizazioari begira*. Bilbao, The Publication Service of the University of the Basque Country.

# Des communiqués de presse du Cnes à la presse généraliste. Vers un observatoire de la diffusion des termes

Anne Condamines[1] et Aurélie Picton[2]

## Résumé

Cet article s'intéresse au phénomène de la dé-terminologisation, c'est-à-dire aussi bien au passage qu'à l'usage des termes dans la langue dite générale. La problématique s'est élaborée à partir d'une demande du Cnes (Centre National d'Etudes Spatiales) qui souhaitait voir comment le spatial imprègne la connaissance commune. Nous avons ré-interprété cette demande en mettant en place un observatoire permettant d'examiner, à partir des méthodes de terminologie outillée, les phénomènes de dé-terminologisation. Cet observatoire est composé d'un corpus d'articles de presse générale quotidienne et de communiqués de presse du CNES (disponibles sur internet). A partir des candidats-termes extraits grâce à Termostat, nous avons pu constituer une version « termes » du corpus de presse, constituée des articles comportant au moins 1 candidat-terme. L'article présente les questions de méthodologie de constitution du corpus et d'analyse outillée que pose l'étude de la dé-terminologisation. Il propose aussi quelques résultats qui ont déjà pu être obtenus grâce à l'analyse comparée (statistique, lexicographique et sémantique) dans les corpus de l'observatoire, du fonctionnement de certains termes-candidats (particulièrement ceux construits à partir de *satellite*) et il expose les perspectives d'analyses, nombreuses, qui pourraient être menées grâce à la constitution de cet observatoire, sur les phénomènes de glissement des termes vers la langue générale. A terme, cet observatoire permettra de caractériser systématiquement les éléments permettant de décrire la dé-terminologisation.

## Mots-clés

Constitution de corpus, Dé-terminologisation, Diffusion des termes, Domaine spatial, Observatoire de textes, Terminologie outillée.

[1] CLLE ERSS, UMR 5263, CNRS et Université Toulouse 2 Le Mirail, anne.condamines@univ-tlse2.fr

[2] TIM, Faculté de Traduction et d'Interprétation, Université de Genève, aurelie.picton@unige.ch

## 1. Introduction

Les phénomènes de néologie par diffusion des termes des langues de spécialité vers la langue générale ont été peu étudiés de manière systématique et lorsqu'ils le sont, il s'agit le plus souvent d'un point de vue lexicographique. Les questions qui se posent alors concernent le choix des termes à sélectionner comme entrées des dictionnaires ainsi que la définition à retenir, alors que le dictionnaire ne s'adresse pas à des experts (Josselin-Leray, 2005; L'Homme, Vandaele, 2007; Maniez et al., 2008).

Dans cet article, nous abordons cette problématique d'une manière nouvelle et, à notre connaissance, inédite. En effet, notre réflexion est née d'une demande pratique de l'agence spatiale française, le Cnes (Centre National d'Études Spatiales). Cet organisme souhaitait savoir comment le spatial *imprègne* la langue commune, question, que, à travers notre regard de linguistes, nous avons interprétée en une problématique linguistique : comment les termes du spatial se diffusent-ils dans la langue générale? Souhaitant mettre en œuvre des méthodes de linguistique de corpus et de terminologie textuelle, nous nous sommes, dans un premier temps, interrogées sur la méthode à définir pour procéder à une analyse du passage des termes dans la langue générale et, au-delà, à construire un véritable observatoire de ce phénomène, qui ne parte pas d'une liste prédéfinie de termes à observer, mais qui permette de faire émerger des textes les termes pertinents à observer sur une période donnée. Cet article relate donc la mise en place de cet observatoire ainsi que les premières études que nous avons pu mener, qui nous permettent d'évaluer les possibilités et les limites de cette démarche. La partie 2 de cet article précise le contexte de notre étude, à savoir la demande du Cnes, l'interprétation de la problématique du point de vue de la terminologie et la réflexion sur la méthode à mettre en place. La partie 3 concerne la mise en place de l'observatoire proprement dit (constitution des corpus et préparation des données). La partie 4 relate, quant à elle, les premières observations que nous avons pu mener. La partie 5 nous permet de tirer les premières conclusions sur l'étude en corpus du phénomène de diffusion des termes et sur la pertinence de l'observatoire construit pour l'observer.

## 2. Origine de l'étude et problématisation en terminologie

Dans cette partie, nous situons notre étude, qui allie prise en compte d'une demande externe de l'agence spatiale française et inscription de cette demande dans une problématique de terminologie textuelle.

### 2.1. Point de départ de l'étude : la demande du Centre National d'Études Spatiales

À l'occasion de ses cinquante ans, le Cnes a souhaité réunir différentes productions et points de vue traduisant l'analyse d'autres disciplines

scientifiques de tous ordres sur la place du spatial auprès du grand public depuis ces 50 dernières années. À cette occasion, nous avons été sollicitées avec une demande assez inédite : le service de communication du centre voulait en effet savoir comment le spatial, en tant qu'ensemble de connaissances, réelles ou imaginaires, se diffuse dans le grand public, de façon plus ou moins consciente (Condamines, Picton, 2011). Pour reprendre la demande dans les termes mêmes qui nous ont été proposés, il s'agissait de savoir comment les connaissances sur le spatial « perfusent » le grand public. En tant que linguistes travaillant sur des domaines spécialisés, le défi était double : d'une part tenir compte de ce besoin, d'autre part, et surtout, lui donner une interprétation qui le rapproche de problèmes scientifiques connus et travaillés en linguistique. A partir de la métaphore de la perfusion, nous avons surtout travaillé la dimension « injonction, lente, régulière et prolongée » (TLFi) mais aussi l'aspect « non perçu » : un patient perfusé ne ressent pratiquement rien. À partir du point de vue de la linguistique de corpus, cadre de travail que nous adoptons ici, il s'agit alors d'essayer de travailler cette notion de « flux », encore difficile à saisir à l'aide de corpus. Pour tenter toutefois de prendre en compte la dynamique du discours, la méthode généralement utilisée consiste à construire un corpus ouvert composé de textes de différentes périodes, à les comparer et à les mettre à jour régulièrement pour obtenir ce que Renouf appelle des corpus dynamiques (Renouf, 1994). Mais cette mise à jour n'est possible qu'avec certains types de textes produits à intervalles réguliers (par exemple les dépêches de presse) et demande des outils spécifiques pour pouvoir traiter finement ces données linguistiques dynamiques. Pour ces raisons, nous privilégions, dans notre approche, des corpus comparables en diachronie courte. Compte tenu de tous ces paramètres, la question était pour nous de nous interroger d'une part sur les problématiques théoriques que nous devions convoquer pour éclairer cette question et, d'autre part, sur les corpus et la méthodologie que nous pouvions construire pour l'examiner. Dans la section suivante, nous présentons différents éléments théoriques en jeu dans cette problématique de diffusion.

### 2.2. Un point de vue terminologique

Bien que rarement systématiquement étudié, ce phénomène de *glissement* (pour reprendre l'idée de flux) de la terminologie vers le lexique général a revêtu plusieurs dénominations dont « banalisation lexicale » par Guilbert (1975) puis Galisson (1978), « dé-spécialisation » (considérée comme équivalente à vulgarisation) chez Gouadec (1990), « déterminologisation », « migration », ou même « dilution » (qui ne concerne que l'aspect sémantique) chez Meyer et al., (1999) ou encore « dédomanialisation » (Rastier, Valette, 2009).

Tout d'abord, dans cette étude, notre « entrée » dans les textes se fait par les termes. Il s'agit d'un choix méthodologique et pratique qui combine à la fois le fait que les termes sont des pivots entre langue et connaissances et que la recherche de chaînes de caractères ou de groupes de chaînes de caractères (des

termes) est en adéquation avec les possibilités des outils. Sur la base de la problématique formulée par le Cnes, le point le plus facile à cerner pour nous était que nous allions être confrontées au passage de termes dans la langue générale et donc *a priori* à un phénomène de néologie, de forme ou de sens, dans la langue générale. Étonnamment si ce passage du spécialisé au général est très souvent cité comme un mode possible de néologie (Sablayrolles, 2000), il a été peu étudié de manière systématique. Les travaux existants portent plutôt sur le phénomène inverse, celui qui permet aux terminologies d'emprunter à la langue générale (Guilbert, 1975; Calberg-Challot, 2007); comme mentionné en introduction, une question qui se pose souvent alors est celle de savoir quels termes doivent être consignés dans un dictionnaire de langue générale.

Pour repérer les phénomènes de néologie dans la langue générale, nombreux sont les travaux qui, dans les années 1990, ont proposé de comparer un corpus de langue « générale » (la presse, le plus souvent) et une liste d'exclusion (les entrées d'un ou plusieurs dictionnaires) afin de repérer les formes non recensées dans le dictionnaire et donc candidates à une entrée et une définition dans un dictionnaire de langue générale (Cabré, De Yzaguirre, 1995; Mathieu, 1998). Si ces travaux présentent un intérêt certain, ils ne répondent pas tout à fait à notre problématique. En effet, notre objectif n'est pas lexicographique, puisqu'il ne s'agit pas de construire un dictionnaire des termes du spatial dans la langue générale, mais plutôt d'ordre méthodologique et théorique puisqu'il vise à comprendre comment, par quels mécanismes linguistiques, s'effectue la diffusion des termes du spatial dans le grand public. Il ne peut donc s'appuyer sur un dictionnaire pré-existant (en tout cas pas uniquement).

Deuxièmement, parmi les nombreuses questions que la problématique de ce glissement soulève, nous pouvons en retenir deux ici. La première est qu'il s'agit le plus souvent d'un phénomène non conscient à la fois des locuteurs et des interlocuteurs. Il se différencie alors de la vulgarisation par le fait qu'il ne relève pas d'un processus délibéré de mise à portée de connaissances scientifiques auprès du grand public (Roqueplo, 1974). Mais parallèlement, le fait qu'un terme se diffuse dans la langue générale est souvent lié à un intérêt accru du public/interlocuteur pour la thématique à laquelle est lié ce terme : la médecine, l'informatique, le nucléaire… le spatial. La néologie, de forme comme de sens, n'est donc possible que si elle est collectivement acceptée :

> Une évolution [de la norme linguistique] ne saurait donc s'imposer que dans la mesure où elle résulte de causes collectives, autrement dit de facteurs affectant l'ensemble du groupe linguistique considéré (pas forcément la communauté entière). (Nyckees, 2000 : 37)

Une autre question qui se pose vient de la nécessité de s'interroger sur la nature du terme « intégré » à la langue générale. Ainsi par exemple, Ungureanu se demande à propos du terme diffusé « est-ce toujours un terme, est-ce devenu un mot ? » (Ungureanu, 2003 : 108). Pour le dire autrement, si un terme a intégré le lexique commun, est-il toujours, pour les locuteurs, chargé de son

sens initial? La « conscience » même de son domaine d'origine est-elle toujours présente?

Il faut aussi s'intéresser à l'impact qu'a cette mobilité vers la langue commune sur le sens lui-même. Dans un grand nombre de cas, on peut penser que le grand public ne connaît pas exactement le sens d'un terme, en tout cas si l'on entend par là qu'il en a une définition experte. En reprenant les propositions de Putnam (1990), et à la suite de Gentilhomme (1994) ou Nicolae et Delavigne (2009), on peut penser que le grand public laisse aux experts la responsabilité du sens exact d'un terme (d'où l'idée « d'usage par délégation », voire par « déférence » proposé par Putnam) mais ne s'empêchent pas d'utiliser ce même terme. C'est même le fait de cette autorisation qui signale que le terme ne relève plus du seul domaine spécialisé, de la seule compétence langagière des experts. On trouve une idée assez proche dans le distinguo qu'établit Wierzbicka (1985) entre concept maximal et concept minimal du terme, repris par Dury (1997 : 26) :

> Le spécialiste détiendrait […] le concept maximal d'un terme ou tout du moins serait celui qui accède aux informations conceptuelles les plus nombreuses […] et le non-spécialiste, par contre, serait celui qui accèderait au concept minimal, minimum vital mais suffisant pour communiquer.

De fait les usages se trouvent en partie modifiés et, comme le fait la linguistique de corpus, on peut faire l'hypothèse que les contextes d'apparition du terme ainsi passé dans l'usage « courant » vont pouvoir être utilisés pour repérer les modifications sémantiques. En reprenant la proposition de Meillet (1921), également soutenue par Sablayrolles (2000), selon les cas, le sens va être soit étendu mais appauvri, soit restreint mais enrichi.

Enfin, la diffusion des termes dans la langue générale peut se faire *via* des figures de style, l'usage perdant parfois toute référence domaniale. Le phénomène de métaphore est de ce point de vue particulièrement important. Par exemple, on peut parler aujourd'hui d'une *pléiade d'artistes*, de *vouloir décrocher la lune*, ou encore de *tirer des plans sur la comète*. Notons que ces cas-là intéressent aussi nos interlocuteurs du Cnes, qui les considèrent (à juste raison) comme des indices de l'influence du spatial sur le grand public. Mais, faute d'en avoir rencontrés de manière significative à ce stade de notre étude (et sur l'intervalle temporel que nous ciblons), dans la suite de cet article, nous ne parlerons pas de ce type de phénomène.

Compte tenu de ce contexte de questionnement théorique et de la demande du Cnes, nous proposons donc dans cet article une réflexion sur la définition d'une méthode permettant d'examiner les glissements du domaine de l'espace vers la langue générale, d'une part pour identifier en corpus les termes concernés et d'autre part pour étudier les phénomènes linguistiques qui sont à l'œuvre. Nous expliquons cette démarche méthodologique dans les sections suivantes.

## 2.3. Construction d'une démarche méthodologique

### 2.3.1. Une approche basée sur la linguistique de corpus

Pour répondre à notre problématique de la diffusion des termes vers la langue générale, nous situons notre approche dans la linguistique de corpus et dans la terminologie textuelle (Pearson, 1998; Condamines, Dehaut, 2011). Nous nous appuyons ainsi sur les progrès des méthodes outillées qui peuvent être mises en œuvre sur les corpus spécialisés. Nous voyons ainsi comment, selon l'objectif de l'étude, nous pouvons faire appel aux trois différents indices - de formes, de distribution de ces formes et de quantification de ces formes - que nous fournissent les outils tels que les extracteurs de termes, les analyseurs syntaxiques ou les concordanciers. (Condamines, Dehaut, Picton, 2012). Enfin, nous inscrivons nos travaux dans des études de diachronie courte (Dury, 2006; Humbley, 2011; Tolédano, Candel, 2006), ce qui nous amène à mettre en place des méthodes dédiées et à nous interroger sur la nature et le rôle des corpus à examiner (Picton, 2009).

Concrètement, notre approche se met ici en place sur la base de la comparaison de corpus. Dans le cadre de l'étude de la néologie, encore peu de travaux d'exploration outillée se basent sur la comparaison directe de corpus de textes, censés permettre de repérer les termes qui sont apparus (ou sont en train de migrer) dans la langue générale; notons toutefois les travaux de l'IULA (Cabré, Nazar, 2011) et ceux de l'OLST (Drouin et al., 2006).
Mais dans le cadre de notre problématique de diffusion de termes, la volonté de comparer des corpus (et non pas un dictionnaire avec un corpus spécialisé), se heurte à la question du choix de corpus représentatifs (Biber, 1993) pour observer le passage des termes dans le temps.

### 2.3.2. Choix du type de corpus

Le choix de baser notre approche sur une comparaison de corpus impose une réflexion importante sur le genre des textes à prendre en compte pour des données représentatives de ce passage entre langue de spécialité et langue générale. De ce point de vue, le rôle de mise à disposition de la connaissance experte (et donc de diffusion de la terminologie) est généralement attribué aux medias (Jacobi, 1986; Moirand, 2007). Dans la mesure où les corpus journalistiques sont parmi les plus accessibles et ce, en grande quantité, nous avons considéré que la presse était un genre adapté pour observer les termes utilisés dans la langue générale. Toutefois, nous n'avons pas opté pour la presse ayant une vocation de vulgarisation mais pour la presse générale. En effet, il nous a semblé que l'usage y était moins lié à une volonté de vulgarisation que celui que l'on pouvait trouver, par exemple dans la presse spécialisée (La Recherche, Sciences et Vie, etc.) et qu'il pouvait donc mieux refléter une utilisation banalisée, en tout cas supposé « compréhensible » par un lecteur de presse « moyen ».

Du point de vue de la langue de spécialité, le recours à des textes spécialisés du domaine spatial a été écartée pour deux raisons principales : premièrement, l'hétérogénéité des spécialités présentes au Cnes (optique, mécanique, astronomie, etc.) rend difficile la circonscription d'un corpus exploitable et représentatif du domaine. L'accès aux données et à des spécialistes de ces domaines à l'agence spatiale est également limité et ardu, d'autant plus que nous nous inscrivons dans une perspective diachronique (Picton, 2009; Dury, 2004). Pour cette raison, nous avons choisi plutôt de nous intéresser à un genre de textes « à l'interface » entre spécialistes et grand public : les communiqués de presse du Cnes. En effet, dans une perspective de diffusion des connaissances, au moins une autre communauté de locuteurs peut intervenir entre les experts et les journalistes de presse quotidienne : les chargés de communication des organismes publics et privés. Il est bien évident que ces chargés de communication n'ont pas seulement une intention de diffusion de la connaissance mais aussi de mise en valeur des compétences des sociétés qu'ils représentent. Toutefois, du point de vue de l'usage des termes, les productions de cette communauté discursive nous ont paru intéressantes (hypothèse que l'on retrouve d'ailleurs également chez Nicolae et Delavigne (2009) par exemple). En effet, cet usage et sa diffusion relève pour une grande part d'un processus non-conscient et n'est donc sans doute que peu affecté par une volonté commerciale. Il nous est apparu ainsi que le basculement de l'usage expert vers l'usage « général » pouvait se repérer dans le passage du discours des chargés de communication à celui des journalistes de presse générale qui sont les principaux destinataires de leurs communiqués. Ces derniers sont de plus intégralement disponibles sur le site du Cnes, ce qui nous a permis de les récupérer de manière automatisée, assez aisément.

## 3. Construction de l'observatoire

### 3.1. Constitution des corpus

Dans cette partie, nous expliquons comment, à partir des extraits de la presse quotidienne, nous avons construit les premières bases d'un observatoire de la diffusion des termes en corpus.

#### 3.1.1. Communiqués de presse (CP)

Le premier corpus constitué regroupe l'ensemble des communiqués de presse mis en ligne par le Cnes sur son site[3], du 1er janvier 2003 au 31 décembre 2011, soit 535 communiqués. Ces textes ont été récupérés automatiquement pour chaque mois de cet intervalle et regroupés par année (Tableau 1). Ce corpus nous sert tout à la fois de base de comparaison pour observer la diffusion des termes, et de base à l'extraction des termes pertinents à observer : nous faisons en effet l'hypothèse que les termes présents dans ce corpus « intermédiaire », à

[3] http://www.cnes.fr (consultée le 31.10.2012)

l'intention de la presse, doivent être les cibles privilégiées d'une diffusion dans la langue générale.

| Années | Nombre d'occurrences |
|---|---|
| CP_2003 | 37 274 |
| CP_2004 | 30 240 |
| CP_2005 | 44 645 |
| CP_2006 | 35 232 |
| CP_2007 | 20 720 |
| CP_2008 | 24 692 |
| CP_2009 | 31 672 |
| CP_2010 | 28 397 |
| CP_2011 | 44 875 |
| **TOTAL** | 297 747 |

*Tableau 1 : Corpus de communiqués de presse*

3.1.2. Presse générale (VT et VG)

Le deuxième corpus de l'étude est composé d'articles de presse générale, du 1er janvier 1998 au 31 décembre 2011, regroupés également par année. Ces articles ont été récupérés *via* la plateforme LexisNexis®[4] qui propose des abonnements à l'ensemble de la presse nationale et internationale en version électronique. Pour cette étude, nous n'avons sélectionné que les quotidiens nationaux français disponibles intégralement sur la plateforme du 1er janvier 1998 au 31 décembre 2011, soit Le Monde, Le Figaro, Les Échos et la Croix. Nous avons extrait les 20 premiers articles retournés par mois par la plateforme pour chaque journal, soit 3360 articles. Néanmoins, sur cette base, nous proposons de travailler sur deux versions de ce corpus de presse générale :

- une « version termes » (désormais VT), qui regroupe les articles des quotidiens qui utilisent au moins un terme du spatial (Tableau 2). Cette sélection permet de ne pas regrouper seulement les articles des rubriques spécialisées (par ex. sciences, astronomie, etc.), mais de rassembler tous les articles qui contiennent au moins un terme candidat. En effet, non seulement les usages que nous ciblons peuvent apparaître dans n'importe quelle rubrique (Une, politique, histoire, etc.), mais, après de premières observations, nous nous sommes aperçues que le découpage en rubriques était très instable d'une édition à une autre ou d'un quotidien à l'autre. Cette sélection des articles à partir d'une liste de termes est précisée dans la section suivante.

[4] http://www.lexisnexis.com/ (consultée le 31.10.2012)

| Années | Nombre d'occurrences |
|---|---|
| **1998** | 497 938 |
| **1999** | 516 584 |
| **2000** | 492 032 |
| **2001** | 469 738 |
| **2002** | 446 198 |
| **2003** | 491 712 |
| **2004** | 459 526 |
| **2005** | 482 481 |
| **2006** | 428 297 |
| **2007** | 523 817 |
| **2008** | 456 197 |
| **2009** | 504 827 |
| **2010** | 460 671 |
| **2011** | 430 835 |
| **TOTAL** | 6 660 853 |

*Tableau 2 : Corpus de presse « version termes »*

- une « version globale » (désormais VG), qui regroupe les articles des quotidiens, sélectionnés aléatoirement (Tableau 3).

| Années | Nombre d'occurrences |
|---|---|
| **1998** | 295 905 |
| **1999** | 336 359 |
| **2000** | 292 501 |
| **2001** | 281 455 |
| **2002** | 367 494 |
| **2003** | 264 701 |
| **2004** | 237 167 |
| **2005** | 346 810 |
| **2006** | 351 970 |
| **2007** | 416 091 |
| **2008** | 373 714 |
| **2009** | 386 459 |
| **2010** | 256 280 |
| **2011** | 303 894 |
| **TOTAL** | 4 512 800 |

*Tableau 3 : Corpus de presse « version globale »*

Nous rediscutons plus loin l'impact que le choix de l'une ou l'autre de ces versions peut avoir sur l'analyse (§ 4.1. et § 5.).

### 3.2. Préparation des données

Pour cette étude outillée, l'ensemble des corpus a été lemmatisé et étiqueté avec TreeTagger (Schmidt, 1995).

Comme mentionné plus haut, les communiqués de presse ont été choisis comme genre de textes à l'*interface* entre les langues de spécialité et la langue générale. Suivant ce choix, nous faisons l'hypothèse que les termes du spatial présents dans ces communiqués sont potentiellement diffusés de manière privilégiée dans la presse et donc, dans la langue générale. Pour cette raison, les candidats-termes du domaine présents dans les communiqués sont extraits automatiquement avec l'extracteur TermoStat (Drouin, 2003). À ce stade de notre recherche, nous arrêtons notre sélection aux 100 premiers candidats-termes proposés par TermoStat. Sur ces 100 candidats, 87 ont été validés, par nous-mêmes et après consultation des experts (par ex. *Cnes, satellite, espace, lanceur, agence spatiale, observatoire, orbite, observation, agence spatiale française, mission, fusée, lancement*).

Cette extraction de termes vise donc un double objectif : premièrement, elle nous permet de cibler directement les termes pertinents à observer dans les corpus de presse, évitant ainsi de choisir des unités à observer *a priori,* sur la base de l'intuition seule ; deuxièmement, cette extraction permet d'orienter la constitution des corpus de presse « version termes » (VT), puisque les articles retenus pour ce corpus doivent contenir au moins un terme de la liste extraite avec TermoStat pour être sélectionnés (cf. § 3.1.1.).

## 4. Premières observations

Ces différents éléments méthodologiques offrent une base très riche pour l'étude de la diffusion des termes à travers plusieurs types d'observations croisées. Premièrement, le fait de disposer d'une grande quantité de données supposées représentatives permet d'obtenir des résultats quantitatifs sur la diffusion des termes, éléments encore peu disponibles actuellement en terminologie. Deuxièmement, d'un point de vue qualitatif, des observations linguistiques fines sont également possibles grâce aux outils d'exploration des textes pour l'analyse des différents sous-corpus. Ce regard croisé est garant en parallèle d'une réponse double à la problématique posée dans cette recherche : premièrement, il offre des éléments de réponse à la demande pratique du Cnes (quels termes du spatial sont diffusés et comment le sont-ils auprès du grand public?); deuxièmement, il permet de proposer de nouvelles pistes d'investigation sur des questions théoriques et empiriques en terminologie. Pour montrer cela, nous proposons dans cet article de présenter plusieurs cas d'observation. Afin de guider cette présentation, nous organiserons nos observations autour du terme *satellite* (et de ses composés), qui est le deuxième terme repéré par TermoStat. Après un aperçu global des tendances de diffusion dans nos corpus (§ 4.1.), nous explorerons respectivement quatre cas de figure :

- le suivi de l'évolution de concepts reliés au cours du temps (§ 4.2.)
- le repérage de fonctionnalités qui intéressent le grand public (§ 4.3.)
- la comparaison du mode de construction des termes complexes (§ 4.4.)
- l'observation de changements sémantiques (§ 4.5.).

Ces différentes pistes examinées vont nous permettre d'illustrer la richesse des phénomènes en jeu, tels que l'on peut les saisir à travers l'observatoire que nous proposons. Nous montrons comment peuvent se mettre en place les analyses sur la diffusion des termes en corpus comparables, par des allers-retours entre hypothèses, basées sur des connaissances linguistiques *a priori*, et prise en compte des occurrences des termes et de leurs contextes.

### 4.1. Observations quantitatives globales

Le fait de disposer de corpus comparables *a priori* représentatifs et de grande taille offre un regard rarement adopté, car difficilement saisissable, sur la « quantification » de certains aspects de la diffusion des termes. Notamment, à partir de l'extraction des 100 premiers termes extraits par TermoStat sur le corpus CP, nous pouvons mesurer la proportion de termes présents dans la langue générale dans nos deux corpus. Ces données permettent de mesurer la diffusion dans les corpus, mais également de s'assurer de leur représentativité.

Ainsi, premièrement, dans le corpus VT, 100% des termes extraits sont présents. Si ce résultat pouvait être attendu dans la mesure où la constitution-même du corpus VT repose sur cette liste de termes, elle permet de nous assurer que tous sont bien représentés dans la presse[5]. Ces termes représentent 254 505 occurrences dans ce corpus, soit 3.9% des occurrences totales. Ces premiers chiffres montrent donc une proportion tout à fait remarquable de termes dans ce corpus de langue générale[6].

Dans le corpus VG par contre, seuls 71% des termes extraits sont présents. Mais l'élément le plus significatif réside dans le fait que ces termes ne représentent plus que 9 263 occurrences, soit 0.9% des occurrences totales du corpus. De plus, s'il n'est peut-être pas très surprenant de ne pas y retrouver d'occurrences de termes très spécifiques (par ex. *aéronomie* ou *lidar*), il est beaucoup plus surprenant de constater l'absence de projets spatiaux-phares

---

[5] Il aurait été possible en effet que certains termes fréquents soient surreprésentés et bloquent la présence des autres termes dans les 200 textes sélectionnés par mois.

[6] Notons toutefois que ces chiffres s'appuient sur des occurrences non désambiguïsées. Certaines occurrences d'unités telles que *satellite*, *étoile* ou *atmosphère* ne sont donc pas (ou ne sont plus aujourd'hui) liées au domaine spatial (par ex. « une danseuse étoile »). Néanmoins, nous reviendrons sur le choix de désambiguïser ou non dans la dernière partie, dans la mesure où en diachronie, cette étape peut entraîner un certain biais dans l'analyse des résultats.

(*emergesat* ou *egnos*), ainsi que des termes peu spécifiques et très fréquents dans les CP, tels que *observation spatiale*, *recherche spatiale* et *système spatial*.

Ces premiers chiffres permettent donc de souligner un comportement très différent entre les deux corpus. En effet, si l'on s'attendait à cibler des articles contenant des termes du spatial de manière privilégiée dans VT, nous ne nous attentions pas à une telle absence de termes dans VG. De fait, ces résultats obtenus dans le corpus VG montrent peut-être plus la manière dont l'information peut se perdre dans la masse de données des médias. En effet, VG et VT ayant été constitués à partir de la même base LexisNexis®, il semblerait qu'avec la sélection totalement aléatoire de textes dans la masse d'articles de presse, l'on perde des occurrences pertinentes. À titre d'exemple en effet, alors que le terme *Cnes* apparaît plus de 4070 fois dans VT, celui-ci n'apparaît plus que 32 fois dans VG. Cette perte de données limite donc considérablement les observations à ce stade. Il apparaît alors clairement que d'autres questions doivent être réglées concernant la constitution d'un corpus de presse générale : ce corpus est-il vraiment représentatif de la masse des termes dans la langue générale? Sinon, comment échantillonner les corpus de presse de manière représentative pour une problématique diachronique telle que la nôtre? Ou encore, comment manipuler autant de données avec les outils actuels, tout en permettant une analyse linguistique fine? En attendant de pouvoir offrir des réponses satisfaisantes à ces questions, dans le cadre de cet article nous mettons ce corpus de côté dans la suite des observations et concentrons nos analyses sur la comparaison des corpus CP et VT.

### 4.2. Suivi de l'évolution de concepts liés

À travers la comparaison croisée de corpus diachroniques en langue générale et en langue de spécialité, il est intéressant d'arriver à saisir la manière dont les technologies spatiales se diffusent auprès du grand public. Parmi ces technologies, les satellites artificiels sont des concepts centraux bien connus du public français (notamment à travers la filière SPOT[7]) qui, depuis les années 80, connaissent un grand nombre d'évolutions techniques. Parmi ces évolutions, nos corpus font ressortir la question de la miniaturisation de ces objets. On trouve en effet aujourd'hui des mini-satellites (de 100 à 1000kg), des micro-satellites (de 10 à 100kg), des nano-satellites (de 1 à 10kg) et des pico-satellites (moins de 1kg). À titre d'illustration, en regardant leur traitement dans les ressources dictionnairiques générales (Tableau 4), on remarque très bien la différence de statut de ces termes entre langue générale et langue de spécialité : si ces technologies sont bien définies et répertoriées dans le Grand Dictionnaire Terminologique ou France Terme, elles ne le sont que partiellement dans le Petit Robert ou le Larousse. L'intérêt alors de travailler à partir de corpus est que cela peut permettre de mieux saisir dans quelle mesure la langue générale a été imprégnée de ces termes.

---

[7] http://www.cnes.fr/web/CNES-fr/9183-gp-spot-25-ans-de-carriere-25-millions-d-images.php (consultée le 31.10.2012)

| | Larousse en ligne | PR2012 | GDT | FranceTerme |
|---|---|---|---|---|
| ***satellite*** | oui | oui | oui | oui |
| ***mini-satellite*** | oui | non | oui | oui |
| ***micro-satellite*** | oui | non | oui | oui |
| ***nano-satellite*** | non | non | oui | oui |
| ***pico-satellite*** | non | non | oui | oui |

*Tableau 4 : Traitement des termes de la famille « satellite » dans les ressources terminographiques et lexicographiques (2012)*

*Figure 1 : Évolution de fréquence des nouveaux types de « satellite » - corpus CP*

*Figure 2 : Évolution de fréquence des nouveaux types de « satellite » - corpus VT*

L'observation de ces termes en corpus permet d'offrir une vision beaucoup plus riche de leur attestation en langue générale, et notamment de saisir l'évolution depuis 1998, en parallèle avec leur évolution dans les communiqués

de presse. La Figure 1 présente la répartition de ces termes dans le corpus CP, depuis 2003; la Figure 2 présente la répartition de ces mêmes termes dans le corpus VT depuis 1998.

De ces graphiques émergent plusieurs informations, qui permettent de mesurer, du moins partiellement, le parallèle entre l'actualité du Cnes et la langue générale. Ces données offrent donc un premier angle de réponse à la demande pratique. Parmi les éléments saillants :

- Premièrement, tous ces termes (à l'exception de *pico-satellite*) sont attestés dans nos corpus depuis 1998 (et 1999 pour *nano-satellite*) : toutes ces technologies sont développées depuis les années 80 et ont été diffusées dans la presse avant 2000. Nous ne pouvons donc pas observer dans nos corpus de phénomène de néologie formelle ou totale qui serait liée aux activités du Cnes dans ces exemples;
- Deuxièmement, la répartition des termes varie grandement dans le temps et les courbes présentent plusieurs pics de fréquence, notamment pour *micro-satellite*, en 1998, 2000 et 2004. Si les données ne permettent pas de parler de néologie, elles permettent toutefois de comprendre que ces pics sont dus à des actualités fortes dans le paysage spatial. Ainsi, depuis fin 1998, le Cnes développe sa première filière de micro-satellites, appelée Myriade[8], qui s'accompagne notamment d'appels à projets et de rassemblements scientifiques importants jusqu'en 2000. Ces actualités se répercutent alors immédiatement dans les corpus de presse;
- Enfin et en lien avec le point précédent, ces pics montrent qu'il y a donc bien un lien de diffusion, au moins partiel, entre presse générale et communiqués de presse. En effet, des parallèles importants peuvent être relevés entre les deux graphiques, qui prouvent l'influence – au moins partielle, mais nous y reviendrons - que le Cnes a sur la diffusion de ces termes. Par exemple, les deux graphiques présentent un pic commun en 2004, année du lancement du premier micro-satellite du Cnes, appelé Demeter[9]. Cette information est passée par les communiqués de presse et dans la foulée, dans la presse elle-même.

Ces différents éléments illustrent donc la manière dont des corpus diachroniques comparables permettent de visualiser plusieurs aspects de la diffusion des termes dans la langue générale. Ils montrent notamment le parallèle entre l'actualité spatiale française et la diffusion des termes spatiaux dans la presse, entre les communiqués de presse et la presse, mais également les limites de l'impact des communiqués de presse du Cnes sur la diffusion des termes : s'il est clair que le Cnes est un acteur majeur de l'actualité spatiale en France (comme le prouvent les différents pics de fréquence associables aux

[8] http://smsc.cnes.fr/MYRIADE/Fr/ (consultée le 31.10.2012).
[9] http://smsc.cnes.fr/DEMETER/Fr/ (consultée le 31.10.2012).

projets du Cnes), il est également clair que les informations spatiales, et donc les termes, peuvent être diffusées dans la presse à partir d'autres médias : autres communiqués d'agences scientifiques majeures (CNRS, ESA, etc.), reprises d'autres nouvelles, etc. Il serait donc pertinent de retravailler le corpus CP pour l'ouvrir à d'autres types de textes et améliorer la représentativité de ces données.

### 4.3. Repérer les fonctionnalités qui intéressent le grand public

Toujours en focalisant nos observations sur le terme *satellite* mais avec un point de vue qualitatif, nous avons pu mettre rapidement en évidence que, parmi les contextes les plus fréquents de *satellite*, on note la présence de structures évoquant les fonctionnalités et les applications possibles liées à l'utilisation des satellites. Il s'agit principalement de la structure [GN par *satellite*] comme dans *observation de la terre par satellite*, le GN étant le plus souvent une nominalisation déverbale[10] et la préposition *par* introduisant le *satellite* comme le moyen par lequel est mise en œuvre la fonctionnalité indiquée par le GN. Cet usage fréquent constitue déjà un indice sur le fait que les lecteurs sont supposés être intéressés par les possibilités offertes par l'utilisation des satellites.

L'examen de cette structure, de la nature et de la répartition des GN dans chacun des corpus va nous permettre d'aborder la question des liens entre phénomènes langagiers et intérêt du public dans le phénomène de la néologie. Le tableau 5 rend compte des 6 premiers GN (par nombre d'occurrences) que nous avons trouvés selon les sous-corpus (VT), dont la majorité semble liée aux concepts de navigation et de télécommunications par satellite. Notons que cette structure apparaît 760 fois dans le corpus de presse VT et 69 fois dans le corpus des communiqués de presse (soit, environ deux fois plus dans les communiqués de presse (2,3 pour mille mots) que dans le corpus VT (1,1 pour mille mots)).

| | Le Monde | Les Échos | La Croix | Le Figaro | VT | CP CNES |
|---|---|---|---|---|---|---|
| 1 | *navigation* | *radionavigation* | *navigation* | *navigation* | *navigation 25,7%* | *navigation 16,7%* |
| 2 | *radionavigation* | *navigation* | *positionnement* | *positionnement* | *radio-navigation 20,3%* | *télécommunication 16,7%* |
| 3 | *positionnement* | *télécommunication* | *télévision* | *localisation* | *positionnement 10,5%* | *observation de la terre 7,8%* |
| 4 | *observation* | *positionnement* | *localisation* | *télévision* | *télévision 7%* | *localisation 7,8%* |
| 5 | *télécommunication* | *télévision* | *radionavigation* | *télécommunication* | *télécommunication 6,8%* | *internet 7,8%* |
| 6 | *localisation* | *téléphone* | *guidage* | *téléphone* | *localisation 4,6%* | *positionnement 6%* |

*Tableau 5 : Groupes nominaux (GN) les plus fréquents dans la structure [GN par satellite]*

[10] Bien que ce ne soit pas toujours le cas, par exemple pour *bourse* ou *chaîne*, cas sur lequel nous revenons plus loin.

Ce tableau montre que, dans tous les cas, *navigation* ou *radionavigation* apparaissent comme la fonctionnalité la plus présente dans les corpus. Ce constat peut permettre deux conclusions : d'une part, on remarque que cette fonctionnalité, très présente dans les communiqués est aussi parfaitement repérée dans la presse, et même plus : si l'on regarde la répartition d'un point de vue statistique, on s'aperçoit que 46% des occurrences de [GN *par satellite*] concernent la navigation ou la radionavigation dans la presse quotidienne et 20,1% dans les communiqués de presse. Nous sommes donc dans une sorte d'usage « en miroir » quant au premier rôle accordé à cette fonctionnalité, qui est certainement attribuable au développement du GPS. Cette hypothèse est d'autant plus probable que les GN *localisation* et *positionnement* sont également liés à cette technologie. D'autre part, cette homogénéité pourrait être un argument en faveur d'une interprétation de diffusion : le grand public semble être capable d'utiliser ces termes, et plus généralement, cette structure, en tout cas de les comprendre. Il y aurait donc une sorte de continuité dans le passage de l'usage scientifique à l'usage banalisé, *via* l'utilisation de ces termes par les communiqués de presse.

Si l'on s'intéresse à présent de manière globale aux 53 GN qui peuvent apparaître dans la structure [GN *par satellite*] (corpus CP et VT confondus), on se rend compte que seulement 22 sont utilisés dans les CP. Pourtant, parmi les 31 qui ne sont pas dans ces CP, 6 sont utilisés dans au moins trois des quotidiens de VT dont *bouquet, chaîne, détection, guidage, téléphonie, téléphone*. Si plusieurs hypothèses peuvent être proposées pour expliquer ce décalage, l'observation étendue de ces cas montre que ces termes semblent correspondre à différents degrés d'évolution et d'intégration des applications satellitaires auprès du grand public dans le temps : nous présentons trois cas d'évolution différents à travers les exemples de *télévision par satellite, navigation par satellite* et les termes liés au thème de la médecine par satellite.

4.3.1. Terme diffusé : *télévision par satellite*

Comme nous venons de le voir, les GN les plus fréquents dans les corpus sont tous liés à différents types de télécommunications, la télévision en tête. Cet exemple est particulièrement intéressant dans la mesure où il illustre un exemple d'intégration très clair d'une application du spatial auprès du grand public et d'appropriation du terme par le locuteur non spécialiste depuis quelques années déjà (et bien avant 1998, date de nos premiers corpus de presse). Cet aspect se manifeste linguistiquement par une succession d'ellipses à partir du terme *bouquet de chaines de télévision par satellite*, qui aujourd'hui ne se retrouve repris dans nos corpus majoritairement que par les termes *bouquet de télévision satellite, bouquet satellitaire, ou bouquet satellite*.

4.3.2. Terme en cours de diffusion : *navigation par satellite*

Le deuxième exemple est celui de *navigation* qui est, sur la période 2003-2011, la technologie satellitaire la plus représentée dans les corpus. Cet exemple est particulièrement riche et intéressant pour notre étude puisqu'il s'agit d'un terme/concept en pleine diffusion et évolution. Cette évolution est marquée dans nos corpus d'une part par une augmentation du nombre d'occurrences de ce terme et d'autre part, par une diversification de ses contextes d'apparition. Nous présentons l'analyse complète de ce cas dans la section 4.5. où nous détaillons la manière dont l'observatoire permet de saisir l'évolution sémantique en cours de ce terme.

4.3.3. Termes aux prémices d'une diffusion : le thème de la médecine par satellite

Enfin, inversement, plusieurs termes utilisés dans les CP par le Cnes ne sont quasiment jamais repris dans les quotidiens. En effet, si les concepts de la navigation et de la télévision par satellite sont repris de manière significative, certaines autres fonctionnalités sont aussi présentes dans les CP mais inégalement représentées dans VT. C'est le cas par exemple des termes liés aux applications médicales par satellite. Dans les CP, 4 GN peuvent relever de cette catégorie : *dépistage des complications (1), télé-médecine (2), télé-santé (2), sauvetage (3)*, soit 8 occurrences sur les 69 qui correspondent à la structure [GN *par satellite*] (11,6%). Or, cette catégorie reste encore très peu présente dans VT : seulement 6 occurrences sur 760 GN (0,79% ; *dépistage des complications* (1), *télémédecine* (2), *échographie* (1) et *téléconsultation* (2)). Notons que certains quotidiens, tel que Le Monde, n'utilisent d'ailleurs aucun de ces syntagmes. Il semble donc qu'il y ait un décalage important entre les CP d'une part, et les quotidiens d'autre part, sur la thématique de la médecine par satellite, d'autant plus qu'aucune autre structure syntaxique n'a pu être rencontrée dans nos corpus pour exprimer ces concepts (telles que [le satellite X permet la téléconsultation] ou encore, les [échographies peuvent être réalisées grâce au satellite X] ou, et c'est une structure moins prévisible, qui semble spécifique de ce corpus : [satellite de GN] (avec GN = *télécommunication, localisation, observation de la terre*…)).

## 4.4. Comparer le mode de construction des termes complexes

Une autre piste de recherche, plus générale, est apparue, toujours à partir de l'étude des structures [GN *par satellite*]. Il s'agit du mode de construction des termes complexes, et en particulier de la différence potentielle entre la construction des termes complexes dans les langues de spécialité et dans la langue générale. En effet, en travaillant sur la structure [GN par satellite], nous nous sommes aperçues qu'elle pouvait alterner avec la forme [GN satellite] par exemple:

*observation par satellite/observation satellite*
*localisation par satellite/localisation satellite*
*télécommunication par satellite/télécommunication satellite*

Cette alternance a attiré notre attention et nous avons voulu voir comment, plus généralement, se faisait la répartition des structures [GN prep GN] vs [GN GN] (par exemple : *station sol/station au sol, données satellite/données du satellite*). Le Tableau 6 rend compte de cette répartition. Notons que nous sommes parties des structures [GN GN] pour lesquelles nous avons fait l'hypothèse que la structure concurrente [GN prep GN] était possible.

| | **CP** | **VT** |
|---|---|---|
| ***NN*** | 120 | 171 |
| ***N prep N*** | 122 | 582 |
| **Proportion de *NN*** | 49.6% | 22,7% |

*Tableau 6 : Répartition des variantes selon leurs modes de construction dans VT et CP*

Les résultats obtenus permettent bien de mettre en évidence une tendance : alors que 4 fois sur 5, la langue générale utilise les termes GN prep GN, on ne les sélectionne qu'1 fois sur 2 seulement, dans les textes spécialisés.

L'ellipse du déterminant et de la préposition est souvent considérée comme un indice de terminologisation (Collet, 1997; Portelance, 1991). Dans le même ordre d'idée, pour la théorie des sous-langages, ces formes de « déviances » (Kittredge, Lehrberger, 1983), appelées ainsi car elles ne correspondent pas aux constructions attendues, font partie intégrante de la problématique. Mais aucune étude n'a, à notre connaissance, étudié le processus inverse, de rétablissement du déterminant et de la préposition, dans les cas de dé-terminologisation. Or, il semble bien que, de manière globale, lors du passage des termes dans le discours général, les prépositions soient rétablies.

Le chantier est ouvert car beaucoup de questions restent en suspens : certaines prépositions sont-elles particulièrement rétablies ? Inversement, d'autres ne sont-elles jamais rétablies ? Ne peut-il y avoir d'erreur dans la restitution de la préposition (*communication satellite* est-il l'ellipse de *communication par satellite* ou de *communication du satellite*)? Y'a-t-il des régularités en fonction des quotidiens ? Néanmoins, cet observatoire non seulement permet de mesurer cette tendance (aspect quantitatif) mais il permettra aussi d'approfondir cette problématique d'un point de vue qualitatif.

### 4.5. Observer des changements sémantiques

Pour terminer, nous revenons sur le cas de *navigation satellite*. De manière encore plus fine en effet, il est possible avec cet observatoire de repérer des changements sémantiques dont l'origine repose précisément sur la diffusion des termes dans la langue générale. Ainsi, nous l'avons vue, le terme *navigation*

*satellite* se diffuse largement dans la langue générale, notamment à travers le développement et la banalisation des GPS. Les modifieurs de *navigation* répertoriés à ce jour dans les ressources lexicographiques sont *maritime, aérienne* et *informatique*. Les combinaisons entre *navigation* et ces différents adjectifs méritent toutefois un commentaire. En effet, si on perçoit la parenté sémantique entre *navigation* et *maritime* puis la métaphore, régulièrement attestée[11] entre l'eau et l'espace dans *navigation aérienne,* la création *navigation informatique* demande un effort supplémentaire pour considérer l'internet comme un lieu fluide, dans lequel on peut se déplacer (comme l'eau ou l'espace). Aujourd'hui, 4 contextes apparaissent, qui montrent une évolution sémantique émergente de ce paradigme vers l'ajout de la navigation *terrestre* dont par exemple :

« système de navigation des véhicules terrestres » (VT)
« nous conjuguons les données GPS avec les données de la réflectométrie pour obtenir, notamment en matière de navigation terrestre, une précision jusque-là inégalée et aussi une plus grande fiabilité » (VT)

On peut penser que la création de *navigation terrestre* est le fruit d'un double processus : d'une part, une troncation dans le terme « navigation des véhicules terrestres » et, d'autre part, l'influence du paradigme *maritime, aérien,… terrestre*. Quoi qu'il en soit, la combinaison créée a pour effet d'augmenter la couverture sémantique de *navigation* dont le sens pourrait être, dans le contexte spatial : guidage par satellite de tout objet se déplaçant dans l'air, l'espace ou sur terre.

## 5. Éléments de discussion et remarques conclusives

Dans cet article, nous avons présenté un regard original de linguistique de corpus et de terminologie textuelle sur une question encore rarement systématiquement étudiée : la diffusion des termes dans la langue générale.

À partir d'une demande pratique du Centre national d'études spatiales, nous avons proposé une réflexion méthodologique sur les outils et les données envisageables pour analyser de manière complète et fine les phénomènes en jeu. Si les pistes que nous abordons à ce jour sont encore parfois embryonnaires, elles restent toutefois nombreuses et riches pour répondre à des besoins pratiques mais aussi pour mettre en œuvre des analyses encore inédites à ce jour, souvent faute de moyens. Nous l'avons illustré dans la section 4, de nombreux aspects, tant quantitatifs que qualitatifs peuvent être questionnés et offrent un portrait inédit et ouvert du passage des termes d'une langue de spécialité à la langue générale. Ces différentes raisons nous mènent, malgré l'imperfection de l'environnement que nous présentons, à proposer le terme d'*observatoire*.

[11] Voir par exemple *vaisseau spatial.*

D'un point de vue méthodologique, le défi que nous nous posons reste de taille et beaucoup de paramètres sont à affiner. Parmi les principaux soulevés dans cette recherche, nous en retiendrons trois à ce stade. Le premier concerne la pertinence du genre « presse » pour refléter les usages en langue générale. Comme nous l'avons dit, si sa représentativité a parfois été remise en cause, la presse n'en reste pas moins un genre souvent utilisé, disponible et en quantité, trois qualités encore rares dans d'autres genres pour analyser la langue générale. Mais au-delà de cet aspect, dans notre cas, ce sont surtout les masses de données qui nous posent question : les données disponibles sont-elles trop abondantes? Limitent-elles la représentativité des données pour observer la diffusion des termes? Ou alors, comment échantillonner, en gardant l'équilibre entre quantité et représentativité par rapport à notre problématique de diffusion? Ces questions restent encore pour nous difficiles à trancher, mais doivent être prises en compte de manière centrale dans la suite de nos travaux.

Le deuxième est que s'il est clair que la présence des termes que nous avons observés est influencée par les activités du Cnes diffusées, elle peut l'être également par d'autres médias tels que des communiqués d'agences scientifiques majeures autres comme le CNRS ou l'ESA (European Space Agency), ou encore, obtenues par d'autres sources, traduites, etc. Il faudra donc veiller à retravailler le corpus CP pour le compléter par d'autres sources et améliorer ainsi sa représentativité.

Le troisième problème de taille dans notre environnement est un problème encore une fois relativement classique dans les approches sur (grands) corpus : celui de la désambigüisation. En effet, nos corpus ne sont pas désambigüisés et il est clair que nombre des termes que nous observons sont polysémiques (*espace, satellite, atmosphère, étoile*, etc.). De fait, nombre de contextes que nous obtenons ne sont pas pertinents. Cependant, notre perspective diachronique implique une nouvelle contrainte sur cette question : puisque le sens est susceptible d'évoluer dans le temps, dans quelle mesure la désambigüisation systématique serait-elle pertinente et efficace? Nous l'avons vu, la diffusion des termes entrainent des changements de sens (et notamment des métaphores) que nous cherchons précisément à observer et que nous ne pouvons pas toujours anticiper. De fait, plusieurs tests devront également être mis en place pour en mesurer l'impact.

De manière générale, si nous croyons que notre observatoire pose des pistes originales et prometteuses pour l'analyse de la diffusion des termes, les ressources et les méthodes que nous impliquons doivent encore être questionnées et ajustées pour offrir une analyse mieux contrôlée. Cependant, nous restons convaincues que cet observatoire met à disposition un cadre méthodologique global et reproductible qui permet d'articuler analyses globales (quantitatives) et locales de tous ordres : linguistiques, lexicométriques, statistiques, etc.

Dans cet article, nous avons d'ores et déjà ouvert plusieurs pistes de travaux à explorer. D'un point de vue quantitatif, nous avons montré que la comparaison CP/VT montre tantôt une harmonisation, tantôt un décalage. Il est nécessaire alors d'interroger ces résultats en fonction d'événements qui ont pu avoir lieu dans le domaine de l'espace et de leur impact sur le grand public. Cet aspect du rôle de l'extra-linguistique devra (et pourra) être interrogé plus systématiquement dans son rôle sur les phénomènes de néologie. D'un point de vue qualitatif, plusieurs axes de recherche vont être approfondis. Tout d'abord, d'un point de vue global, le rétablissement de la préposition et du déterminant dans le passage des CP au corpus de presse, semble un indice assez net de déterminologisation, qui serait le pendant de la suppression de ces éléments dans le passage du général au spécifique. Mais, nous l'avons vu, cette hypothèse pose beaucoup de questions qui pourront être examinées grâce à l'observatoire que nous avons construit. Un autre phénomène qui a été mis au jour concerne la conjonction d'éléments d'explication pour justifier l'apparition d'un terme comme *navigation terrestre*. Nous pensons que nous pourrons trouver d'autres néologismes de ce type, ce qui contribuera à établir une typologie de la néologie liée au passage de la langue spécialisée à la langue générale. Enfin, sur un autre point (mais le propre de ce type de recherche exploratoire est de faire émerger des questions sur des plans que nous n'avions pas imaginés dans un premier temps), il est apparu une structure que l'on pourrait certainement considérer comme un marqueur de fonctionnalité : [GN1 par GN2] et son pendant [GN2 de GN1]. Dans la perspective de systématiser la recherche d'information de ce type, il sera nécessaire d'évaluer la pertinence de ce marqueur, dans les corpus du spatial mais aussi dans des corpus d'autres domaines.

D'un point de vue théorique et épistémologique, nous pensons que ce cadre est donc garant d'une observation non seulement plus systématique, reproductible et fiable, mais également inductive. En d'autres termes, cet observatoire permet de mettre au jour des phénomènes de diffusion pertinents et centraux à observer mais non connus ou soupçonnés *a priori*. Ceci permettra, nous l'espérons, d'aller vers une meilleure compréhension de ces phénomènes et surtout vers la mise au jour d'une typologie plus exhaustive et fine du passage des termes d'une langue de spécialité à la langue générale.

Et en effet, si cette question de la diffusion est parfois abordée aujourd'hui par les linguistes et terminologues, elle reste encore mal connue comme en témoigne l'hétérogénéité des dénominations qui la concernent. À l'aide de meilleurs outils, il sera alors possible de mieux saisir les liens entre *migration*, *diffusion*, *vulgarisation*, *déterminologisation* ou encore *néologie*, tous phénomènes à la frontière si difficile à saisir entre langue de spécialité et langue générale.

## Remerciements

Cette étude a été menée dans le cadre d'une convention avec le Centre National d'Etudes Spatiales que nous remercions pour la confiance dont il nous a témoigné.

## Références

ALEXANDRU C., GAUDIN, F. (2005), Les contextes à la source du terme, In *Mots, termes et contextes*, 59-67, Bruxelles, EAC-AUF. (http://perso.univ-lyon2.fr/~thoiron/JS%20LTT%202005/pdf/Gaudin.pdf, consultée le 31.10.2012)

BIBER D. (1993), Representativeness in Corpus Design*, Literary and Linguist Computing,* Vol. 8(4), 243-257.

CABRÉ M.T., DE YZAGUIRRE L. (1995), Stratégie pour la détection semi-automatique des néologismes de presse, *Traduction Terminologie Rédaction*, Vol. VIII(2), 89-100.

CABRÉ M.T., NAZAR R. (2011), Towards a New Approach to the Study of Neology, *Neologica*, Vol. 6, 63-80. (http://www.termisti.org/semin2011/cabrenazar.pdf, consultée le 31.10.2012)

CALBERG-CHALLOT M. (2007), Quand une langue de spécialité emprunte au langage courant : le nucléaire, étude de cas, In HUMBLEY J. (dir.), *Aspects de la recherche en langues de spécialité*, *Cahier du CIEL 2007-2008*, 71-85, Paris, Université Paris VII, UFR EILA.

COLLET T. (1997), La réduction des unités terminologiques complexes de type syntagmatique, *Meta* : *journal des traducteurs*, Vol. 42(1), 193-206.

CONDAMINES A., DEHAUT N. (2011), Mise en œuvre des méthodes de la linguistique de corpus pour étudier les termes en situation d'innovation disciplinaire : le cas de l'exobiologie, *Meta* : *journal des traducteurs,* Vol. 56(2), 266-283.

CONDAMINES A., PICTON A., (2011), Les mots du spatial chez le grand public, une histoire entre imaginaire et réalité, In AZOULAY G., PESTRE D. (dirs.) *Polyphonies spatiales*, 190-192, Paris, Gallimard.

CONDAMINES A., DEHAUT N. PICTON A. (2012), Rôle du temps et de la pluridisciplinarité dans la néologie sémantique en contexte scientifique. Études outillées en corpus, In GÉRARD C., KATABEK J. (dirs.) *Néologie sémantique et analyse de corpus, Cahiers de Lexicologie*, Vol. 101, 161-184, Paris, Garnier Classiques.

DROUIN P. (2003), Term Extraction Using non-Technical Corpora as a Point of Leverage, *Terminology*, Vol. 9(1), 99-117.

DROUIN P., PAQUIN A., MÉNARD N. (2006), Extraction semi-automatique des néologismes dans la terminologie du terrorisme, In *Actes des 8e Journées internationales d'Analyse statistique des Données Textuelles, JADT 2006*, en ligne http://lexicometrica.univ-paris3.fr/jadt/jadt2006/PDF/033.pdf (consultée le 31.10.2012)

DURY P. (1997), *Étude comparative et diachronique de l'évolution*, Thèse de doctorat en Linguistique – Lexicologie et Terminologie multilingues – Traduction, Université Louis-Lumière, Lyon 2.

DURY P. (1999) Les variations sémantiques en terminologie : étude diachronique et comparative appliquée à l'écologie, In DELAVIGNE V., BOUVERET M. (dirs.) *Sémantique des termes spécialisés*, 17-33, Rouen Publications de l'Université de Rouen.

DURY P. (2004), Building a Bilingual Diachronic Corpus of Ecology: The Long Road to Completion, *Icame Journal,* Vol. 28, 5-16.

DURY P. (2006), La dimension diachronique en terminologie et en traduction spécialisée : le cas de l'écologie, In GAUDIN F., CANDEL D. (dirs), *Aspects diachroniques du vocabulaire*, 109-124, Rouen, Publications des Universités de Rouen et du Havre.

GALISSON R. (1978), *Recherches de lexicologie descriptive, la banalisation lexicale.* Paris, Nathan.

GENTILHOMME-KOTYRINE Y. (1994), Regards sur la terminologisation en lexicologie, *Meta* : *journal des traducteurs,* Vol. 39(4), 546-560.

GOUADEC D. (1990), *Terminologie : Constitution des données*. Paris, AFNOR.

GUILBERT L. (1975), *La créativité lexicale*, Paris, Larousse.

HUMBLEY J. (2011), Vers une méthode de terminologie rétrospective, *Langages,* Vol. 183(3), 51-56.

JACOBI D. (1986), *Diffusion et vulgarisation, itinéraires du texte scientifique*, Besançon, Presses Universitaires de Franche-Comté.

JOSSELIN-LERAY A. (2005), *Place et rôle des terminologies dans les dictionnaires généraux unilingues et bilingues. Étude d'un domaine de spécialité : volcanologie*. Thèse de doctorat en Linguistique – Lexicologie et Terminologie multilingues – Traduction, Université Louis-Lumière, Lyon 2.

KITTREDGE R., LEHRBERGER J. (dirs.) (1983), *Sublanguage: Studies of Language in Restricted Semantic Domains*, Berlin/New York, W. de Gruyter

L'HOMME M.-C., VANDAELE S. (dirs.) (2007), *Lexicographie et terminologie. Compatibilité des modèles et des méthodes*, Ottawa, Presses de l'Université d'Ottawa.

MANIEZ F., DURY P., ARLIN N., ROUGEMONT, C. (dirs.) (2008), *Corpus et dictionnaires de langues de spécialité,* Travaux du CRTT, Grenoble, Presses Universitaires de Grenoble.

MATHIEU Y. (1998), Étude pour une extraction automatique de néologismes, In CLAS A., MEJRI S., BACCOUCHE T. (dirs.), *La Mémoire des mots*, Actes V$^{e}$ Journées Scientifiques du réseau Lexicologie, Terminologie, Traduction, Tunis, 25-27 septembre, 455-461, Montréal, Aupelf.

MEILLET A. (1921), *Comment les mots changent de sens,* Année sociologique 1904-1905, édition de 1921, en ligne, (http://fr.wikisource.org/wiki/Comment_les_mots_changent_de_sens (consultée le 31.10.2012).

MEYER I., MACINTOSH K. (1999), 'L'étirement' du sens terminologique : aperçu du phénomène de la déterminologisation, In BÉJOINT H., THOIRON P. (dirs.), *Le sens en terminologie*, 198-217, Lyon, Presses Universitaires de Lyon.

MOIRAND S. (2007), *Les discours de la presse quotidienne. Observer, analyser, comprendre*, Paris, Presses universitaires de France, Linguistique nouvelle.

NICOLAE C., DELAVIGNE V. (2009), Naissance et circulation d'un terme : Une histoire d'exoplanètes, *Textes et corpus*, Vol. 4, Actes de sixièmes journées de Linguistique de Corpus, Lorient, septembre 2009, 143-155, en ligne http://www.licorn-ubs.com/jlc6/ACTES/Nicolae_JLC09.pdf (consultée le 31.10.2012).

NYCKEES V. (2000), Changement de sens et déterminisme socio-culturel, In FRANÇOIS J. (dir.), *Théories contemporaines du changement sémantique*, 31-58, Leuven, Peeters.

PEARSON J. (1998), *Terms in Context*, Amsterdam/Philadelphie, John Benjamins

PICTON A. (2009), *Diachronie en langue de spécialité. Définition d'une méthode linguistique outillée pour repérer l'évolution des connaissances en corpus. Un exemple appliqué au domaine spatial.* Thèse de doctorat en Sciences du Langage, Université Toulouse 2.

PORTELANCE C. (1991), Fondements linguistiques de la terminologie, *Meta : journal des traducteurs*, Vol. 36(1), 64-70.

PUTNAM H. (1990), *Représentation et réalité*, Paris, Éditions Gallimard.

RASTIER F., VALETTE M. (2009), De la polysémie à la néosémie, *Le français moderne*, MEJRI S. (dir.), *La problématique du mot*, Vol. 77, 97-116.

RENOUF A. (1994), Corpora and Historical Dictionaries, In LANCASHIRE I., RUSSON WOOLDRIDGE T. (dirs.), *Early Dictionary Databases*, Center for Computing in the Humanities, 219-235, Toronto, University of Toronto.

ROQUEPLO P. (1974), *Le partage du savoir, science, culture, vulgarisation*, Paris, Le seuil.

SABLAYROLLES J.-F. (2000), *La néologie en français contemporain*, Paris, Honoré Champion.

TOLEDANO V., CANDEL D. (2006), Mouvance terminologique, l'espace d'une courte diachronie, dans les dictionnaires et dans la presse : le cas de "mondialisation-globalisation", In CANDEL D., GAUDIN F. (dirs.), *Aspects diachroniques du vocabulaire*, 157-183, Rouen, Presses Universitaires de Rouen et du Havre.

UNGURANEU L. (2003), *L'interpénétration langue générale-langue spécialisée dans le discours d'internet*, Thèse de doctorat, Université Paris 13/Université Technique de Moldova.

SCHMID H. (1995), Improvements in Part-of-Speech Tagging with an Application to German, In Actes de l'atelier *ACL SIGDAT-Workshop*, Dublin, Irlande, en ligne, ftp://ftp.ims.uni-stuttgart.de/pub/corpora/tree-tagger2.pdf (consultée le 31.10.2012)

WIERZBICKA A. (1985), *Lexicography and conceptual analysis,* Ann Arbor: Karoma.

# Implantation of English terms including the -ING morpheme in French, Spanish and Italian: A corpus-based study of the debates of the European Parliament

François Maniez[1]

## Summary

We briefly describe the means through which the French Ministry of Culture attempts to encourage the use of French equivalents to some English terms ending in *-ing*, and we then examine data extracted from the Europarl corpus (Tiedemann 2009) in order to ascertain the extent to which translators either use translation equivalents or simply borrow the English term in French, Spanish and Italian. We also try to determine to what extent the length of the expression ending in *-ing* has an impact on the degree of implantation of the English borrowing in those languages. The results show that in Romance languages, there is a tendency to borrow single word units such as *benchmarking*, although the implantation of such borrowings may be short-lived when a suitable translation equivalent is suggested. Our conclusion is that EU simultaneous translators contribute to the spread of English *V-ing* terms or expressions in their target language, and that Spanish and French tend to borrow fewer words containing the *-ing* morpheme than Italian.

## Key-words

Borrowing, Europarl corpus, ING morpheme, neology, Romance languages

## 1. Introduction

Many new terms created by borrowing English words in Romance languages contain a word beginning with a verb base and ending with the *-ing* morpheme. Such loanwords very often stimulate the neological process, as new words using the lexical stock of the receiving language are coined. However, national institutions in countries such as France have tried to stem the tide of such borrowings by suggesting alternative words through commissions whose work is funded by the French Ministry of Culture (*Commission Générale de Terminologie et de Néologie,* henceforth CGTN). Widespread use of the coinages that result from the commissions' work may follow, in which case dictionaries will eventually reflect their adoption by the general public some time after their implantation

[1] Centre de Recherche en Terminologie et Traduction, Université Lumière Lyon 2, francois.maniez@univ-lyon2.fr.

has started. For instance, Humbley (2008) reports that only *brainstorming* was included in the 1986 version of the *Petit Robert* dictionary, while both *brainstorming* and *remue-méninges* (the term recommended by the CGTN) were included in the 2007 version of the same dictionary with cross-references to each other. Entries for the new words coined by the commission typically give a definition as well as the foreign word they are supposed to replace. Here is the entry for the word *remue-méninges* taken from the *Vocabulaire de l'économie et des finances* that was published in 2006 by the CGTN:

> **remue-méninges,** n.m.
> ♦ ***Définition :*** Technique de groupe destinée à stimuler l'imagination des participants en vue de leur faire produire le maximum d'idées dans le minimum de temps. ♦ ***Équivalent étranger :*** brainstorming.
> Source : *Journal officiel* du 22 septembre 2000.

Many real (*listing*, *shopping*) and spurious English borrowings (*footing, surbooking*) coexist with the officially recommended terms, whose implantation has met with varying degrees of success. For instance, the French equivalents suggested for *opinion shopping* (chalandage de firmes d'audit) and *treaty shopping* (chalandage fiscal) in the Journal Officiel of 22/09/2000 respectively returned less than 100 and less than 3,000 hits when queried on the Google search engine more than twelve years after they were suggested by the CGTN.

The frequent use of the *-ing* morpheme in Romance language neology has been attributed to its nominalizing function (Picone 1996). It should however be noted that many uses of such borrowings in French are singular or truncated forms that are not considered as acceptable in standard English (*un jogging* for *a jogging suit*, *un parking* for *a parking lot, un passing* for *a passing shot*). The use of the *-ing* morpheme added to a Romance language word base may give rise to neology of form and meaning: in Spanish, the *-ing* morpheme has occasionally been added to a Spanish noun base (*puenting* = *bungy-jumping* (*salto elástico*), in which the root for the noun *puente* (bridge) is used) while some English words have taken on new meanings in French (dated Fr. *footing* for *jogging*). Some French words also use verb bases that exist in English, as is the case for *forcing* and *pressing*, two terms from the field of sport which have now passed into general language and are very often used in conjunction with the verb *faire* (*faire le forcing/pressing*). Here is the etymology and history section devoted to the word *forcing* in the online edition of the *Trésor de la langue française* (henceforth TLF):

> **forcing :** [**Étymol. et Hist.** 1916 boxe; 1953 au fig. « effort intensif ». Empr. à l'angl. *forcing,* subst. verbal du verbe *to force* se rattachant, par l'a. fr., au fr. *forcer**, employé dans le domaine des [sports] pour désigner une attitude d'attaque intensive.

The use of **forcing** in that sense does not exist in English, and the expressions *exert pressure* or *put pressure on* may be used to convey the same meaning, as

witnessed from the following excerpts from the Europarl corpus (the corpus is available on the OPUS web site and its makeup is described in Tiedemann 2009).

| | |
|---|---|
| Must we constantly be forced to **exert such pressure** to guarantee the entry into force of a treaty and its institutional reforms? | Sommes-nous condamnés à constamment devoir **faire le forcing** pour garantir la mise en œuvre d'un traité et de ses réformes institutionnelles ? |
| Liability must include GMOs, particularly at a time when the United States is **putting pressure on** to try to get the moratorium on imports lifted in the context of the WTO negotiations. | La responsabilité doit inclure les OGM surtout au moment où les États-Unis **font le forcing** pour lever le moratoire sur les importations dans le cadre des négociations de l'OMC. |

*Table 1: Two examples of English translations for faire le forcing in the Europarl corpus*

The case of *pressing* is a bit different, since the word is used with a meaning that is somehow related to French usage in the English basketball term *full court pressing defense*, a defensive strategy in which one player harasses the ball handler in the backcourt while the rest of the team maintains a zone defense. The expression *pressing defense* can be traced back to the 1940s on the Google Books search engine, and was soon abbreviated to *full court press*, which can be traced back to the 1950s. The following excerpt from the TLF definition of *pressing* mentions two possible origins for the word, which could have been formed from the verb *press* or directly borrowed from English.

> **pressing : B.** *Fam., SPORTS*. Pression persistante exercée par l'adversaire. […] *faire attaquer le porteur du ballon et* [...] *utiliser le « pressing » et l'agressivité au service de cette récupération du ballon* (J. MERCIER, *Footb.*, 1966, p.64). [...] 2 prob. formé sur l'angl. *to press* « exercer une pression sur, appuyer sur, presser » d'où « attaquer, assaillir, ne pas laisser de répit à » ou directement emp. à *pressing* dont l'empl. dans le domaine du sport n'est cependant att. qu'en 1976.

What the TLF editors mean here is that the English word *pressing* was not used as a noun (as it is in French) until 1976, since its use as an adjective in the expression *pressing defense* predates its use as a noun by about thirty years, so that the use of the noun *pressing* in French might be the result of a simple truncation of the type mentioned earlier for such words as *jogging*, *parking* or *passing*.

## 2. Objectives and method

We tried to determine to what extent simultaneous interpreters in such institutions as the European Parliament contribute to the implantation of

English *V-ing* terms or expressions in target languages through their literal or near-literal translation in Romance languages. Our aim also was to ascertain whether some Romance languages show a higher degree of "resistance to invaders" and to what extent the length of the expression has an impact on the degree of implantation. For that purpose, we used alignments available in the multilingual corpus Europarl (Tiedemann 2009), which includes all the reports of the debates of the European Parliament between 1996 and 2003 in 11 European languages. We queried the corpus for French, Spanish and Italian words and phrases containing a word beginning with a verb/noun base and ending with the *-ing* morpheme (e.g. *benchmarking*, *gender mainstreaming*). We found that in French, Italian and Spanish, there is a tendency to borrow single word units (*benchmarking*), although implantation of such borrowings may not last more than a few years.

## 3. Study of a few -ING borrowings in French, Italian and Spanish

### 3.1. *Benchmarking*

Since the borrowing *benchmarking* was one of the most frequently used in all three Romance languages under study, we analyzed its use over three parliamentary terms (1999-2004, 2004-2009 and 2009-2014) in the French, Italian and Spanish versions of the minutes of the debates of the European parliament. For technical reasons, it was only possible to ascertain the number of debates in which the term was used and not the actual number of times it was used in total for each term. However, the second column of Table 2 gives the actual number of tokens of the word as recorded in the Europarl corpus, which covers another period (1996-2003) that overlaps with the fist parliamentary term we studied.

| Language | 1996-2003 (nb of hits) | 1999-2004 (nb of debates) | 2004-2009 (nb of debates) | 2009-2014 (nb of debates) |
|---|---|---|---|---|
| **English** | 279 | 92 | 63 | 23 |
| **Spanish** | 140 | 46 | 2 | 1 |
| **French** | 170 | 48 | 11 | 2 |
| **Italian** | 228 | 82 | 18 | 1 |

*Table 2: Compared use of* benchmarking *over three parliamentary terms in French, Italian and Spanish*

*Benchmarking* happens to be the foreign equivalent for a recommended French term (*référenciation*) that was included in a list of 100 terms published by the CGTN in 2004 (*Cent termes français du vocabulaire technique recommandés par la CGTN*); two other synonyms were also provided for that term (*étalonnage* and *parangonnage*). Table 3 shows frequencies for the use of *benchmarking* and its recommended French translation equivalents in the last two terms. The results

for the parliamentary term that started in 2009 cannot be deemed statistically significant, but those that were obtained for the 2004-2009 period clearly show that one of the suggested synonyms (*étalonnage*) was widely used after the publication of the CGTN's list, while the term which heads the entry (*référenciation*) was used only once. As for the word *parangonnage*, its implantation was perhaps doomed from the start, since it is formed after a word which is itself quite rare, the French noun *parangon*, which migrated into English circa 1548 as *paragon*. A query for *parangonnage* on the Google search engine gave less than 8,000 hits in January 2013, and most of its occurrences seemed to be instances of use in its original sense in the field of printing.

| | benchmarking (EN) | référenciation | étalonnage | parangonnage |
|---|---|---|---|---|
| **2004-2009** | 63 | 1 | 22 | 0 |
| **2009-2014** | 23 | 0 | 2 | 0 |

*Table 3: Compared use of* benchmarking *and its recommended French translation equivalents in the last two terms*

The following excerpts from the Europarl corpus show that Spanish and Italian occasionally use one-word translation equivalents which are cognates of the English words *reference* and *parameters*:

- Perhaps that is what this was actually all about, hence the rejection of the **benchmarking** method.
- Peut-être que tel était le but réel, d'où le rejet de la méthode de **l'étalonnage**.
- Tal vez era eso de lo que se trataba y de ahí el rechazo del método de la **referencia**.
- Forse è proprio questo il punto, e si spiega così il rifiuto del metodo dei **parametri**.

Another excerpt from the same corpus shows use of translation equivalents which are longer than the source language term, a phenomenon that is frequently observed when the base of the *-ing* form is a compound:

- The problem can be partially solved by **benchmarking**.
- Le problème peut être résolu en partie grâce à **l'étalonnage**.
- El problema puede ser parcialmente resuelto por medio del **uso de puntos de referencia**.
- Il problema può essere parzialmente risolto adottando **parametri di riferimento**.

In many cases, the translation equivalents used for *benchmarking* in Spanish (*evaluaciones comparativas*, *análisis comparativos, estudios comparativos, evaluación mediante parámetros de referencia, sistema de puntos de referencia*) and Italian (*parametri comparativi*, *metodi comparativi, parametri di riferimento, applicazione di parametri, definizione di criteri*) come in the form of a complex noun phrase in which one or more prepositional phrases are embedded.

### 3.2. *Gender mainstreaming* and *gender budgeting*

The case of the expression *gender mainstreaming* is slightly different from that of *benchmarking* since the official recommendation for its French equivalent (*paritarisme*) was made public after most of the debates of our corpus had occurred, as part of the *Vocabulaire des sciences humaines* that was published in the *Journal officiel* of March 4, 2006:

> **paritarisme,** n.m.
> *Domaine :* Sciences humaines/Sciences sociales.
> *Définition :* Action en faveur de l'égalité entre les femmes et les hommes.
> *Équivalent étranger :* gender mainstreaming.

Evidence from more recent parliamentary debates shows that the term *paritarisme* seems to have been adopted, and the fall in the number of occurrences of *gender mainstreaming* in the French version of the debates (Table 4) is probably correlated with the increase in the use of the recommended French equivalent.

| Language | 1996-2003 (nb of hits) | 1999-2004 (nb of debates) | 2004-2009 (nb of debates) | 2009-2014 (nb of debates) |
|---|---|---|---|---|
| **English** | 526 | 68 | 66 | 27 |
| **Spanish** | 36 | 19 | 3 | 1 |
| **French** | 52 | 28 | 15 | 2 |
| **Italian** | 69 | 36 | 23 | 4 |

*Table 4: Compared use of* gender mainstreaming *over three parliamentary terms in French, Italian and Spanish*

The drop in the number of uses of the English expression seems to have occurred faster in Spanish (3 occurrences in 2004-2009) than in Italian (23 for the same period), a fact which is consistent with the higher rate of borrowing from English that is usually witnessed in Italian. In the Europarl corpus, the translation equivalents chosen for *gender mainstreaming* are usually quite long, a fact that bears witness to the difficulty that was originally posed by the source language term:

- Yet, closer inspection reveals that the practical implementation of **gender mainstreaming** in the EU's external policies is still weak.
- En y regardant de plus près, on remarque toutefois que l'application concrète du **principe de prise en considération systématique des questions d'égalité entre hommes et femmes** dans les politiques externes de l'UE reste faible.
- Aún así, una inspección más de cerca revela que la aplicación práctica de **la integración de la perspectiva de género** a las políticas exteriores de la UE todavía es poco importante.
- Eppure un attento esame mostra che l'attuazione pratica dell'**integrazione della dimensione di genere** nelle politiche esterne dell'Unione è ancora scarsa.

In the Europarl corpus, the high degree of variation in translation equivalents bears witness to the difficulty of selecting one single expression: *igualdad de hombres y mujeres*, *trato igualitario de géneros*, *integración de las perspectivas de género* and even the lengthy *medidas destinadas a integrar la igualdad de hombres y mujeres en todas las políticas comunitarias* are found. Another expression using the word *gender* as a modifier has created difficulties for translators: the term *gender budgeting*.

| **Language** | **1996-2003 (nb of hits)** | **1999-2004 (nb of debates)** | **2004-2009 (nb of debates)** | **2009-2014 (nb of debates)** |
|---|---|---|---|---|
| **English** | 35 | 9 | 18 | 6 |
| **Spanish** | 14 | 6 | 1 | 1 |
| **French** | 20 | 6 | 4 | 0 |
| **Italian** | 26 | 6 | 11 | 2 |

*Table 5: Compared use of* gender budgeting *over three parliamentary termsin French, Italian and Spanish*

The official recommendation for a French translation equivalent was published in the *Journal officiel* of March 29, 2003 and suggested the use of *établissement des budgets publics selon la perspective de genre*. This particularly elaborate expression seems not to have stuck in usage, since there were only 33 hits for the expression for a Google search conducted in January 2013. Table 5 shows differing trends for the 2004-2009 term, as the use of the English borrowing nearly disappeared in Spanish, slightly decreased in French in spite of a higher figure for English, and almost doubled in Italian, thus confirming the tendency of the latter language for such borrowings that was seen in Table 4.

The following excerpts from the Europarl corpus show that French and Spanish usually employ translation equivalents for *gender budgeting* that roughly translate as "integration of a gender perspective in budgeting", whereas Italian

either uses the original English expression or a shorter translation equivalent, *bilancio di genere*:

- A useful way of doing this is through **gender budgeting** […]
- Il peut s'avérer utile de se mettre à l'œuvre via **l'intégration de la dimension du genre dans les budgets nationaux** […]
- Una forma útil de hacerlo es mediante **la integración de la perspectiva de género en el presupuesto** […]
- Un'utile metodologia è quella del **gender budgeting** […]

- Austria, incidentally, leads the way in this and will have a law in place as from 2009 – **gender budgeting** – […]
- L'Autriche, soit dit en passant, ouvre la voie en la matière et disposera d'une loi dès 2009 sur **l'intégration de la dimension de genre dans le processus budgétaire**, […]
- Austria, por cierto, es líder en esto y empezará a aplicar una ley en 2009 − **dimensión del género en la elaboración de presupuestos** − […]
- L'Austria, a tale proposito, dà l'esempio in questo contesto e dal 2009 entrerà in vigore una legge, sul **bilancio di genere […]**

The relatively low number of occurrences of *gender budgeting* in French in the latest parliamentary terms may however be due to an effort to avoid borrowing the English term, as web usage seems to suggest that its use is even more widespread in French than in Italian (69,000 hits for "le gender budgeting" as opposed to 14,000 for "il gender budgeting" and a mere 53 for "el gender budgeting". While such French expressions as *intégration de la dimension de genre dans le processus budgétaire* return very few results on Google (30), shorter equivalents fare much better and actually show some degree of implantation in academic texts (108,000 results for *budgétisation sensible au genre*). As for the expression *bilancio di genere*, it seems to have worked its way into the Italian language with 1,840,000 results.

### 3.3. *Naming and shaming*

Coordinated compounds are quite frequent in English, and many of them (*drink and drive*, *shake and bake*) involve some kind of alliteration (Renner 2006). The rate of borrowing for the most frequently used expression of that type that contains the *-ing* morpheme in the Europarl corpus is low, a fact which may be due to the presence of a non-lexical word in the expression, but also to the relative rarity of *shame* used as a verb and the consequent opacity of its meaning in the target languages involved. The expression was nonetheless considered to be used frequently enough in French to justify the suggestion of a French equivalent by the CGTN, which was published in the *Journal officiel* of January 19, 2010:

**mise au pilori**
*Domaine :* Tous domaines.
*Synonyme :* stigmatisation, n.f.
*Définition :* Pratique consistant à publier le nom de personnes physiques ou morales impliquées dans des activités tenues pour répréhensibles.
*Équivalent étranger :* naming and shaming.

The results in Table 6 show that in spite of a steady increase in the use of the expression over the past 13 years in the English part of the corpus, the number of borrowings has remained quite low in all three Romance languages under study.

| **Language** | **1996-2003 (nb of hits)** | **1999-2004 (nb of debates)** | **2004-2009 (nb of debates)** | **2009-2014 (nb of debates)** |
|---|---|---|---|---|
| **English** | 26 | 11 | 27 | 18 |
| **Spanish** | 2 | 2 | 1 | 0 |
| **French** | 2 | 2 | 1 | 1 |
| **Italian** | 1 | 2 | 1 | 0 |

*Table 6: Compared use of* naming and shaming *over three parliamentary terms in French, Italian and Spanish*

In the absence of a clear translation equivalent (at least until 2010 in the case of French), such scarcity of borrowing gives rise to a high level of variation not only in French, but also in the other two languages involved, as witnessed by the following excerpts from the Europarl corpus:

- If you the Court will not do your job of **naming and shaming**, then we will have to do it for you.
- Si el Tribunal no cumple con su deber de **nombrar y afear la conducta de los responsables**, nosotros deberemos hacerlo por ustedes.
- Si vous, la Cour, **ne dévoilez pas ces noms et ne dressez pas une liste de "coupables"**, nous devrons alors le faire à votre place.
- Se la Corte non intende fare il proprio lavoro **nominando e svergognando i trasgressori**, allora dovremo occuparcene noi.

- There must be a system of **naming and shaming** for those businesses that are careless.
- Debe haber un sistema que permita **señalar con el dedo** a las empresas negligentes.
- Il faut mettre en place un système consistant à **désigner nommément et à stigmatiser** les entreprises indélicates.

- Occorre un sistema per **indicare e stigmatizzare** le imprese che danno prova di negligenza.

- What is important here is public health, rather than some kind of **naming and shaming**.
- Lo que importa es la salud pública, no **publicar los nombres de las empresas para su escarnio**.
- Ce qui importe dans ce dossier c'est la santé publique et non de **jeter l'opprobre** d'une manière ou d'une autre.
- Ciò che è importante qui è la salute pubblica , non **mettersi a distribuire nomi e discredito**.

- Mr President, today I would like to start a process of **naming and shaming** those Member States that fail to implement European directives.
- Señor Presidente, hoy voy a iniciar un proceso de **denuncia** de aquellos Estados miembros que no ponen en práctica las directivas europeas.
- Monsieur le Président, je souhaite commencer aujourd'hui par **citer, afin de leur faire honte**, les États membres qui ne transposent pas les directives européennes.
- Signor Presidente, mi sembra giunto il momento di **additare al pubblico biasimo** quegli Stati membri che trascurano di applicare le direttive europee.

### 3.4. *Level playing field*

Multi-word expressions that do not involve coordination seem to be borrowed more frequently than coordinated compounds. Such is the case of the expression *level playing field*, a sports metaphor which is very often used in politics. The fact that it may be most commonly associated with football might explain why the expression shows a relatively high frequency of use in the Spanish and Italian parts of the corpus since football is quite a popular sport in Spain (the expression is found 33 times in Italian, 25 times in Spanish and 16 times in French, for a total of 330 occurrences in the English part of the corpus). Whether the expression is borrowed literally in another language (as is the case in Spanish and Italian in the example below) or translated by another expression (as is the case for French), it is often followed by a paraphrase:

- The purpose of the directive is supposed to be to create a **level playing field** in Europe.
- El objetivo de la misma pretende ser crear en Europa un **level playing field**, es decir, igualdad de armas.
- Le but de cette directive est de créer un **terrain d'égalité**, c'est-à-dire, une égalité des armes dans toute l'Europe.

- La direttiva dovrebbe mirare alla creazione di un **level playing field** in Europa, ossia garantire che le parti siano ad armi pari.

### 3.5. *Phasing out*

The term *phasing out* is another term which is often borrowed by Romance languages in the Europarl corpus. The figures obtained (31 occurrences in Spanish, 30 in French and 40 in Italian for a total of 513 in English) suggest that phrasal verbs might be more likely to be borrowed than other two-word expressions, a fact which remains to be confirmed through the study of more numerous examples:

- […] the possibility of **phasing out** the special conditions must also be discussed.
- […] hay que hablar también sobre un **phasing out** de las condiciones especiales.
- […] il faut aussi parler de la période de **suppression progressive** des mesures spécifiques supplémentaires.
- […] bisogna prevedere un **phasing out** delle condizioni speciali.

As suggested by the previous example, the expression tends to be borrowed more often in Spanish and Italian than in French, and it is more likely to be borrowed when it is used in the *-ing* form (208 occurrences out of 513) in English, even though use in the passive voice in English may also trigger such borrowings:

- Either they are priority hazardous substances and must be phased out or they are not.
- se trata de sustancias prioritarias peligrosas y entonces pertenecen a un **phasing out** o no.
- Soit il s'agit de substances dangereuses prioritaires et elles doivent faire l'objet d'un **phasing out**, soit elles ne le sont pas.
- si tratta di sostanze pericolose prioritarie e, in quanto tali, sono destinate al **phasing out**, oppure non lo sono.

### 3.6. *Remailing*

Use of the word *remailing* is worth mentioning for several reasons. It is actually the only *-ing* form that we could find in the corpus that is actually used as frequently or even more in some of the Romance languages under study (7 occurrences in Italian and 9 occurrences in Spanish) as it is in English (7 occurrences), as the borrowing is sometimes used to translate the words *redirection* and *reposting*:

- A reserved sector, a universal service obligation, which is attacked, even undermined, by unauthorised **redirection** can never meet with our approval.
- No puede merecer nuestra conformidad una zona reservada, un compromiso de servicio universal, que se vea contrarrestado, cuando no minado, por un **remailing** inadmisible.
- Un domaine réservé, une obligation de service universel contrecarrée voire minée par un **repostage** illicite ne peut recevoir notre adhésion.
- Un'area riservata e un obbligo al servizio universale indeboliti, anzi minati da un inammissibile **remailing** non possono incontrare la nostra approvazione.

- Mr President, Mr Commissioner, ladies and gentlemen, **reposting** is surely an aspect of the practice of dumping.
- El **remailing** es con toda seguridad una práctica competencial clasificable en la categoría de dumping.
- Monsieur le Président, Monsieur le Commissaire, chers collègues, le **repostage** constitue certainement une pratique de dumping.
- Signor Presidente, signor Commissario, onorevoli colleghi, la **reimpostazione** è certamente una forma di concorrenza attraverso il dumping.

Even though users of the word *remailing* in French were never "named and shamed" by the French Academy or the CGTN, the word is not to be found in the French section of the Europarl corpus, where the equivalents *repostage* and *réexpédition* are used. The high rate of borrowing observed in both Spanish and Italian might be due to the fact that the term is easily understandable because of the previous borrowing of *email*, although that argument would hold true for French as well.

## 4. Conclusion

EU simultaneous translators probably do contribute to the spread of English *V-ing* terms or expressions in their target language, even though many of their borrowings are ultimately replaced by translation equivalents that eventually take hold, in some cases with the help of an official policy that encourages their use. Our figures show that Spanish and (to a lesser degree) French tend to borrow fewer words containing the *-ing* morpheme than Italian, but it remains to be determined whether this is specific to this particular type of borrowing, as the permeability of Italian to English words in general is a well-known fact, especially in such specialized fields as information technology. Another finding of our study, however intuitive it might seem at first sight, is the fact that the longer the expression using words containing the *-ing* morpheme, the lesser the chance that implantation of the borrowing will follow. Once again, this is probably true of borrowings in general. Finally, the fact that phrasal verbs seem

to be more frequently borrowed than other two-word expressions needs to be tested on a larger scale.

## Bibliography

ALVAR EZQUERRA M. (1995), *La formación de palabras en espagnol*, Madrid, Arco Libros.
BARREAU J.-L. (2001), Des emprunts de l'espagnol péninsulaire aux autres langues européennes, *Cahiers de l'Institut de Linguistique de Louvain*, Vol. 27 (3-4), 89-112.
DEROY L. (1956), *L'emprunt linguistique*, Paris, Les Belles Lettres.
GUILBERT L. (1975), *La créativité lexicale*, Paris, Larousse.
GROSSMANN M., RAINER F. (dirs.) (2004), *La formazione delle parole in italiano*, Tübingen, Niemeyer.
HUMBLEY J. (2008), Emprunts, vrais et faux, dans le *Petit Robert 2007*, In PRUVOST J. (dir.), *Les journées des dictionnaires de Cergy : Dictionnaires et mots voyageurs. Les 40 ans du Petit Robert, de Paul Robert à Alain Rey*, 221-238, Herblay, Éditions des Silves.
MAKRI-MOREL J. (2009), *La création lexicale en espagnol péninsulaire contemporain : étude néologique, typologie des procédés et réflexions*, Lille, ANRT.
MIRANDA J. A. (1994), *La formación de palabras en español*, Salamanca, Colegio de España.
MITTERAND H. (1986), *Les mots français*, Paris, Presses Universitaires de France.
PICONE M. D. (1996), *Anglicisms, Neologisms and Dynamic French*, Amsterdam, Benjamins.
RENNER V. (2006), *Les composés coordinatifs en anglais contemporain*, PhD dissertation, Université Lumière Lyon 2.
TIEDEMANN J. (2009), News from OPUS – A collection of multilingual parallel corpora with tools and interfaces, In NICOLOV N., ANGELOVA G., MITKOV R. (eds.), *Recent Advances in Natural Language Processing V: Selected papers from RANLP 2007*, 237-248, Amsterdam, Benjamins.
TOURNIER J. (1991), *Précis de lexicologie anglaise*, Paris, Nathan.

### Internet sources

*Cent termes français du vocabulaire technique recommandés par la CGTN* (2004, MCC, DGLFLF), *www.dglflf.culture.gouv.fr/publications/vocabulaires/100termes.pdf*
*Trésor de la langue française,* online edition, http://atilf.atilf.fr/
*OPUS, the open parallel corpus*, http://opus.lingfil.uu.se/
*Vocabulaire de l'économie et des finances* (2006), http://fr.scribd.com/doc/120507304/Vocabulaire-de-l%E2%80%99economie-et-des-finances

# Neologismos, políticas lingüísticas feministas y norma académica en los glosarios en línea Gentyll del ámbito del deporte

Mercedes Bengoechea[1]

## Resumen

Nuestro equipo lleva tiempo estudiando la penetración de las políticas lingüísticas antisexistas en diversos ámbitos. Pese a las críticas feministas, detectamos que algunos de los recursos lexicográficos y terminológicos disponibles manifiestan los mismos defectos por los que se les ha criticado: ausencia de términos para denominar a las mujeres y/o denominación sexista para éstas. Además de ponerlo de manifiesto, y asumiendo la necesidad de disponer de términos con los que denominar de forma no sexista a los dos sexos, decidimos crear una serie de glosarios inglés-español, alternativos a los existentes, que se hicieran eco de las políticas lingüísticas antisexistas y que recogiesen los agentivos aplicados tanto a mujeres como a hombres en distintos campos de actividad: Gentyll on-line glossaries/Glosarios Gentyll en línea (http://gentyll.uah.es/glossaries.html).

En este artículo tratamos de mostrar cómo resolvimos las dificultades lexicográficas encontradas para la elaboración de los glosarios, tanto en inglés como en español. Las políticas no sexistas en ambas lenguas otorgan especial importancia a la denominación de los sexos, pero difieren en las soluciones propuestas: neutralización en inglés, y feminización en español. En consecuencia, las soluciones adoptadas en el glosario español fueron diferentes a las del glosario inglés.

En el caso inglés, el procedimiento de recogida de los neologismos que han venido a sustituir a los agentivos marcados por el sexo (*-man, -woman, -ess, -ette*...) consistió en buscar los términos sustitutivos: a) que las propias guías de uso no sexista del inglés recomiendan y que raramente se recogen en las bases terminológicas, y/o b) que poco a poco se han ido creando o utilizando en las sociedades anglófonas.

En la recogida de neologismos en español se conjugaban varios factores, entre los que sobresalen el androcentrismo cultural imperante, que lleva a aplicar de forma inconsciente el término masculino existente por la connotación negativa o peyorativa del femenino, y la falta de un uso mayoritario anterior. En ocasiones la lengua ofrecía además varias posibles realizaciones del femenino para términos minoritarios o no registrados. En tales casos, nuestro objetivo fue ofrecer, de entre todas las posibilidades potenciales, un neologismo que,

[1] Universidad de Alcalá, mercedes.bengoechea@uah.es

siguiendo los últimos criterios de la RAE, respondiera simultáneamente a las políticas lingüísticas feministas planteadas en los últimos años en España.

**Palabras clave**

Lexicografía sexista, guías feministas, denominaciones profesionales no sexistas, términos sexuados en español, inglés neutral, glosarios con perspectiva de género, agentivos.

## 1. Introducción: agentivos femeninos en los recursos lexicográficos y terminológicos

Pese a que en las últimas décadas los feminismos han criticado el sexismo y androcentrismo que imperan en repertorios terminológicos y diccionarios (p. ej. Gershuny, 1977; Forgas, 1986; Kramarae, 1992; Pauwels, 1998; Vargas et al, 1998; Bowker, 2001; Lledó, 2004; Medina Guerra, 2004), nuestro equipo detectó que algunos de los recursos lexicográficos utilizados en la traducción manifiestan los mismos defectos por los que habían sido criticados. Aun hoy en día en ocasiones se puede encontrar una denominación sexista para las mujeres o una conspicua ausencia de términos profesionales y agentivos femeninos para ellas (Bengoechea y Cabellos, 2012; Bengoechea, Cabellos y Simón, 2013).

Tres razones pueden explicar la ausencia de agentivos femeninos (y de algunos masculinos) de muchos diccionarios, bases de datos terminológicas y glosarios. En primer lugar puede deberse a que la lexicografía contemporánea basada en corpus refleja los usos reales de la lengua y no los potenciales. Por ejemplo, hasta hace bien poco tiempo agentivos para mujeres como "obispa" o agentivos para hombres como "matrón" sólo habrían existido en español en potencia; la configuración social de los sexos impedía su existencia real[2].

Una segunda razón puede encontrarse en el hecho de que, incluso cuando la existencia del agentivo está documentada, la lexicografía (sobre todo la dedicada a la lengua común) presta mayor atención a los usos mayoritarios que a los muy minoritarios, por lo que tiende a ignorar esa existencia. Tal sería el caso del término femenino "fresadora" en español, que puede encontrarse en portales de ofertas de empleo[3], pero que, sin embargo, no ha sido reconocida en el diccionario normativo del español (a partir de ahora DRAE[4]):

---

[2] En la futura 23ª edición del diccionario normativo del español (que en junio de 2012 ofrece la Web de la RAE) figurará "matrón"; no así "obispa", que sin embargo sí figura en RAE y Asociación de Academias de la Lengua Española (2009: I, §2.6m, 109).

[3] Por ejemplo: infoempleo.com, infojobs.net, laboris.net, computrabajo.es (búsqueda: 12 junio 2012).

[4] Se citará como DRAE (1992, 2001 o 2014) en el cuerpo del texto, pero aparecerá en las referencias bibliográficas como RAE (Real Academia Española) (1992, 2001 o 2014).

> **fresador. 1. m.** Operario encargado de manejar las diferentes clases de máquinas para fresar. (DRAE, 2001. Sin enmendar para la futura versión de 2014. Consulta: junio 2012).

Una tercera razón radica en que quienes compilan los recursos lexicográficos y terminológicos tienen poco interés en incorporar las políticas lingüísticas no sexistas. Es precisamente el afán por hacer efectivas esas políticas la razón por la que nuestro equipo acometió la elaboración de los glosarios on-line *Gentyll*. Nuestro deseo era ofrecer a los recursos lexicográficos existentes[5] o futuros los agentivos femeninos (y algunos masculinos) que no figuraban hasta ahora, siguiendo las políticas de uso no sexista de las lenguas.

A lo largo de este trabajo explicaremos cuáles son esas políticas, los factores sociolingüísticos tenidos en cuenta para elaborar los glosarios de agentivos y, cuando se trataba de un neologismo, las directrices seguidas en su creación.

## 2. Políticas lingüísticas no sexistas en español[6] (e inglés)

Las críticas feministas a los usos androcéntricos de las lenguas y sus consiguientes propuestas de uso alternativo, junto a la transformación social de nuestras sociedades hacia una mayor igualdad entre los sexos, han tenido como resultado transformaciones en la lengua inglesa (Cooper, 1989; Rubin, Greene, Schneider, 1994; Pauwels, 1997a, 1997b, 2003) y española (Nissen, 2002; Velasco, Vázquez, Ibáñez, 2009; Bengoechea, 2011). Ya en los años ochenta Fasold (1988), Cooper (1989) o Fasold et al. (1990) categorizaron como políticas lingüísticas las propuestas feministas de reforma verbal de la lengua inglesa, las cuales han sido recogidas y argumentadas por Frank y Treichler (1989), Pauwels (1998, 2003) o Hellinger y Bußmann (2001-2003). Por ser suficientemente conocidas las argumentaciones y las propuestas, no nos extenderemos en ellas.

En el caso del español en España podemos hablar con toda propiedad de políticas lingüísticas, puesto que la *Ley Orgánica 3/2007 de 22 de marzo, para la igualdad efectiva de mujeres y hombres*, (a partir de aquí, denominada Ley de Igualdad) recoge la obligatoriedad de utilizar el lenguaje no sexista en los textos administrativos (artículo 14.11), en los proyectos de las TIC y sociedad de la información financiados con fondos públicos (artículo 28.4) y en los medios de comunicación (artículos 37 y 38)[7].

---

[5] Muchos de los cuales constituyeron la base de nuestros glosarios.

[6] Las políticas lingüísticas que nos guían son las que rigen en España; por tanto nuestros glosarios recogen únicamente términos del español europeo.

[7] No era la primera normativa en que el uso no sexista del lenguaje se recomendaba u obligaba (Bengoechea, 2011; Centenera Sánchez-Seco, 2011a). Esa obligatoriedad es aún materia de controversia y el seguimiento de tal supuesta obligatoriedad es a menudo relativo, por supuesto (Bengoechea, 2011; Centenera Sánchez-Seco, 2011b).

Uno de los ámbitos al que la Ley de Igualdad afecta especialmente es el ámbito del empleo, aunque para ese dominio, recogido en el artículo 5, la ley no hace referencia explícita al lenguaje:

> Artículo 5. Igualdad de trato y de oportunidades en el acceso al empleo, en la formación y en la promoción profesionales, y en las condiciones de trabajo.
> El principio de igualdad de trato y de oportunidades entre mujeres y hombres, aplicable en el ámbito del empleo privado y en el del empleo público, se garantizará, en los términos previstos en la normativa aplicable, en el acceso al empleo, incluso al trabajo por cuenta propia, en la formación profesional, en la promoción profesional, en las condiciones de trabajo, incluidas las retributivas y las de despido, y en la afiliación y participación en las organizaciones sindicales y empresariales, o en cualquier organización cuyos miembros ejerzan una profesión concreta, incluidas las prestaciones concedidas por las mismas. [...] (L.O. 3/2007)

Pese a no hacer mención expresa de políticas lingüísticas, o tal vez por ello, este artículo llevó a modificar la *Ley 56/2003, de 16 de diciembre, de Empleo*, cuyo artículo 22.bis queda redactado en los siguientes términos en la *Ley 35/2010, de 17 de septiembre, de medidas urgentes para la reforma del mercado de trabajo*:

> Artículo 22 bis. Discriminación en el acceso al empleo.
> 1. Los servicios públicos de empleo, sus entidades colaboradoras y las agencias de colocación en la gestión de la intermediación laboral deberán velar específicamente para evitar la discriminación tanto directa como indirecta en el acceso al empleo. Los gestores de la intermediación laboral cuando, en las ofertas de colocación, apreciasen carácter discriminatorio, lo comunicarán a quienes hubiesen formulado la oferta.
> 2. En particular, se <u>considerarán discriminatorias las ofertas referidas a uno de los sexos</u>, salvo que se trate de un requisito profesional esencial y determinante de la actividad a desarrollar. En todo caso se considerará discriminatoria la oferta referida a uno solo de los sexos basada en exigencias del puesto de trabajo relacionadas con el esfuerzo físico. (*Subrayado nuestro*)

En la aplicación de esta norma e incluso con anterioridad a su aprobación[8], se ha entendido que los anuncios de empleo deben utilizar un lenguaje del que no quepa deducir discriminación contra mujeres u hombres, lo que implicaría que todos ellos hacen uso de lenguaje no sexista. En base a ello la Inspección de Trabajo ha realizado campañas de detección en Internet o en los medios de comunicación de anuncios que puedan ser considerados no neutros, llamando a las empresas intermediarias de colocación, o a la compañía anunciadora directamente, requiriendo cambios o sancionando por discriminación. Uno de los cambios solicitados es el anuncio del puesto de trabajo en sus formas

[8] Como demuestra el artículo del académico de la RAE Arturo Pérez Reverte, "Chantaje en Vigo", aparecido en el semanario *XLSemanal* - 06/12/2009, en el que califica de "neonazi" a la inspectora de trabajo que obligó a que una oferta de empleo de "auditor" incluyese también la forma femenina: *auditor/a*.

masculina y femenina, de forma que se entienda sin ningún género de dudas que puede solicitarlo tanto un hombre como una mujer.

Otro de los ámbitos al que han afectado las políticas lingüísticas ha sido el deporte. Como práctica derivada de la Ley de Igualdad, las retransmisiones deportivas femeninas deberían ser práctica común. Así lo recomendaba el Parlamento Europeo (2003a), a través de la Comisión de Derechos de la Mujer e Igualdad de Oportunidades en mayo de 2003, donde planteó una serie de pautas para garantizar la igualdad de derechos en el deporte (Romero Granados, 2008). Las mismas peticiones y sugerencias fueron aprobadas posteriormente el 5 de junio de 2003 en la Resolución del Parlamento Europeo sobre las mujeres y el deporte (2002/2280(INI). Entre otras, el Parlamento Europeo:

- Requiere a los medios de comunicación que velen por la cobertura equilibrada del deporte femenino y masculino, así como por una representación no discriminatoria de las mujeres en el deporte. [...]
- Propone, en el marco de la revisión de la Directiva 89/552/ CCE "Televisión sin fronteras", inscrita en el artículo 3 bis relativo a la retransmisión de acontecimientos de gran importancia, la necesidad de que todos los estados miembros incluyan y garanticen la retransmisión de las competiciones deportivas femeninas. [...]

(Parlamento Europeo, 2003b)

Pero, para cumplir las recomendaciones y la normativa, el personal técnico de la prensa, las televisiones y las radios necesitarían tener a su disposición glosarios que ofreciesen los sustantivos que designan a las jugadoras y su posición en el campo. En 2009 estos glosarios no existían.

Conscientes de esa necesidad, nuestro equipo presentó un proyecto de investigación al Ministerio de Ciencia e Innovación en el que, entre otros objetivos, se planteaba la elaboración de tales glosarios (proyecto FEM2009-10976). Partiendo de la base de que, para poder implementar con éxito las políticas lingüísticas, las empresas de trabajo temporal, las compañías intermediarias de colocación y el ejercicio del periodismo requieren términos con los que denominar de forma no sexista a los seres humanos, nuestro equipo decidió crear una serie de glosarios en español, alternativos a los existentes, que recogiesen los agentivos no sexistas aplicados tanto a mujeres como a hombres en distintos campos de actividad. El primer glosario que creamos fue el de deportes, cuya versión inicial[9] en español se recogió en la publicación del Instituto de la Mujer *Hablamos de deporte* (Alfaro, Bengoechea, Vázquez, 2010).

[9] Se han realizado después modificaciones al glosario para dar cabida a los cambios sociales y a las enmiendas a la 22ª edición del diccionario normativo del español que pueden consultarse en la Web de la Real Academia Española (http://www.rae.es/rae.html).

Como nuestra preocupación también incluía el ámbito traductológico, se decidió que los glosarios fuesen bilingües español-inglés. Dado que las sociedades anglófonas tienen larga tradición de políticas lingüísticas no sexistas que han impulsado instituciones diversas e implementado los gobiernos, las universidades, los medios de comunicación o las editoriales, nuestros glosarios en inglés trataban de recoger sus recomendaciones, algo que las bases de datos lexicográficas no habían hecho o no lo habían hecho de forma consistente (Bengoechea y Cabellos, 2012; Bengoechea, Cabellos, Simón, 2013). De momento están finalizados los del ámbito del deporte (2.900 entradas), de ocio y turismo (1.900 entradas) y del derecho (2.500 entradas); y a lo largo de 2013 se presentarán los de política, medicina y educación: *Gentyll on-line glossaries/Glosarios Gentyll en línea* (http://gentyll.uah.es/glossaries.html). Los glosarios están sujetos a constante revisión, al entender el equipo que los cambios sociales afectan particularmente a esta área del lenguaje y por la intención de incorporar algunas de las nuevas políticas de la RAE.

En el resto de este trabajo nos centraremos particularmente en el glosario inglés-español del ámbito deportivo, que consideramos suficientemente representativo de la labor desarrollada. Trataremos de mostrar cómo resolvimos las dificultades lexicográficas encontradas para la inclusión de neologismos, especialmente en español. Una de nuestras más importantes fuentes fueron las guías de uso no sexista en ambas lenguas, las cuales otorgan especial importancia a los agentivos para mujeres, pero difieren en las soluciones propuestas: neutralización en inglés, y feminización en español. En consecuencia, las soluciones adoptadas en el glosario español fueron diferentes a las del glosario inglés.

## 3. Agentivos y neologismos no sexistas en los glosarios deportivos en inglés

Al igual que sucede en el caso español, en la mayoría de los países de habla inglesa existe normativa en contra de la discriminación por sexo, que ha sido interpretada de forma que obliga a un uso no sexista de la lengua, especialmente en temas laborales. Tal sería *Sexual Discrimination Act 1975* en Gran Bretaña, cuya interpretación llevó a prohibir cualquier tipo de discriminación en las ofertas de empleo (Kennedy, 1996). Para obrar en consecuencia, en los países anglófonos se recurrió a agentivos no sexistas.

La idea directriz de las políticas no sexistas de agentivos en inglés es su neutralización formal respecto al sexo, de manera que la forma aplicada a mujeres y a hombres sea idéntica (ejemplo: *chair*, para presidenta o presidente de un comité), si bien se admiten términos con alternancia en las desinencias masculina **y** femenina para cada uno de los sexos respectivamente (*chairman/chairwoman*). Los agentivos femeninos con desinencias derivativa (*-ess, -ette, -ine*), asimétricos respecto a la forma masculina (p.ej.: *usher/usherette*), son

considerados sexistas. En general, el debate se ha centrado en si elegir un término de nueva creación (*chairperson*) o un término ya existente y relacionado semánticamente con el que se intenta sustituir (*president*) (Pauwels, 2003: 559).

Para nuestro glosario, el procedimiento de recogida de los neologismos que han venido a sustituir a los agentivos marcados por el sexo (*-man, -woman, -ess, -ette*...) ha consistido en buscar los términos sustitutivos que las propias guías de uso no sexista del inglés recomiendan –y raramente se recogen en las bases terminológicas–, o aquellos que poco a poco se han ido creando o utilizando en las sociedades anglófonas. Las fuentes han sido fundamentalmente tanto las directrices de las políticas antisexistas de las diversas editoriales y medios de comunicación (Maggio, 1997; Miller, Swift, 1995; MLA, 2003; Schwartz, 1995; libros de estilo de la BBC y de los medios norteamericanos; etc.) como nuevas incorporaciones al inglés recogidas en webs institucionales (United States Department of Labor, 2010-2011; "Guidelines for non discriminatory job advertising and recruitment" de Australia; las guías de la State Service Commision de Nueva Zelanda; portales de universidades estadounidenses, canadienses, neozelandesas, australianas, irlandesas o británicas; etc.). Realizamos además búsquedas en Internet con el buscador Google, utilizando claves de búsqueda como "feminist", "non sexist naming", "non discriminatory language", "gender fair language", "gender neutral", "non biased language" + "guidelines", "gender" + "language" + "policy"...

Estos neologismos no suponen muchas entradas en número en nuestros glosarios. Algunos de ellos llevan el suficiente tiempo en la lengua inglesa como para que ya hayan sido recogidas en diccionarios y bases de datos. Lo que éstos no contemplan son los usos atendiendo a la variación sociolingüística por motivos de políticas antisexistas; los nuestros sí recogen esta variación, puesto que ese constituye uno de los propósitos de nuestro trabajo. A modo de ilustración, véase en Figura 1 cómo aparecen en nuestro glosario inglés las varias denominaciones acuñadas para sustituir en los deportes de lucha a *mat chairman* ("jefe/jefa de tapiz"): "mat chair", "mat chairperson", "mat chairwoman". Junto a las denominaciones, las entradas recogen las recomendaciones antisexistas:

*Fig. 1. Gentyll >Sports >'mat chair'*

## 4. Agentivos y neologismos no sexistas en los glosarios deportivos en español

Proporcionar la forma femenina de términos en el ámbito del deporte aplicados a mujeres dentro y fuera de canchas y pistas fue el principal objetivo que nos marcamos en los glosarios en español. El segundo, ofrecer los masculinos de aquellos términos estereotípicamente considerados propios de mujeres (p.ej. *azafato de pista*, en tenis).

Al revés que en el inglés, desde las primeras guías publicadas (Consejo de Europa, 1986; Unesco, 1990; NOMBRA, 1994) la recomendación antisexista respecto a los agentivos en español es utilizar la forma que corresponda al sexo de la persona de referencia, lo que se ha leído como necesidad de disponer de formas femeninas en la lengua[10]. Por ese motivo, las agencias igualitarias se ocuparon ya en los años 90 de editar glosarios con los términos profesionales, cargos y ocupaciones de mujeres que aparecían exclusivamente en masculino en el DRAE (por ejemplo, Lledó Cunill, 1995). Frecuentemente esto se realizó

[10] Así lo entendió también la Orden de 22 de marzo de 1995 del Ministerio de Educación y Ciencia, por la que se adecua la denominación de los títulos académicos oficiales a la condición masculina o femenina de quienes los obtengan, figurando en esa Orden la denominación de títulos para mujeres como "médica especialista", "técnica superior", "ingeniera técnica", etc.

mediante la explícita feminización en -a del término masculino correspondiente (*pilota*). Posteriormente, y conforme las mujeres expandían los dominios de su actividad profesional y el DRAE se abría lentamente a la diferencia sexual en sus ediciones de 2001 y 2014 –parte de la cual ya está disponible en la web de la RAE–, aparecieron nuevas publicaciones con glosarios profesionales y laborales de las agencias de igualdad (Lliteras et al., 2004; Lledó Cunill, 2006; Guerrero Salazar, 2007) y de otras instituciones (Instituto Cervantes, 2011). Todas ellas constituyen la base de nuestros glosarios, que tratan de compaginar las propuestas feministas de denominaciones no sexistas con los usos normativos dictados por la Real Academia Española (fundamentalmente, RAE y Asociación de Academias de la Lengua Española, 2009), como se expondrá a continuación. Para términos cuya forma femenina (y alguna masculina) no figuraba ni en el DRAE, ni en los glosarios especializados, ni en las guías feministas, buscamos su realización a través de Google con el fin de detectar su existencia.

Debemos hacer hincapié, sin embargo, en la dificultad de evaluar tanto la fiabilidad de los datos encontrados en Internet, como el grado de consenso respecto a las formas encontradas, algo que se ha observado frecuentemente en trabajos diversos sobre lexicografía y terminología (Scholze-Stubenrecht, 2001; Kilgarriff, Grefenstette, 2003; Docherty, 2007; Ludeling, Kyoto, 2008; etc.). En el caso de agentivos, es notable la cantidad de femeninos no normativos que pueden encontrarse en la escritura popular, especialmente en blogs y chats informales de España y América: *la concursanta, la contrincanta, la representanta*, *la veraneanta, la testiga, la visitanta, la sumillera*, etc. De hecho, esta proliferación confirma la tendencia del español que apuntara Julio Casares: «una tendencia constante de la lengua a diferenciar los géneros por medio de desinencias especiales y según normas analógicas. [...] [E]l pueblo ve en la *a* la terminación femenina por excelencia y [...] esta corriente [...] se nos muestra cada día más poderosa» (1961: 191-193). Pero, puesto que no siempre la tendencia es así entendida por todas las capas de la población o las instituciones que dictan la norma, todos los hallazgos de Internet deben ser estudiados individual y cuidadosamente.

Por supuesto, la mayoría de las entradas en español de nuestro glosario deportivo no son neologismos, sino términos plenamente documentados y de larga vida en la lengua, tanto la forma femenina como la masculina: *el tenista/la tenista; el jugador/la jugadora*. Sin embargo, otras de las entradas en femenino eran más difíciles de documentar. El uso minoritario de algunas de ellas podía rastrearse en Internet (tal sería el caso de *la delantera centro*), mientras otras no aparecían ni en las guías feministas, ni en los recursos electrónicos (p.ej. *la extrema nata*). Este ejemplo constituye un caso ilustrativo de la situación que se produciría en 2009 al intentar nombrar a aquella jugadora de fútbol ofensiva o delantera que jugase en la banda en fútbol. La forma femenina de "el extremo" no figuraba en 2009 en los glosarios terminológicos del deporte (p.ej. TERMCAT, consulta 2009, 2010; Ordoño Muñoz, 2004). El diccionario de la

RAE tampoco lo aportaba –ni probablemente lo hará en 2014– en su acepción 11 del sustantivo[11]:

> **extremo.** [...] **11.** m. En el fútbol y otros deportes, miembro de la delantera que, en la alineación del equipo, se sitúa más próximo a las bandas derecha o izquierda del campo. (DRAE 2001)

La forma femenina que ofreciese nuestros glosarios sería, por tanto, un 'nuevo femenino'.

Podría pensarse que en tales ocasiones los 'nuevos' femeninos no serían neologismos realmente, sino simplemente resultado de aplicar una indiscutible regla morfológica de la lengua a la forma masculina: *el achicador de espacios/la achicadora de espacios*; *el carrilero/la carrilera*. Pero lo cierto es que la morfología y la sociolingüística del género gramatical en español son significativamente complejas. Ni siquiera el "mating problem" que parece solucionar la clásica descripción de Harris (1991) o la alternancia -o/-a es consistente, puesto que la RAE y la Asociación de Academias de la Lengua Española (2009) describen decenas de casos especiales que no cumplen la alternancia. Así pues, esta cierta inconsistencia morfológica y la complejidad sociolingüística de un tema que ocupa portadas en los medios y está sujeto a reivindicaciones y controversias nos llevan a pensar que algunas de las formas femeninas pueden considerarse casos de neologismos.

Sabemos que quizá resulte chocante definir como neologismo la mera feminización (*buza*) o masculinización (*matrón, amo de casa*) de un vocablo existente, pero nos tomamos la libertad de entender por neologismo –en lo que quizá sea una definición más amplia que las definiciones al uso– cualquier forma nueva que aparece en una lengua, en ciertos casos como nueva creación, que se crea para nombrar algo (un puesto) o alguien que (quizá) no existía anteriormente, pero al que o a la que hay necesidad de nombrar. Esta nueva forma se genera a partir de las estructuras morfosintácticas del español.

Que en el caso de la forma femenina de ciertos agentivos deportivos podemos hablar con propiedad de neologismos lo prueba un artículo aparecido simultáneamente a la publicación de nuestros glosarios del deporte (Gómez Torrego, 2010). Si no eran neologismos, el artículo prueba al menos que su presencia en la lengua era minoritaria y sujeta a debate:

> Y si son mujeres las componentes de un equipo de fútbol (y cada vez hay más), ¿habrá que hablar de la extrema derecha, la extrema izquierda, o será mejor decir la extremo derecho (¿también derecha?), la extremo izquierdo (¿también

[11] Si bien en los últimos meses de 2011 y los primeros de 2012 la forma "extrema" ha sido incorporada ya a otros diccionarios y glosarios (p.ej. *Concise Oxford Spanish Dictionary*), mientras el Diccionario VOX-Larousse ha convertido "extremo" en sustantivo común.

> izquierda?)?; ¿y deberíamos hablar de la portera o de la portero; de la delantera o de la delantero; de la media centro o de la medio centro? En nuestra opinión, como un partido de fútbol se plantea como una batalla entre dos ejércitos (más adelante hablaremos del léxico belicista de esta jerga) y las Academias de la Lengua Española en su *Diccionario panhispánico de dudas* han decidido que los sustantivos que designan a mujeres en el ámbito militar queden como comunes en cuanto al género (la soldado, la cabo, la sargento…), lo más razonable sería denominar a las mujeres futbolistas como la portero, la medio, la delantero, la extremo, etc. Incluso, a pesar de que las Academias proponen la forma femenina la árbitra, conculcando –eso sí, de forma deliberada- la regla que obliga a poner el artículo 'el' delante de sustantivos femeninos que empiezan por (h)a- tónica (el ama; el aya…), convendría decir mejor la árbitro, en consonancia con las otras palabras mencionadas. Sería una propuesta coherente y evitaría dicha conculcación, que no deja de ser excepcional. (Gómez Torrego, 2010: 152)

Para la inclusión de una u otra forma de neologismo femenino tuvimos que tomar en consideración una serie de factores sociolingüísticos a cuya complejidad hemos hecho referencia en párrafos anteriores.

## 5. Factores sociolingüísticos considerados

A continuación se detallan los factores, a veces contradictorios entre sí, que nuestras propuestas tuvieron que considerar. Para cada uno de los factores se proporcionan ejemplos que los esclarecen:

1. Cuando comenzamos la elaboración de nuestros glosarios en 2009 existían dos o más realizaciones potenciales del femenino para ciertos términos especializados, minoritarios o no registrados.

Por ejemplo, y por volver al término anteriormente mencionado, teníamos la posibilidad de la forma *la extremo* (que se asemejaría al ámbito militar y mantendría la posible metonimia del masculino, al referirse a la futbolista que domina el extremo del campo) o *la extrema* (por feminización 'regular' del término masculino).

2. En ciertos casos donde no era fácil encontrar una forma femenina consolidada en la lengua, el uso mayoritario tendía a ser la forma masculina. Esto se debía unas veces a la connotación sexista, negativa o peyorativa del término femenino (*perita*, *asistenta...)*, otras veces al androcentrismo cultural imperante y otras veces a la falta de actualización de una minoría de profesionales de la traducción, quienes tienden a aplicar normas caducas (ejemplo: "la árbitro") por desconocimiento de la última decisión académica (*árbitro, tra*, DRAE, 2001).

Un ejemplo sería el término *líbero/líbera*, en voleibol, de la que documentalmente pueden encontrarse la forma masculina en -o y la forma femenina en -a para jugadoras. Así, aunque medios como el diario asturiano *La*

*Nueva España* utilizan para mujeres la forma *líbera*[12], el DRAE mantiene exclusivamente la forma masculina referida a hombres, sin proporcionar indicación de la forma femenina:

> **líbero²**. (Del it. *libero*). **1.** m. En algunos deportes, jugador defensivo sin posición fija.
>
> (DRAE, 2001).

Mientras, una búsqueda en Internet con el artículo "la" demuestra que indudablemente la forma masculina para mujeres es dominante en Internet, tanto en voleibol como en fútbol:

- *"la líbero"+ fútbol* > más de **50.000 entradas**
- *"la líbera"+ fútbol* > 38 entradas, principalmente en portales americanos

- *"la líbero"+ voleibol* > más de **40.000 entradas**
- *"la líbera"+ voleibol* > 6 entradas, principalmente en portales americanos

(Búsqueda mediante Google, 22 de junio de 2012)

*Tabla 1. Búsqueda en Internet de forma femenina líbera o líbero (fútbol y voleibol)*

3. La falta de un uso documentado anterior se conjugaba en muchos casos con un tercer factor: la tendencia androcéntrica del discurso –incluido el discurso de la lexicografía (Bengoechea, 2008) – a utilizar el masculino como aposición del término "mujer".

Para ilustrarlo remitimos al siguiente ejemplo, tomado de Internet el 24 de junio de 2012, buscando "árbitro", "mujer árbitro", "mujer árbitra" y "árbitra". Pese a ser ya normativo el femenino "árbitra", el buscador halló, como puede verse, que mayoritariamente se emplea al masculino (*árbitro*), en aposición a "mujer" (65.900 entradas) o no (45.900 entradas), siendo minoritaria la forma femenina *árbitra* (27.300 entradas).

[12] Véase, a modo de ejemplo, http://www.lne.es/deportes/1921/grupo-gervasport-asegura-permanencia/610394.html.

| | | |
|---|---|---|
| • *"la árbitro"* | -> | 45.900 resultados |
| • ***"la mujer árbitro"*** | **->** | **65.900 resultados** |
| • *"la árbitra"* | -> | 27.300 resultados |
| • *"la mujer árbitra"* | -> | 62 resultados |

(Búsqueda mediante Google. 24 de junio de 2012)

*Tabla 2. Búsqueda en Internet de forma femenina árbitra o árbitro*

4. El cuarto factor en nuestra consideración fue decidir una política coherente en los casos en los que la norma académica muestra una cierta inconsistencia.

Un caso ilustrativo lo encontramos en el primer ejemplo de la tabla 2 ("la árbitro"). Podría pensarse que se trata de un error gramatical extendido debido a la relativa falta de uso del término, pero puede tratarse también de duda ante la inconsistencia de la RAE en sus entradas, acepciones y definiciones, lo que no contribuye a que la forma femenina se solidifique y afiance. Nótese cómo en el DRAE, "árbitro" parece ser un sustantivo masculino si se usa como núcleo del grupo nominal, pero común cuando va en aposición (*el/la juez árbitro*):

> **árbitro, tra.** [...] **2.** m. y f. Persona que en algunas competiciones deportivas cuida de la aplicación del reglamento.
>
> **juez árbitro. 1.** com. *Dep.* En competiciones deportivas, árbitro principal que toma las últimas decisiones.
>
> (DRAE, adelanto de la futura edición de 2014)

5. Un quinto factor –y quizá el más relevante– fueron los cambios en la norma. La política de la RAE en los últimos diez años no solo ha sido relativamente inconsistente, ha sido mudable, o cuanto menos vacilante, para ciertos términos registrados y normalizados. Este hecho ha contribuido a la inseguridad de la comunidad hablante, particularmente cuando las guías feministas se decantaban por una de las variantes (en justicia, debe reconocerse que las decisiones de la RAE y de las propias guías feministas frecuentemente reflejan la vacilación de la sociedad española en materia de género social y gramatical y una rapidísima transformación a la que resulta arduo seguir el paso). Nuestros glosarios han tratado de adaptarse a las muchas enmiendas realizadas en el DRAE para su 23ª edición, si éstas están disponibles en la Web. Las enmiendas continuarán produciéndose hasta la publicación definitiva de la 23ª edición en 2014, lo que nos obligará a mantenernos en constante alerta para incorporar los cambios a nuestros glosarios.

Un ejemplo de vacilación tanto por parte de la RAE como por parte feminista puede verse en "gerente (deportivo)", que en la edición 21ª del DRAE (1992) era un sustantivo común. En la edición 22ª del DRAE pasa a tener formas masculina y femenina, para hombre y mujer respectivamente (*gerente/gerenta*),

aunque se admite la forma en masculino para el femenino. Sin embargo en 2014 el DRAE clasificará el sustantivo que ha sido masculino durante diez años (*gerente*) como común, mientras mantendrá una forma femenina (*gerenta*), la cual pasará a ser un americanismo[13]:

> **gerente.** (Del lat. *gerens, -entis*). **1.** com. Persona que dirige los negocios y lleva la firma en una sociedad o empresa mercantil, con arreglo a su constitución. (DRAE, 21ª edición, 1992)
>
> **gerente, ta.** (Del lat. *gerens, -entis*). **1.** m. y f. Persona que lleva la gestión administrativa de una empresa o institución. MORF. U. t. la forma en m. para designar el f. *Ana es gerente.* (DRAE, 22ª edición, 2001)
>
> **gerente.** (Del lat. *gerens, -entis*). **1.** com. Persona que lleva la gestión administrativa de una empresa o institución.
>
> **gerenta. 1.** f. *Am.* Mujer que lleva la gestión administrativa de una empresa o institución. (DRAE, avance 23ª edición)

Por su parte, el glosario feminista de Lledó Cunill de 2006 ofrece también las dos formas, pero en este caso la preferencia es *gerenta*, forma que no se reconocía en la edición de su glosario de 1995.

6. Pese a que la *Nueva Gramática de la Lengua Española* (RAE y Asociación de Academias de la Lengua Española, 2009) simplificó y solucionó en gran medida las dudas ante las desinencias de los cargos y denominaciones en masculino y femenino, las excepciones a la regla son tan numerosas que contribuyen a la controversia existente y en ocasiones a la disparidad entre las guías feministas, la RAE y el uso social.

Por ejemplo, pese a que para la RAE *piloto* es un sustantivo común que carece de forma femenina alternativa (RAE y Asociación de Academias de la Lengua Española 2009: I, §2.5m, 102), al igual que otros sustantivos que designan grados de la escala militar (2009: I, §2.6c, 105), las guías feministas proporcionan una posible forma femenina además de la forma común. Y así lo recogen algunas webs, donde aparece *pilota* como uso muy minoritario[14].

7. En algunos casos para los que la RAE no se había pronunciado, quizá contábamos con propuestas de nuevas formas femeninas a nuestra disposición

---

[13] Y que carecerá de etimología, como suele suceder en el DRAE con términos aplicados a mujeres. Véase Forgas Berdet (2004)

[14] Por su parte, pese a que los sustantivos terminados en -ista son comunes, la RAE mantiene *modisto* como especial anomalía. Quizá por analogía también se haya masculinizado alguna forma femenina con morfema de género femenino en -isa: es fácil oír en ambientes festivos, informales o populares formas como *pitoniso* (18.900 resultados en una búsqueda en Google el 18 de junio de 2012).

(Lliteras et al., 2004; Lledó Cunill, 2006; Guerrero Salazar, 2007). Sin embargo, no todas ellas habían gozado de la misma aceptación social.

Por ejemplo, *fabricanta* no había cuajado y *detectiva* parecía haber prendido solo en Cataluña y en algunos países americanos.

8. Finalmente, tuvimos que plantearnos los casos de formas femeninas desde hace tiempo presentes en el sistema y autorizadas por el DRAE, pero desprestigiadas (*asistenta*) o rechazadas por las propias mujeres a quienes se aplicarían. Este rechazo puede deberse a la cacofonía (*la alera*), a puro androcentrismo (*la jueza de línea*), a una cierta connotación sexista del término (*la perita*) o a estar su espacio semántico ocupado por conceptos abstractos (*la técnica, la práctica del puerto*). En el uso de todas ellas se percibe vacilación, como reconocen la RAE y la Asociación de Academias de la Lengua Española (2009).

## 6. Directrices sociolingüísticas y morfológicas que guiaron los neologismos de *Gentyll*

La pauta general adoptada para la creación de neologismos puede resumirse de la siguiente forma: teniendo en cuenta los ocho factores enumerados, a partir de las bases de datos terminológicas del deporte y de diccionarios especializados existentes (que en 2009-2010 proporcionaban tan solo la forma para varones), nuestros glosarios se proponen ofrecer la(s) forma(s) femenina(s) –o, en contados casos, la forma masculina– que, siguiendo las más recientes políticas de la RAE, respondiera(n) simultáneamente a las políticas lingüísticas feministas planteadas en los últimos años en España. Es una opción relativamente conservadora –como suele serlo la propia práctica de la traducción– que, excepto en casos excepcionales (el criterio 6.2.4.), sigue la norma académica y al mismo tiempo se hace eco de las propuestas feministas y de la necesidad de las mujeres de ser nombradas.

A continuación se desgranan las directrices que sirvieron de guía:

### 6.1. Términos simples

6.1.1. Si existía documentación en el DRAE sobre la forma femenina o común

(a) Cuando existía documentación, en ésta sólo aparecía una forma femenina o una forma común, que ya estaba aceptada por la RAE y coincidía con la forma propuesta en las guías feministas, no había problema: ni se trataba de un neologismo propuesto por nuestro equipo, ni había controversia. Ejemplos: *la técnica-el técnico, la/el púgil, la/el pívot, la perita-el perito, la/el sparring, la árbitra-el árbitro, la abanderada-el abanderado, la/el central, la bateadora-el bateador, la/el pelotari, la/el artífice, la azafata-el azafato, la/el base, la/el suplente*, etc.

(b) Si el DRAE ofrecía más de una posibilidad (una forma común y una femenina: *la jefa/jefe, la jueza/juez, la principianta/principiante, la presidenta/presidente*...), en nuestros glosarios la forma femenina prima sobre la común, pero se ofrece también ésta en una nota de uso. Ejemplo: *vicepresidenta* de un club deportivo (Figura 2).

Any
'vicepresidenta': 3 registros encontrados

| # | Deporte | Mujer | Hombre | Uso Esp | Woman | Man | Uso Ing |
|---|---|---|---|---|---|---|---|
| 1 | Deportes en general | La vicepresidenta | El vicepresidente | 1 | The deputy chair | The deputy chair | - |
| 2 | Deportes en general | La vicepresidenta | El vicepresidente | 2 | The deputy chairwoman | The deputy chairman | 1 |
| 3 | Deportes en general | La vicepresidenta | El vicepresidente | 3 | The vice president | The vice president | - |

**Notas de uso (términos en español):**
1. También *la vicepresidente*.
2. También *la vicepresidente*.
3. También *la vicepresidente*.

**Notas de uso (términos en inglés):**
1. Sex-specific terms with *woman* or *man* are less and less frequent, being substituted by terms that do not specify sex (which are more comprehensive, more accurate and more consistent). Avoid the use of sex-specific terms (singular or plural) as indefinite, sex-inclusive occupational titles.

*Fig. 2. Gentyll >Deportes > 'vicepresidenta'*

En los casos excepcionales en que la forma femenina es muy restringida respecto a la común (p.ej. "gerenta", "vicegerenta"), se ofrece como preferible la forma común y en segundo lugar la forma femenina, pero se mantiene ésta incluso cuando el DRAE la ha hecho desaparecer del léxico común (p.ej. "jineta", Figura 3).

(c) En los casos en que el DRAE ofrece únicamente una forma común, pero contamos además con una forma femenina documentada suficientemente en las guías feministas y en Internet (*la piloto/pilota, la copiloto/copilota, la miembro/miembra, la timonel/timonela, la tripulante/tripulanta*...), se ofrece en primer lugar y como forma preferible la recogida en el DRAE, y como segunda opción la forma ofrecida por las guías.

(d) Cuando se trataba de agentivos terminados en -nte/-nta cuyas formas femeninas han sufrido un proceso de desprestigio y minorización (*asistenta, ayudanta*), y para los que el DRAE recoge la forma masculina como común, se ha propuesto la forma femenina en segundo lugar y la común en primer lugar, como forma preferible: *la ayudante/ayudanta, la asistente/asistenta técnica.*

Any
'jinete': 8 registros encontrados

| # | Deporte | Mujer | Hombre | Uso Esp | Woman | Man | Uso Ing |
|---|---|---|---|---|---|---|---|
| 1 | Deportes ecuestres | La amazona | El jinete | - | The rider | The rider | - |
| 2 | Deportes ecuestres | La amazona | El jinete | - | The horsewoman | The horseman | - |
| 3 | Deportes ecuestres | La jinete/jinetA | El jinete | 1 | The rider | The rider | - |
| 4 | Deportes ecuestres | La jinete/jinetA | El jinete | 2 | The woman rider | The man rider | 1 |
| 5 | Deportes ecuestres | La jinete/jinetA | El jinete | 3 | The horsewoman | The horseman | 3 |
| 6 | Deportes ecuestres | La jinete/jinetA de doma | El jinete de doma | 4 | The dressage rider | The dressage rider | - |
| 7 | Deportes ecuestres | La jinete/jinetA de salto | El jinete de salto | 5 | The show jumper | The show jumper | - |
| 8 | Deportes ecuestres | La jinete/jinetA de concurso completo | El jinete de concurso completo | 6 | The three-day event rider | The three-day event rider | - |

**Notas de uso (términos en español):**
1. *Jinete* solo en Andalucía y América.
2. *Jinete* solo en Andalucía y América.
3. *Jinete* solo en Andalucía y América.
4. *Jinete* solo en Andalucía y América.
5. *Jinete* solo en Andalucía y América.
6. *Jinete* solo en Andalucía y América.

**Notas de uso (términos en inglés):**
1. Sex-specific terms with *woman* or *man* are less and less frequent, being substituted by terms that do not specify sex (which are more comprehensive, more accurate and more consistent). Avoid the use of sex-specific terms (singular or plural) as indefinite, sex-inclusive occupational titles.
2. Sex-specific terms with *woman* or *man* are less and less frequent, being substituted by terms that do not specify sex (which are more comprehensive, more accurate and more consistent). Avoid the use of sex-specific terms (singular or plural) as indefinite, sex-inclusive occupational titles.
3. Sex-specific terms with *woman* or *man* are less and less frequent, being substituted by terms that do not specify sex (which are more comprehensive, more accurate and more consistent). Avoid the use of sex-specific terms (singular or plural) as indefinite, sex-inclusive occupational titles.

*Fig. 3. Gentyll >Deportes >'jinete'*

Any
'copiloto': 2 registros encontrados

| # | Deporte | Mujer | Hombre | Uso Esp | Woman | Man | Uso Ing |
|---|---|---|---|---|---|---|---|
| 1 | Canoa | La copiloto/copilota | El copiloto | 1 | The copilot | The copilot | - |
| 2 | Deportes de motor | La copiloto/copilota | El copiloto | 2 | The co-rider | The co-rider | - |

**Notas de uso (términos en español):**
**1.** Preferible *la copiloto*.
**2.** Preferible *la copiloto*.

*Fig. 4. Gentyll >Deportes >'copiloto'*

Any
'spotter': 3 matches found

| # | Sport | Woman | Man | Sp Usage | Woman | Man | Eng Usage |
|---|---|---|---|---|---|---|---|
| 1 | Archery | La spotter | El spotter | - | The spotter | The spotter | - |
| 2 | Archery | La observadora | El observador | - | The spotter | The spotter | - |
| 3 | Gymnastics | La ayudante/ayudanta | El ayudante | 1 | The spotter | The spotter | - |

**Usage notes (Spanish terms):**
**1.** Preferible *la ayudante*.

*Fig. 5. Gentyll >Sports > 'spotter'*

(e) Hemos considerado que en deportes el término inglés *official* («person holding public office », *Concise Oxford Dictionary*, 2008) no corresponde al grado militar. La forma femenina que ofrecemos es, por tanto, *oficiala*.

6.1.2. Si no existía documentación en el DRAE sobre forma femenina o común: NEOLOGISMOS[15]

A excepción del criterio 6.2.4., en nuestros neologismos hemos tratado de hacer compatibles las directrices de la RAE y de las guías feministas, como ya se ha dicho. Hemos recogido todas las formas encontradas en la documentación feminista, pero, cuando entraban en conflicto con la morfología que dicta la última norma académica (RAE y Asociación de Academias de la Lengua Española, 2009), hemos dado preferencia a ésta en los neologismos, dejando la forma feminista como segunda opción.

(a) Siguiendo la morfología general del español, hemos considerado comunes los términos masculinos terminados en -tra, -ista, -ta (RAE y Asociación de Academias de la Lengua Española, 2009: I, §2.5a-d, 97-99). Ejemplo: *la kayakista*.

Así mismo hemos considerado comunes los términos compuestos propios o univerbales, tanto los que responden al esquema morfológico Verbo-Nombre (*la cazatalentos, la cazagoles, la guardavalla, la pasapelotas, la recogebates, la recogeflechas, la tirapenaltis*...), como los que responden a Nombre-Nombre (*la cicloturista*) o a Adjetivo-Nombre (*la mediapunta, la mediocentro, la superclase, la todoterreno*...) (RAE y Asociación de Academias de la Lengua Española, 2009: I, §2.5ñ, 103).

(b) Cuando no aparece en el DRAE la forma femenina de un sustantivo terminado en -o, contemos o no con documentación en las guías feministas y venga o no avalada por su aparición en Internet, en general se propone la

[15] Naturalmente, no todos los términos que no aparecen en el DRAE son neologismos; puede tratarse de terminología deportiva que el DRAE no siempre recoge.

alternancia -o/-a para las formas masculina y femenina: *la alera, la acólita, la armera, la artillera, la buza, la líbera* (RAE y Asociación de Academias de la Lengua Española, 2009: I, §2.6f, 106-107), etc.

(c) Los sustantivos agudos masculinos con terminaciones en -r, -s, -n y -z se han feminizado mediante desinencia en -a. Ejemplo: *patrona* (de barco).

También forman el femenino en -a en nuestros glosarios los masculinos agudos terminados en -el, -il y -ol, cuyos femeninos toman respectivamente desinencia en -ela (como "bedela"), en -ila (recordemos que "alguacil" ha adoptado desinencia para el femenino normativamente al figurar "alguacila" en el DRAE, 23ª ed.), o en -ola (como "española"). Ejemplo: *proela.*

Las nuevas formas femeninas cuyo masculino no agudo termina en consonante -r, -s o -t se han considerado en general comunes (RAE y Asociación de Academias de la Lengua Española, 2009: I, §2.5ñ-t, 103-105). Para "líder" proporcionamos dos femeninos, *la líder y la lideresa*, ésta última forma femenina ya es recogida en el DRAE especialmente para América y en política. Por tanto, lo ofrecemos en segundo lugar: *la líder/lideresa* (en atletismo) (Figura 6).

Ady
'lideresa': 3 registros encontrados

| # | Deporte | Mujer | Hombre | Uso Esp | Woman | Man | Uso Ing |
|---|---|---|---|---|---|---|---|
| 1 | Arte | La líder lideresa de parrilla | El líder de parrilla | | The group leader | The group leader | - |
| 2 | Atletismo | La líder lideresa | El líder | | The leader | The leader | - |
| 3 | Deportes en general | La líder lideresa | El líder | | The leader | The leader | - |

**Notas de uso (términos en español):**

**1.** Preferible *la líder* en deporte; *lideresa* es un término de reciente cuño que se aplica principalmente en América y en política en España.

**2.** Preferible *la líder* en deporte; *lideresa* es un término de reciente cuño que se aplica principalmente en América y en política en España.

**3.** Preferible *la líder* en deporte; *lideresa* es un término de reciente cuño que se aplica principalmente en América y en política en España.

*Fig. 6. Gentyll >Deportes > 'lideresa'*

(d) Los agentivos terminados en -e se ofrecen en general como comunes (RAE y Asociación de Academias de la Lengua Española, 2009: I, §2.5g-l, 100-102): *la ariete, la base, la interbase, la suplente, la pivote...*, excepto cuando existe documentación en las guías feministas y en Internet de una forma femenina en -a (*cadeta, naveganta, lugartenienta...*), en cuyo caso se ofrece en primer lugar la forma común y en segundo lugar la femenina. Ejemplos:

*La cadete/cadeta*
NOTA: Preferible la primera opción

*La lugarteniente/lugartenienta*
NOTA: Preferible la primera opción

*La navegante/naveganta*
NOTA: Preferible la primera opción

(e) Los agentivos terminados en -i (tónica o átona) se ofrecen como comunes (RAE y Asociación de Academias de la Lengua Española, 2009: I, §2.5n, 102-103): *la pichichi*.

(f) USO METAFÓRICO DE SUSTANTIVOS Y DE LOCUCIONES. Hemos convertido en comunes los términos metonímicos, tenga el sustantivo original forma masculina o femenina: *la cerrojo* (en fútbol), *la cerebro* (del equipo), *la comodín, la motor* (del juego), *la uno, la ocho, la cinco*, etc., incluso cuando se trata de una locución: *la cabeza de carrera, la cabeza de serie, la peso mosca, la cinturón negro, la maillot amarillo, la número uno*... (RAE y Asociación de Academias de la Lengua Española, 2009: I, §2.6j, 108; 2.7h, 112; 11.2i, 742). Las excepciones son: *el as, el astro, el descarte*.

Existen en varios deportes denominaciones metonímicas para jugadores o jugadoras que se refieren a sus posiciones en el campo. Unas veces son sustantivos simples o compuestos, otras veces son locuciones nominales: *punta, exterior, interior, lateral, centro, medio, mediocentro, mediocampo, pilar, poste, poste alto, poste bajo, poste medio* y *extremo*. Hemos considerado tales términos comunes[16]. Siguiendo la norma, todas esas formas para mujeres (sean términos simples, compuestos o locuciones) se acompañan de adjetivo en femenino: *la medio derecha, la lateral izquierda, la extremo izquierda, la mediocentro* polaca*, la poste alto* española...

(g) EXTRANJERISMOS. Las nuevas voces extranjeras y los nuevos extranjerismos masculinos se han convertido en comunes: *la crack*[17]*, la pívot*, *la catcher* (en béisbol), *la hooligan*...[18] (A veces también extranjerismos femeninos: *el cheerleader*). Muchas de tales formas, pero no todas, pertenecen a deportes de lucha, como karate o judo. Ejemplos: *la uke, la tori, la sung, la fukushin, la chong, la hong, la shushin, la pushim*...

(h) NOMBRES EPICENOS. Hemos mantenido como epiceno el sustantivo femenino para el corredor o corredora que en las pruebas de larga distancia se pone en cabeza para imponer un ritmo determinado al resto de participantes: *la liebre*. (RAE y Asociación de Academias de la Lengua Española, 2009: I, §2.7d, 110-111)

---

[16] Pese a estar ampliamente documentada la forma femenina "extrema", especialmente en América.

[17] Aunque la RAE y la Asociación de Academias de la lengua Española (2005) sugieren "crac", los medios mantienen la grafía inglesa.

[18] Se me escapa la razón por la que la RAE ha decidido no acentuar *hooligan*, pero sí *pívot*; como tampoco es fácil entender por qué se ha españolizado *esprinter*, pero no *sparring*.

**6.2. Sustantivos compuestos sintagmáticos o locuciones nominales**

6.2.1. Estructura: Nombre + Adjetivo.

A veces los agentivos son grupos nominales compuestos de sustantivo + adjetivo. En tales casos, para el sustantivo hemos seguido las directrices detalladas en 6.1.2. Además, hemos identificado la clase de adjetivo del que se trata y, conforme a su desinencia, aplicado la regla morfológica pertinente (Harris, 1991: 34-36): *la árbitra auxiliar; la pivote defensiva; la surfera alpina; la asesora técnica....*

6.2.2. Estructura: Nombre + Nombre

Otras veces el acompañante del núcleo del grupo nominal no es un adjetivo, sino otro sustantivo en aposición, como *delantera centro*. Es en este grupo donde se contaba con menos documentación en las guías feministas y en Internet.

Consideramos que el DRAE (2001, 2014) no siempre ha sido consistente en los sustantivos en aposición con su política formulada en la *Nueva Gramática de la Lengua Española* (2009). Un ejemplo sería el caso que hemos mencionado en el apartado dedicado al cuarto factor sociolingüístico tomado en consideración: *juez árbitro*. En nuestros glosarios hemos tratado de mantener la consistencia y, por ello, normalmente se proporcionan dos formas femeninas, tanto para el núcleo nominal como para la aposición: *la directora médica, la jueza cronometradora, la jugadora líbera, la jueza oficiala, la oficiala jefa...*, a no ser que se trate de la aposición metonímica mencionada en 6.1.2. (f): *la delantera centro*.

En casos en los que el DRAE proporciona dos formas para mujeres, una común y una femenina, la forma femenina ha primado en nuestros glosarios sobre la forma común (*la comisaria jefa, la árbitra-jueza*), aunque se proporciona la forma común como alternativa:

> *La comisaria jefa*
> NOTA: También *la comisaria jefe*
>
> *La entrenadora jefa*
> NOTA: También *la entrenadora jefe*
>
> *La jueza árbitra*
> NOTA: También *la juez árbitra*
>
> *La medidora jefa*
> NOTA: También *la medidora jefe*
>
> *La árbitra jueza*
> NOTA: También *la árbitra juez* (Figura 7)

Any
'arbitra jueza': 5 registros encontrados

| # | Deporte | Mujer | Hombre | Uso Esp | Woman | Man | Uso Ing |
|---|---|---|---|---|---|---|---|
| 1 | Boxeo | La árbitra jueza | El árbitro juez | 1 | The boxing official | The boxing official | - |
| 2 | Taekwondo | La árbitra-jueza | El árbitro-juez | 2 | The judge | The judge | - |
| 3 | Taekwondo | La árbitra-jueza | El árbitro-juez | 3 | The pushim | The pushim | - |
| 4 | Taekwondo | La árbitra-jueza | El árbitro-juez | 4 | The assistant umpire | The assistant umpire | - |
| 5 | Taekwondo | La árbitra-jueza | El árbitro-juez | 5 | The referee | The referee | - |

**Notas de uso (términos en español):**
1. También *la árbitra juez*.
2. También *la árbitra-juez*.
3. También *la árbitra-juez*.
4. También *la árbitra-juez*.
5. También *la árbitra-juez*.

*Fig. 7. Gentyll >Deportes >'árbitra jueza'*

La excepción la constituyen las formas en -nte que aparecen en los agentivos como modificadores del núcleo. La mayoría de ellas podrían considerarse un complemento adjetival y no un sustantivo en aposición: *la árbitra saliente, la gimnasta suplente, la jugadora atacante, la jueza entrante*. Tanto si se consideran locuciones, como compuestos de los tipos (N+Adj.) o (N+N), solo se ofrece la forma común -nte y no la feminizada -nta, cuando la hay: *la árbitra ayudante, la árbitra asistente, la directora gerente, la anotadora asistente*.

6.2.3. Formas compuestas o locuciones que incluyen la palabra "hombre"

Las formas compuestas o las locuciones que incluyen la palabra "hombre" han hecho el femenino cambiando "hombre" por "mujer": *mujer rana, mujer-anuncio, la sexta mujer* (en baloncesto).

6.2.4. El masculino como genérico no marcado

El aspecto que puede resultar más controvertido de nuestros glosarios es que no hemos utilizado el masculino como genérico no marcado. Por tanto, para los agentivos que llevan complemento preposicional en el que podría haber figurado un masculino como genérico, se ofrecen dos formas, a aplicar según el sexo de la(s) persona(s) de la(s) que se trate: *la adjunta del juez/ de la juez de salida* (Figura 8). Si se desconoce el sexo o se trata de un término genérico, se ha utilizado la doble forma (masculina y femenina o femenina y masculina: *la jefa de medidores y medidoras*) o, cuando ha sido posible, un sustantivo abstracto (*la jefa de árbitras y árbitros* o *la jefa de arbitraje*).

Es patente en este punto nuestra divergencia con la RAE y la Asociación de Academias de la Lengua Española (2005: 311; 2009: I, §2.2, 85-89), pero se

trataba de un aspecto que forma parte de las políticas lingüísticas no sexistas que tratamos de recoger.

Any
'adjunta de la juez': 1 registros encontrados

| # | Deporte | Mujer | Hombre | Uso Esp | Woman | Man | Uso Ing |
|---|---|---|---|---|---|---|---|
| 1 | Atletismo | La adjunta de la jueza del juez de salida | El adjunto de la jueza del juez de salida | | The recaller | The recaller | |

**Notas de uso (términos en español):**
1. También *la adjunta de la juez*.

*Fig. 8. Gentyll >Deportes >'adjunta de la juez'*

Como nota final debemos decir que para dejar claro el género gramatical y para resolver la posible duda de quienes consulten estos glosarios, los sustantivos aparecen acompañados del artículo determinado singular (*la/el, the*). Hemos mantenido el artículo femenino en sustantivos que empiezan por a- tónica, al igual que en "la árbitra" (RAE y Asociación de Academias de la Lengua Española, 2009: I, §14.2p, 1036): *la ala derecha* (en balonmano).

En todo caso, somos plenamente conscientes de que nuestros glosarios pueden contener errores e inconsistencias. Agradeceríamos la notificación de ellos o cualquier sugerencia a través de la dirección de correo electrónica de contacto, si tienen interés en visitar nuestra página. GentyllWeb@gmail.com

## Agradecimientos

Este trabajo forma parte del proyecto de investigación FEM2009-10976, financiado por el Ministerio español de Ciencia e Innovación. Agradecemos así mismo la beca de movilidad de Caja Madrid.

## Referencias

ALFARO É., BENGOECHEA M., VÁZQUEZ B. (2010), *Hablamos de Deporte,* Madrid, Instituto de la Mujer.

BENGOECHEA, M. (2008), "Textualización de una identidad genérica femenina homogeneizada: rasgos esencialistas en el Diccionario de la RAE", In DAVIES A., KUMARASWAMI P., WILLIAMS C. (dirs.), *Making Waves Anniversary Volume: Women in Spanish, Portuguese and Latin American Studies*, 71-91, Cambridge, Cambridge Scholar Publishing.

BENGOECHEA M. (2011), "Non-sexist language policies of Spanish: An attempt bound to fail?", *Current Issues in Language Planning, Vol.* 12 (1), 25-43.

BENGOECHEA M., CABELLOS M. R. (2012), "Underlying principles of Gentyll English-Spanish non sexist glossaries: A response to a need", In VATVEDT FJELD R., TORJUSEN J. M. (dirs.), *Proceedings of the 15th*

*EURALEX International Congress. European Association for Lexicography*, *EURALEX OSLO 2012*, 1003-1007. http://www.euralex.org/publications/

BENGOECHEA M., CABELLOS M. R., SIMÓN J. (2013), "*Gentyll* on-line glossaries: Professional titles of women and men in a series of fields of activity", Ponencia presentada al *III International Seminar on English and ESP Lexicology and Lexicography*, noviembre 2011 (en prensa)

BOWKER L. (2001), "Terminology and gender sensitivity: A corpus-based study of the LSP of infertility", *Language in Society*, Vol. 30, 589-610.

CASARES J. (1961), "Catedrática", *Cosas del Decir*, 191-193, Madrid, Espasa-Calpe.

CENTENERA SÁNCHEZ-SECO F. (2011a), "La igualdad entre mujer y hombre en el lenguaje normativo*: el estudio de un proceso*", In TAILLEFER DE HAYA L. (dirs.), *Igualdad: de lo legal a lo real,* 111-148, Málaga, Diputación de Málaga.

CENTENERA SÁNCHEZ-SECO F. (2011b), "El alcance práctico del lenguaje no sexista solicitado desde el plano normativo: ¿una cuestión resuelta?", *Cuadernos Electrónicos de Filosofía del Derecho*, Vol. 24. http://ojs.uv.es/index.php/CEFD/article/view/1622

CONSEJO DE EUROPA (1986), *Igualdad de sexos en el lenguaje,* Bruselas, Comisión de terminología en el Comité para la igualdad entre mujeres y hombres. Mayo 1986.

COOPER R. L. (1989), *Language Planning and Social Change,* Cambridge, Cambridge University Press.

DOCHERTY V. J. (2007), "Dictionaries on the Internet: an Overview", *Proceedings of the EURALEX Conference, 2007*, 67-74. http://www.euralex.org/elx_proceedings/Euralex2000/007_Vincent%20J.%20DOCHERTY_Ditionaries%20on%20the%20Internet%20an%20Overview.pdf

FASOLD R. (1988), "Language policy and change: Sexist language in the periodical news media", In LOWENBERG P. H. (dir.), *Language Spread and Language Policy: Issues, Implications, and Case Studies* (GURT '87), 187–206, Washington, DC, Georgetown University Press.

FASOLD R, YAMADA H., ROBINSON D., BARISH S. (1990), "The language-planning effect of newspaper editorial policy: Gender differences in *The Washington Post"*, *Language in Society*, Vol. 19, 521-539.

FORGAS BERDET E. (1986), "Sexo y sociedad en el último *DRAE*", *Universitas Tarraconensis,* Vol. 10, 79-100.

FORGAS BERDET E. (2004), "La representación femenina en las etimologías", In LLEDÓ CUNILL E. (dir.), *De mujeres y diccionarios. Evolución de lo femenino en la 22ª edición del DRAE,* 421-444, Madrid, Instituto de la Mujer.

FRANK F. W., TREICHLER P. A. (1989), *Language, Writing and Professional Writing: theoretical Approaches and Guidelines for Nonsexist Usage*, New York, MLA.

GERSHUNY H. L. (1977), "Sexism in Dictionaries and Texts: Omissions and Commissions", In NILSEN A. P., BOSMAJIAN H., GERSHUNY H.L. (Dirs.), *Sexism and Language,* 143-160, Urbana, Illinois, National Council of Teachers of English.

GÓMEZ TORREGO L. (2010), "Aspectos semánticos del lenguaje del fútbol en España", *Marco ELE, Revista de Didáctica Español como Lengua Extrajera*, Vol. 11, 150-158.

GUERRERO SALAZAR S. (2007), "Esbozo de una bibliografía crítica sobre recomendaciones y guías para un uso igualitario del lenguaje administrativo", In MEDINA GUERRA A. M. (dir.), *Avanzando hacia la igualdad*, 109-122, Málaga, Diputación de Málaga.

HARRIS J. W. (1991), "The exponence of gender in Spanish", *Linguistic Inquiry*, Vol. 22(1), 27-62.

HELLINGER M., BUßMANN H. (dirs.) (2001-2003), *Gender across Languages: The Linguistic Representation of Women and Men,* Amsterdam, John Benjamins.

INSTITUTO CERVANTES (2011), *Guía de Comunicación No Sexista,* Madrid, Instituto Cervantes.

KENNEDY C. (1996), "La Crême de la Crême": Coercion and corpus change – an example from recruitment advertisements", In COLEMAN H., CAMERON L. (dirs.), *Change and Language*, 28-38, Clevedon, Multilingual Matters.

KILGARRIFF A., GREFENSTETTE G. (2003), "Introduction to the special issue on the Web as corpus", C*omputational Linguistics*, Vol. 29(3), 1-15.

KRAMARAE C. (1992), "Punctuating the dictionary", *International Journal of the Sociology of Language*, Vol. 94, 135-54.

LLEDÓ CUNILL E. (dir.) (2004), *De mujeres y diccionarios. Evolución de lo femenino en la 22ª edición del DRAE,* Madrid, Instituto de la Mujer.

LLEDÓ CUNILL E. (2006 [1995]), *Profesiones en femenino. De la A a la Z*, Madrid, Instituto de la Mujer, 2ª edición.

LLITERAS M., SASTRE Mª Á., CARRERA R., APARECIDO LOPES C. (2004), *Género sin dudas en el ámbito profesional*, Valladolid, Junta de Castilla y León.

LUDELING A., KYTO M. (dirs.) (2008), *Corpus Linguistics. An International Handbook*, Berlin, Mouton de Gruyter.

MAGGIO R. (1997), *Talking about People: A Guide to Fair and Accurate Language*, Phoenix, Oryx.

MEDINA GUERRA M. A. (2004), "El diccionario de la Academia y el léxico de los oficios, profesiones y cargos de responsabilidad desde la perspectiva de género", In MARTÍNEZ GARCÍA A. *(*dir.*), Cultura, lenguaje y traducción desde una perspectiva de género*, 115-134, Málaga, Universidad de Málaga.

MILLER C., SWIFT K. (1995), *The Handbook of Non-Sexist Writing*, London, The Women's Press, 3rd edition.

MLA (2003), *Handbook for Writers of Research Papers* (6th ed.), Joseph Gibaldi, ed., New York, Modern Language Association of America.

NISSEN U. K. (2002), "Gender in Spanish. Tradition and innovation", In HELLINGER M., BUßMANN H. (dirs.), *Gender across Languages: The Linguistic Representation of Women and Men,* Vol. 2, 251-280, Amsterdam, John Benjamins.

NOMBRA (Comisión Asesora sobre Lenguaje del Instituto de la Mujer) (1994), *Nombra en femenino y en masculino,* Madrid, Instituto de la Mujer.

ORDOÑO MUÑOZ A. (2004), *32 léxicos trilingües de deportes y medios de comunicación para los Juegos Olímpicos y Mediterráneos*, 3 volumes, Almería, Asociación de Prensa de Almería/TCI.

PARLAMENTO EUROPEO (2003a), "Sobre las mujeres y el deporte", *Resolución A5-00167/2003 del 21 de mayo*, Comisión de derechos de la mujer e igualdad de oportunidades.

PARLAMENTO EUROPEO (2003b), Resolución del Parlamento Europeo sobre las mujeres y el deporte, (2002/2280(INI)). http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+TA+P5-TA-2003-0269+0+DOC+XML+V0//ES

PAUWELS A. (1997a), "Of handymen and waitpersons: A linguistic evaluation of job classifieds", *Australian Journal of Communication*, Vol. 24(1), 58-69.

PAUWELS A. (1997b), "Of 'Mses', 'Chairmen' and 'Retromen: Linguistic change in the media and the community", *Australian Language Matters*, April/May/June, 3.

PAUWELS A. (1998), *Women Changing Language,* London, Longman.

PAUWELS A. (2003), "Linguistic Sexism and Feminist Linguistic Activism", In HOLMES J., MEYERHOFF M. (dirs.), *The Handbook of Language and Gender*, 550-570, Oxford, Blackwell.

RAE (Real Academia Española) (1992), *Diccionario de la Lengua Española* (DRAE), Madrid, Espasa Calpe, 21ª edición.

RAE (Real Academia Española) (2001), *Diccionario de la Lengua Española* (DRAE) Madrid, Espasa Calpe, 22ª edición.

RAE (Real Academia Española) (2014), *Diccionario de la Lengua Española* (DRAE), Madrid, Espasa Calpe, 23ª edición (en prensa, disponible parcialmente en la web de la RAE http://www.rae.es/rae.htm).

RAE (Real Academia Española) Y ASOCIACIÓN DE ACADEMIAS DE LA LENGUA ESPAÑOLA (2005), *Diccionario Panhispánico de Dudas*, Madrid, Santillana.

RAE (Real Academia Española) Y ASOCIACIÓN DE ACADEMIAS DE LA LENGUA ESPAÑOLA (2009), *Nueva Gramática de la Lengua Española.* Tomo I*: Morfología y Sintaxis,* Madrid, Espasa Calpe.

ROMERO GRANADOS S. (2008), "Influencia del patrocinio y de los medios de comunicación en la discriminación del deporte de élite femenino", *Revista Tandem. Didáctica del Deporte*, Vol. 28, julio, 85-95.

RUBIN D., GREENE K., SCHNEIDER D. (1994), "Adopting gender-inclusive language reforms", *Journal of Language and Social Psychology*, Vol. 13 (2), 91-114.

SCHOLZE-STUBENRECHT W. (2001), *Duden Fremdwörterbuch*, Duden-Verlag.

SCHWARTZ M. (1995), *Guidelines for Bias-Free Writing*, Bloomington, Indiana UP.

TERMCAT. http://www.termcat.cat/es/Diccionaris_En_Linia

UNESCO (1990), *Recomendaciones para un uso no sexista del lenguaje*, París, Unesco.

UNITED STATES DEPARTMENT OF LABOR (2010-2011), *Occupational Outlook Handbook* 2010-2011. http://www.bls.gov/oco/

VARGAS A., LLEDÓ CUNILL E., BENGOECHEA M., MEDIAVILLA M., RUBIO I., MARCO A., ALARIO C. (1998), *Lo femenino y lo masculino en el Diccionario de la Real Academia Española*, Madrid, Instituto de la Mujer.

VELASCO M. L., VÁZQUEZ D., IBÁÑEZ M. (2009), *El cambio lingüístico en la educación en los últimos años en España*, Madrid, Instituto de la Mujer.

# Partie 3

## Approches diachroniques

# Les néologismes techniques dans la traduction florentine de 1464 de l'*Opus agriculturae* de Rutilius Aemilianus Palladius

Moreno Campetella[1]

## Résumé

La traduction florentine de l'*Opus Agriculturae* de Rutilius Aemilianus Palladius (Ve siècle), achevée en novembre 1464, transmise par le *ms. Asburnensis 524* de la Bibliothèque Laurenziana de Florence, constitue l'une des tentatives, assez nombreuses au XVe siècle en Italie, de version « vulgaire » d'un texte technique de l'antiquité latine tardive.

Il s'agit de la première traduction italienne de l'œuvre de Palladius, qui figure parmi les plus précoces témoignages européens de l'importance attachée par les savants de la Renaissance à la divulgation des anciens traités agronomiques, aussi bien parmi les professionnels du secteur (fermiers et grands propriétaires fonciers) qu'au sein de quelques cercles culturels restreints, comme celui de la cour des Médicis. Ce texte montre bien, à travers les nombreux néologismes lexicaux et sémantiques qui le caractérisent, le processus de formation de la terminologie technique agricole italique, encore bien hésitante en ce milieu du *Quattrocento*, et, plus généralement, de la future « langue littéraire italienne commune ».

Parmi les nouveautés lexicales observées ici, qui sont souvent le résultat d'une grande littéralité propre aux traductions italiennes du XVe siècle, citons les adjectifs *squallido* (« misérable » dans l'italien moderne) dans le sens de « sec », désignant un terrain aride (*harenae squalentes* dans le texte latin), *scalabroso* pour indiquer l'herbe portant des traces de sécheresse (*scabra herba* dans le texte latin) ou encore le verbe *palare*, désignant l'opération qui consiste à installer des poteaux de soutien pour les plants de vigne.

Les néologismes sont tout aussi répandus parmi les noms de plantes dont les descendants modernes sont attestés uniquement en tant que régionalismes : c'est le cas de *porcellana*, « pourpier », de *semprevivo*, « joubarbe » ou encore de *duràcino*, terme technique désignant plusieurs espèces de fruits (pêches et prunes surtout) à noyau adhérent.

[1] Institut Pierre Gardette - Université Catholique de Lyon / CRTT – Centre de Recherche en Terminologie et Traduction (EA 4162) - Université Lumière Lyon2, mcampetella@yahoo.fr

**Mots clés**

Agronomie (Italie, Europe Occidentale, XVe – XVIIe siècles) – Italien (XVe-XVIIIe siècles) - Palladius (Rutilius Aemilianus) - Traduction (Italie, XVe siècle).

## Introduction

L'italien du XVe siècle, la langue technique en particulier, doit beaucoup aux textes de la littérature latine, classique et tardive, que les humanistes traduisent massivement[2]. L'idiome vernaculaire utilisé par les humanistes de la Péninsule pendant les premières décennies du XVe siècle nous apparaît comme une simple transposition en vulgaire des lexèmes latins, dictée par une littéralité appliquée à l'opération de traduction qui était la règle à l'époque.

Un exemple nous est fourni par la traduction anonyme, conservée par le *ms. Ashburnensis 524* de la Bibliothèque Laurentienne de Florence, de l'*Opus agriculturae* de Rutilius Taurus Aemilianus Palladius (milieu du Ve siècle de notre ère)[3], traduction achevée en 1464. La traduction de l'*Opus* fut très vraisemblablement commanditée par la cour Médicis de Florence. En effet, le milieu humaniste florentin de l'époque passait pour le plus à la pointe de la péninsule dans le domaine agronomique et tous les Médicis, à commencer par le fondateur de la dynastie, Cosme (1398-1464), et son successeur Laurent le Magnifique, étaient très passionnés d'économie rurale.

L'importance attachée aux textes agronomiques dans la Toscane ducale du XVe siècle est confirmée aussi par l'antériorité des premières traductions italiennes de Palladius, toutes florentines, par rapport à celles qui virent le jour dans d'autres pays d'Europe comme la France ou l'Angleterre. En effet, la première édition française semblerait être celle de Jean Darces [4] (*Les treize livres des choses rustiques de Palladius*, Paris, 1554). En Angleterre, une certaine précocité des traductions[5] semble avoir été contrebalancée par un vide de quelques siècles[6],

---

[2] Voir à ce propos Dionisotti, 1967 ; Folena, 1973 ; Van Hoof, 1991 ; Ballard, 1992 ; Giovanardi, 1994 ; Orvieto, 1996.

[3] Svennung, 1928 ; Martin, 1976 : VII-XX.

[4] A ma connaissance, il n'existerait aucun témoin manuscrit rédigé en français classique. Parmi les traductions modernes qui ont précédé l'édition de René Martin (1976) citons celle de Saboureux de la Bonnetrie, *Traductions d'anciens ouvrages latins relatifs à l'agriculture*, tome V, Paris, Didot, 1775, revue et modifiée un siècle plus tard par un latiniste anonyme (*Les agronomes latins*, sous la direction de M. Nisard, Paris, Didot, 1864), et celle de M. Cabaret-Dupaty, *L'économie rurale de Palladius*, collection Panckoucke, Paris, 1843.

[5] *Bibl. Bodl. - Ms. Duke Humfrey d. 2* : il s'agit d'une traduction versifiée (attribuée à tort à Thomas Norton), écrite pour Humfrey, duc de Gloucester (1390-1447) entre 1439 et 1447, très vraisemblablement en 1442-43.

[6] La première édition serait due à Barton Ladge et daterait de 1873 (d'après un manuscrit latin de 1420).

alors qu'en Italie, comme on pouvait s'y attendre [7], elles ont vu le jour très tôt et en plus grand nombre.

## 1. Néologismes lexicaux

L'influence lexicale exercée par le modèle latin sur la version vernaculaire est évidente. Les nombreuses traductions mot à mot effectuées de façon anonyme constituent autant de néologismes, lexicaux ou sémantiques. Outre les adjectifs *scalabroso* (< lat. *scaber*) désignant l'herbe portant les marques de la sécheresse, *digiuno* (< lat. *ieiunus*) qui caractérise la nature « sèche » du tuf ou du gravier, ou encore *squalido* (< lat. *squalens*), employé pour indiquer un terrain infertile parce que peu humide[8], examinés dans une étude précédente, les technicismes, qu'il s'agisse de syntagmes ou de vocables, sont particulièrement abondants.

D'intéressants exemples d'emprunt technique sont le verbe *palare* et l'adjectif *soperchio*, pour désigner respectivement l'opération consistant à installer des poteaux de soutien pour les plants de vignes et la branche d'un arbre qui pousse trop et qui dépasse sa longueur normale :

> *In eisdem locis* [scil.: *maritimis et calidis* ] *palandae aut ligandae sunt vineae, priusquam gemma procedat, cuius concussione vel adtritu incurrit grande dispendium (…) Tempore hoc, si quae sunt in seminariis plantae, circumfodiendae sunt et amputandi eis rami superflui vel radiculae* (*Op. agr.* III 20) = « In quegli medesimi luoghi est da palare o sono da legare le viti prima che la gemma esca fuori per lo percotimento o istropicciamento de la quale incorre gran danno (...) In questo tempo sono da potare e' piantoni se alcuni ne sono negli semenzai et sono da tagliare loro e' rami superchi o radicecte[9] » (*Ashb.* 524, f° 48r)

Ce passage vernaculaire de l'*Opus agriculturae* représente le premier témoignage du verbe *palare* en italien. Le sémantème sera repris par les agronomes des siècles suivants comme Daniele Ricciarelli (première moitié du XVI^e^ siècle), Filippo Baldinucci (1625-1696) ou encore Benedetto del Bene (fin du XVIII^e^

---

[7] L'auteur de l'*Opus Agriculturae* nous apprend qu'il possédait des domaines en Italie (surtout dans les environs de Rome et en Sardaigne), ce qui laisserait penser que c'était un habitant de souche de la péninsule. D'autres éléments encore appuieraient cette hypothèse (Martin, 1976 : VII-XIII). Cela pourrait expliquer la plus grande diffusion des versions vernaculaires du traité en Italie que partout ailleurs en Europe.

[8] *Op. agr.* I 5, 1-2, *Ashb. 524*, f° 5r. Pour l'analyse de ces trois adjectifs voir Campetella, 2009 : 92.

[9] « Dans ces mêmes lieux les vignes doivent pouvoir s'appuyer à des poteaux et il faut les lier avant que les bourgeons n'apparaissent, parce que si ces derniers se heurtaient ou frottaient les uns contre les autres, grands seraient les dégâts qui en résulteraient. »
Les traductions françaises des livres I-II de l'*Opus agriculturae* sont tirées de l'édition critique de René Martin (Martin, 1976), celles des autres livres ont été effectuées par l'auteur de cet article.

siècle) dans sa traduction du *Traité d'Agriculture* de Columelle[10]. L'adjectif *soperchio* a laissé les mêmes traces durables dans la littérature agronomique des décennies suivantes : on le retrouve dans le *Trattato della coltivazione delle viti* de Giovanvettorio Soderini (1526-1597), ou encore chez le susmentionné Benedetto del Bene[11].

Les noms de plantes dont la version vernaculaire de 1464 constitue la première attestation dans toute l'aire linguistique romane sont également nombreux : c'est le cas d'*ervo*[12] (< lat. *ervus*), qui constitue un hapax en italien (aujourd'hui le nom italien de cette plante est *rubiglia*), alors que la forme dérivée d'*ervus* est restée au niveau dialectal[13], ainsi qu'en français (*ers*)[14], en espagnol (*yero*) et en catalan (*er*), de *porcellana*[15], « pourpier » (< lat. *portulaca*, avec métathèse consonantique), ou encore de *semprevivo*, « joubarbe »[16], qu'on retrouve seulement, en dehors du passage de l'*Opus agriculturae*, dans la traduction de l'*Histoire Naturelle* de Pline de Cristoforo Landino, publiée à Venise en 1534. Dans le reste de l'aire linguistique romane *sempervive* réapparaît chez Ronsard et, au niveau oral, en provençal[17]. Digne d'intérêt est aussi *duràcino*, terme technique désignant plusieurs espèces de fruits (pêches et prunes surtout) à noyau adhérent : « *Etiam tunc tuberes seruntur et inseruntur et ossa duracinorum* » (*Op.*

---

[10] *GDLI* XII : 383.

[11] *GDLI* XX : 539.

[12] *Fetus frequentant, si hordeum torrefactum vel fabam vel ervum saepe consumant* (*Op. Agr.* I 24, 3) = « Figliano [scil. : i colombi] spesso se orzo arostito o fava mangiano overo ervo » (*Ashb. 524* f° 16r) (« Ils [scil. : les pigeons ramiers] pondent un grand nombre d'œufs, si on leur donne souvent à manger de l'orge grillée, des fèves ou de l'ers »).

[13] Tosc. *lero*, avec agglutination de l'article (*REW* 2910).

[14] La première attestation date de 1538.

[15] *Oculos portulacae suco forinsecus et mulieris lacte curemus…* (*Op. Agr.* I 27, 3) = « Gli occhi [delle galline] medicano con sugo di porcellane et lacte di femina (*Ashb. 524* f° 1r) (« Il faut en soigner [scil. : des poules] les yeux en les baignant extérieurement avec du jus de pourpier et du lait de femme »).
Le seul autre témoignage écrit du vocable en italien date de 1569, plus d'un siècle après le passage de l'*Opus*. Il s'agit d'une note du journal de voyage d'Alfonso Ulloa. Dans l'aire linguistique romane, seul un document français de 1544 (*pourceline*) et le *Dictionnaire* de Furetière de 1690 (*pourcelaine*) ont conservé la même forme attestée par la traduction florentine de 1464, alors qu'en ancien français est attestée la forme *pourchaille* (*FEW* IX : 226 ; *REW* 6679), qui reflète l'orthographe du mot latin, tout comme la totalité des dialectes gallo-romans (ex. gasc. *bourdoulaigue*, prov. *bortolaiga* etc. – voir *REW* 6679) et le français standard, où le nom de la plante est aujourd'hui *pourpier*. L'ancien italien et ses dialectes ont également conservé la graphie latine : a.it. *porcacchia* ou *procacchia* (*REW* 6679).

[16] *Contra erucas, semina quae spargenda sunt sempervivi suco madefiant* (*Op. Agr.* I 35, 3) = « Contro a bruchi e' semi che si seminano si bagni con sugo di semprevivo » (*Ashb. 524* f° 21v) (« Contre les chenilles il faut tremper dans du jus de joubarbe les graines destinées à l'ensemencement »).

[17] *FEW* XI : 444.

*Agr.* III 25, 32), « Ora si pongono e' tuberi et nestansi et l'ossa di duràcino[18]» (*Ashb. 524* f° 56r).

Bien que son aire de diffusion soit presque exclusivement italo-romane[19], ce qui laisserait supposer une origine locale et populaire du terme technique, *duràcino* reste assez productif même en dehors de cette zone : dans une source provençale de 1460 *durayquier* signifie déjà une « espèce de prunier » (dont les fruits ont vraisemblablement un noyau adhérent), alors que dans un autre document de 1562 *duraine* – tout comme la *duracine* du *Dictionnaire* de Cotgrave de 1611 – désigne une « cerise à chair ferme[20] ».

D'autres néologismes palladiens n'ont pas laissé de trace dans la littérature italienne postérieure au point qu'on pourrait, encore une fois, se demander s'il s'agit de vrais néologismes (de vrais hapax) ou bien s'ils sont imputables à une incompréhension de l'original latin de la part du traducteur. Ainsi en est-il de l'*ulpico*[21] (« ciboulette ») qu'on retrouve tout de même dans le *Dictionnaire* de Cotgrave (1611 – *ulpic*[22]), de l'adjectif *uliginoso*[23] (« très humide », « imbibé d'eau », utilisé pour désigner presque exclusivement un certain type de terre) qui a donné *uligineux* chez Rabelais (1546) et dans le *Dictionnaire* de Cotgrave de 1611 (*FEW s .v.*), ou encore du mot *cavolo*[24] (< lat. *caulis*), qui n'indique pas ici

---

[18] « A cette époque de l'année on met en terre les tubercules, on procède aux greffes et à la plantation de tous les fruits à noyau adhérent – cette dernière opération s'effectue à partir de ces mêmes noyaux. »
Voir aussi *persica duracina* (*Op. Agr.* XII 7, 4) = « duràcine pesche » (*Ashb. 524* f° 118r), « pêche à noyau adhérent, pêche Pavie ».

[19] Vegl. *drukno*, it. *ciliegia / uva duracina*, venez. *durazega*, romagn. *duron*, gen. *duaza*, lomb. *düraš*, sic. *duraka* etc. (*REW* 2803).

[20] *FEW* III : 187.

[21] Hoc etiam mense ulpicum bene et allium seremus et cepullas (…). Lactuca et beta et porrus et cappar seri possunt et satureia et nasturtium. Intyba etiam et raphanos nunc aliqui serunt (Op. Agr. IV 9, 5) = « Questo mese porremo bene l'ulpico et l'aglio et cipolle (...). Lattuga, bietola, porro et capperi si possono seminare et santoreggia et agrecto. Alcuni seminano intiba et raphano e' quali usano la state » (Ashb. 524 f° 63r) (« Ce mois-ci on sèmera la ciboulette, l'ail, les oignons (...) la laitue, les bettes, les poireaux, les câpres, la sarriette et le cresson alénois. Il y en a qui sèment aussi les endives et le raifort »).

[22] *FEW s.v.*

[23] *Omne triticum solo uliginoso post tertiam sationem in genus siliginis commutatur* (*Op. Agr.* I 6, 6) = « Ogni grano in terra non humidissima dopo la terza sementa si muta in segale » (*Ashb. 524*, f° 7r) (« Tout froment semé trois fois dans un sol humide se change en une sorte de siligo»).

[24] « Pour que les graines des raves puissent se développer, une fois la plante extraite du terrain, il faudra enlever toutes les feuilles et pratiquer sur la tige une incision d'un "demi doigt" de long ».

le chou, comme c'est le cas en italien standard, mais la tige, plus ou moins charnue, de la rave[25] :

> *Ut vero semina maiora redigantur, erupta rapa foliis omnibus purgabis et ad dimidii digiti crassitudinem in caule succides* (*Op. Agr.* VIII 2, 3) = « Acciò che gli semi diventino maggiori, divelta la rapa, togline via tucte le foglie et taglierai a grossezza di mezo dito nel cavolo suo » (*Ashb. 524* f° 92v)

Les lexèmes examinés jusqu'ici, pures translittérations dans la plupart des cas des termes latins, montrent une adhérence totale du traducteur au texte source, ce qui pourrait laisser supposer une incompréhension de l'original qu'il a sous les yeux et, par conséquent, une connaissance imparfaite de la science agronomique. D'autres néologismes mettent en évidence, au contraire, la parfaite maîtrise de la matière qu'il est en train d'aborder : dans cette autre catégorie de néologismes, les vocables de l'original latin ont été, en effet, remplacés par des technicismes appartenant vraisemblablement au basilecte agricole – les passages de la traduction florentine constituent les seuls témoignages de leur existence – et le traducteur a même ajouté à la traduction, souvent littérale, du lexème du texte source un vocable dont l'existence devait être presque exclusivement orale, nature orale qui devait rendre plus intelligible la totalité du passage. Cela serait en ligne avec le statut des humanistes de la cour princière des Médicis de la première génération (1380-1470) : ces hommes de lettres sont en même temps des fonctionnaires bien avertis et doués d'une conception particulièrement pratique de la gestion des affaires politiques et financières de la république de Florence. En outre, dans la plupart des cas, une bonne gestion de la *res publica* devait coïncider avec une administration d'autant plus performante de son propre patrimoine, majoritairement constitué de propriétés foncières.

Un exemple en ce sens nous est fourni par l'hapax verbal *saectolare* – « tailler (la vigne) », « ébourgeonner » – qui, rajouté à la traduction du passage latin, constitue une explication technique supplémentaire, vraisemblablement à l'usage des professionnels :

> *Erit tamen optimi putatoris inferius sarmentum quod bono loco natum fuerit, reparandae vitis causa semper tueri et ad unam vel duas gemmas relinquere* (*Op. Agr.* III 12, 2) = « Sarà optimo s'el potatore el tralcio disocto el quale sarà bene nato per cagione di racconciare la vite sempre guardarlo et lasciarlo ad una o due gemme: questo si chiama saectolare [26] » (Ashb. 524 f° 42v)

On retrouve le même genre de « complément technique » dans un autre passage où il est question des branches fruitières de la vigne. Au mot *braccia*

---

[25] Certains dialectes de la région Centre, en France, connaissent aussi des vocables dérivés de *caulis* : *chaulé* « tige de navet ou de pomme de terre ».

[26] « La meilleure façon de tailler la vigne sera celle qui consistera à laisser en place les sarments bien développés qui poussent dans la partie inférieure du cep, en ne gardant qu'un ou deux yeux : on appelle cette façon de procéder "ébourgeonnage". »

(littéralement « bras », c'est-à-dire « les sarments ») le traducteur rajoute l'adjectif *madornali*, « principaux ». Les *braccia madornali* désignent donc les branches charpentières – ou « fruitières » – de la vigne. Dans ce contexte spécifique, pour caractériser la branche d'un cep de vigne l'adjectif *madornale* est inusité avant ce passage de l'*Opus* :

> *Ubi quattuor pedibus supra terram levatae steterint, quaterna brachia habeant* (*Op. Agr.* III 12, 6) = « Quando saranno levate quattro piedi sopra terra si aranno quattro braccia cioè quattro buoni tralci madornali procedenti dalle quattro parti della vite[27]» (*Ashb. 524* f° 42v)

Un autre néologisme technique directement puisé dans le vocabulaire agricole oral (et qui sera destiné à une grande fortune dans la littérature agronomique) est le vocable *garzullo*, (« la partie la plus tendre, le cœur de la laitue ou du chou »), traduction du latin *caulis*. Le vocable est dérivé de *garzuolo* (< lat. **cardiolum*, « petit cœur ») ; il s'agit probablement d'un hellénisme originaire du territoire de l'Esarcat de Ravenne (DEI *s.v* .), étant donné qu'en grec ancien *kardìa* indiquait déjà le « cœur », pour ainsi dire, d'un arbre (« aubier »). René Martin, dans son édition critique de l'*Opus agriculturae*, l'a traduit, à tort, par « tige » (*caulis*), comme dans un autre passage examiné plus haut (*Op. agr.* VIII 2, 3)[28]. Mais l'auteur de la version vernaculaire a bien perçu la différence sémantique entre les deux lexèmes, d'où son choix d'utiliser un vocable différent, par rapport à celui employé dans le passage du livre VIII[29]**.** L'occurrence palladienne constitue probablement la première attestation de ce sémantème dans les langues romanes[30]. Terme technique propre au basilecte

---

[27] « Quand la vigne aura atteint la taille de quatre pieds, on aura quatre "bras", c'est-à-dire quatre branches fruitières ("coursonnes") bien développées, qui auront poussé à partir des quatre sarments porteurs du cep. »

[28] Voir passage relatif à la note n° 24.

[29] *Latior fit si rara ponatur vel, cum producere incipiet caulem, eo leviter inciso gleba prematur aut testa... Item multis seminibus condita nascetur, si caprini stercoris bacam subula subtiliter excavariset in ea semen lactucae, nasturtii, ocimi, erucae, radicis inmiseris, tunc involutam fimo bacam terra optime culta brevi scrobe demerseris.* (*Op. Agr.* II 14, 2-3) = « Più larga est et si fa s'ella est posta rada o come ella comincia a producere la foglia quella leggermente si fenda nel garzullo et con terra si corichi et con testo... Ancora nasceranno condite et saporose di molti semi se sottilmente tu scaverai con una lesina uno ca[l]cherello et in quello mecterai seme di lattuga et di nastruzo et d'orzimo et di ruca et di radice et questo cacherello involto con letame sotterrerai in una piccola fossa d'octima terra divelta» (Ashb. 524 f° 32r) (« La laitue devient plus grosse si on la sème claire ou si, après avoir pratiqué une légère incision dans sa tige au moment où elle commence à sortir de terre, on la comprime avec une motte de terre ou une tuile... On peut aussi en obtenir à partir de plusieurs graines différentes : pour cela, on creuse délicatement une crotte de chèvre avec une alène, on met à l'intérieur des graines de laitue, de cresson alénois, de basilic, de roquette et de radis, on enrobe cette crotte de fumier et on l'enfonce dans un trou creusé dans un terrain parfaitement cultivé »).

[30] La deuxième plus ancienne attestation écrite d'un mot sémantiquement proche est celle du *Dictionnaire* de Cotgrave (1611) : ici *cœur* est la « partie intérieure d'un fruit ».

oral, celui de la menuiserie dans ce cas, est aussi *piallaccio*, littéralement « planche rabotée » (< (*tavola*) *piallata*), qui traduit le pluriel *plancae* (*roboreae*) (« planches, parquet, de chêne »), bien moins spécialisé : « *Plancae roboreae subponantur stationibus equorum* » (*Op. Agr.* I 21), « Piallacci di quercia sieno messi di socto alle stalle de' cavalli socto la paglia[31] » (*Ashb. 524* f° 15r). Le mot *piallaccio* était encore bien vivant dans le lexique de la menuiserie il y a une soixantaine d'années (*GDLI s.v.*).

Un autre néologisme lexical est l'hapax verbal est *allascarsi*, dérivé du latin *laxor* (forme intensive de *labeor*), « glisser, s'écrouler » (désignant ici l'effondrement des parois d'un puits à cause des infiltrations d'eau de la nappe phréatique) :

> *Sed dum fodiatur puteus, si terra non stabit vitio generis dissoluti aut umore laxabitur, tabulas obicies directas undique* (*Op. Agr.* IX 9, 3) = « Se l'acqua sarà limosa s'amenda con mescolamento di sale et infino che si cava el pozzo se la terra non farà che rovina per suo vitio o per umidità s'alasca, mecterai davanti d'ogni parte tavole[32] » (*Ashb. 524* f° 98r)

Le verbe *allascarsi* est dérivé de l'adjectif *lasco* (« relâché, détendu ») dont la diffusion, essentiellement dialectale, touche l'aire centrale de la péninsule italique, jusqu'à Naples, auquel on a rajouté le préverbe « ad- ». Avant ce passage dont il est question ici, le verbe *allascarsi* est propre au seul domaine de la marine à voile, où il désigne une corde qui a tendance à se détendre, un nœud qui se desserre. Son appartenance au basilecte oral agronomique me semble très probable.

Intéressant est aussi l'hapax verbal *biutare* ou *inbiutare*, qui désigne l'opération consistant à soigner certaines plaies du tronc des arbres en le recouvrant avec une pâte – la *biuta* – ou une bouillie à base de graisses végétales ou animales, de fumier bovin et de l'argile :

> 1. *Si truncus vitis sole aut pluviis aut noxiis animalibus est cavatus, purgamus, quicquid est mortuum, plagasque ea amurca linimus et terra : quod proderit adversus praedicta* (*Op. Agr.* III 12, 5) = « Se tu non disideri che la vigna si rinuovi et s'el pedale della vite est cavato dal sole o da pioggia o da nocenti animali purgheremo ciò che vè morto et fracido et ungeremo et biutaremo quelle piaghe con morchia et

---

Voir aussi lim. *cœur*, « trognon d'une pomme », *cural* (Agen), « cœur d'un fruit », prov. *couraou*, « cœur d'une pastèque », sav. *coralion*, *coraillon* (Neuchâtel), « cœur d'un fruit » (voir *FEW* II : 1171). Même dans le français courant le *cœur* est souvent celui de la laitue ou du chou (*FEW ibid.*).

[31] « Les stalles des chevaux auront un plancher de chêne. »

[32] « Si l'eau est pleine de boue, il faudra y ajouter du sel et, pendant toute la durée des travaux de creusement du puits, il faudra stabiliser les parois en installant des planches en bois, pour les empêcher de s'effondrer à cause de l'eau contenue dans la terre. »

terra mescolata la qual cosa gioverà contro le predecte piaghe[33] » (*Ashb. 524* f° 43r)

2. *Aliqui diversa dixerunt, vel singula in vasculis fictilibus picatis atque oblitis claudi vel argilla involvi vel solos pediculos creta adlini*… (*Op. Agr.* III 25, 17) = « Altri furono che dissono in altra guisa o che si chiudessono in vasi inpeciati et impiastrati o inviluppare in argilla o solamente e' picciuoli inbiutare di creta...[34] » (*Ashb. 524* f° 54v)

Ce néologisme n'a pas d'autre témoignage écrit que ceux examinés ci-dessus : même sa base de dérivation (*biuta*) n'est attestée que par le *Dizionario italiano* (1861-1879) de Niccolò Tommaseo (*s.v.*), qui la classe comme un lexème propre à la région de Pistoia en Toscane. Il est totalement absent des autres langues romanes et le *DEI* (*s.v.*) le considère comme propre à l'aire italo-romane (tosco-émiliano-ombrienne). Il serait dérivé du latin vulgaire *abluta*, participe passé substantivé féminin du verbe *abluere* (« enlever avec de l'eau »). L'appartenance du lexème au basilecte serait confirmé par l'existence de vocables apparentés en albanais (*bljute*), en serbo-croate (*bluta*), en néo-grec (*bluta*). Déjà au VIIIe siècle *bluta* indiquait l'albumen utilisé pour faire briller les mosaïques.

## 2. Néologismes sémantiques

Tout aussi nombreux sont les néologismes sémantiques : ainsi le verbe *spiccare* passe-t-il par un processus de spécialisation sémantique, du sens très général de « prendere » (existant depuis le XIIe siècle au moins - *GDLI* XIX : 881 ss.) à celui de « explanter un arbre » (pour le replanter ailleurs) : « *Si provincia indiget olivetis et non est, unde planta sumatur, seminarium faciendum est* » (*Op. Agr.* III 18, 6), « Se la provincia ha difecto d'oliveti et non vè onde si spicchino e' piantoni, si est da fare semenzaio[35]» (*Ashb. 524* f° 47r).

A ma connaissance, il n'existe aucune autre attestation de ce sémantème, excepté peut-être un passage de Francesco Sansovino (*Istoria dell'Imperio universale dei Turchi*, Venise 1654, p. 203, cit. dans *GDLI* XIX : 882) où il parle d'un soldat lequel « spicca un cocomero » (« qui cueillit une pastèque »).

Un intéressant hapax sémantique est le vocable *vergola*, qui désigne, dans le passage cité ci-dessous, l'aigrette caractéristique des paons, alors que les

[33] « Si le pied de la vigne se retrouve nu à cause des dégâts provoqués par le soleil, la pluie ou les animaux nuisibles sur la terre tout autour, il faudra éliminer toute partie morte ou pourrie et appliquer sur les plaies éventuelles une fine couche de graisse ou d'huile ou une pâte faite avec les résidus du pressurage des olives et de la terre. »

[34] « Selon d'autres experts [la meilleur façon de conserver les pommes] consisterait à les mettre dans des pots recouverts de poix, ou bien à les enduire d'une couche de glaise ou encore à les passer dans une pâte à base d'argile. »

[35] « Si la région manque d'oliveraies et qu'il n'y a donc aucune possibilité d'explanter et transplanter de jeunes plantes, il faudra penser à créer une pépinière. »

témoignages écrits précédents (*GDLI* XXI : 752), remontant au XIVe siècle, désignent un « tissu dont les franges sont brodées avec des fils dorés » : « *Masculi ova et pullos suos persequuntur velut alienigenas, priusquam illis cristarum nascatur insigne* » ( *Op. Agr.* I 28, 2), « E' maschi perseguitano l'huova et i polli suoi si come gli altrui prima che le vergole nelle creste nascano loro[36] » (*Ashb. 524* f° 17r).

Un processus de transformation sémantique par extension de l'acception primaire est à l'œuvre dans le verbe *annicare*, forme toscane pour le moderne *annicchiare* dans le sens, totalement inconnu par ailleurs, de « recouvrir de terre », littéralement « protéger, mettre à l'abri (la racine de l'arbre) dans une niche (*nicchia*) de terre » :

> *Radicibus debet adfundi calx vivus temperatus cum creta... inparis numeri poma cydonea pro magnitudine eius ponenda et obruenda firmantur* (*Op. Agr.* III 25, 23) = « Si dee gictare in sulle radici [sicl. : del melo cotogno] viva calcina temprata con creta (…) et socterravi a piede mele cotogne in numero caffo secondo la grandezza sua et s'anicherà » (*Ashb. 524* f° 54v)[37]

Dans ce cas, sa nature de technicisme oral agronomique me semble à privilégier, même si un fait de style ne serait pas à exclure *a priori*, en considération du fait que, dans un autre passage, le même verbe latin *obruere* est traduit par le terme beaucoup plus prosaïque de *ricoprire* (« recouvrir »)[38].
Des cas similaires sont représentés par les verbes *diffare* et *incapparsi* ou par le substantif *cacherello*. L'hapax sémantique *diffare* (forme toscane dérivée de *disfare*, par assimilation régressive du S), comme le latin *dissipare* dont il est la traduction, passe du sens de « détruire en éparpillant », « éparpiller », « mettre en pièces » à celui, plus technique, de « briser (les mottes de terre) à la pioche » ou « travailler (la terre) à la houe » : « *Glebae omnes dolabris dissipandae sunt* » (*Op. Agr.* II 3, 2), « Le zolle tucte si deono diffare col marrecto[39] » (*Ashb. 524* f° 28v). Bien qu'il s'agisse d'une traduction littérale, comme d'autres qu'on a examinées plus haut, je pencherais là aussi pour un technicisme agronomique : sa proximité sémantique avec le synonyme *rompere*, très répandu dans la traduction de l'*Opus agriculturae* et possédant cette acception de « briser la terre à la houe » dès le XIVe siècle, le laisserait supposer. Ce sens a pu se développer à partir de celui de « réduire en poudre (de la craie) », dont la première

---

[36] « Les mâles s'en prennent à leurs œufs et à leurs propres petits, comme s'ils n'étaient pas à eux, jusqu'à ce que leur soit poussée l'aigrette caractéristique. »
[37] « Il faudra couvrir les racines du cognassier avec de la chaux vive mélangée avec de la glaise... il faudra également enterrer au pied de l'arbre des coings en nombre impair – la quantité dépendra de la taille de l'arbre – et recouvrir le tout avec de la terre. »
[38] *Semina asparagi pingui et stercorato loco in singulis fossis pones et leviter obrues* (*Op. Agr.* IV 8, 11) = « In luogo grasso et letaminato porrai [scil. : i semi degli asparagi] et lievemente le ricuopri» (*Ashb. 524* f° 64r) (« Les graines des asperges, il faut les planter dans un terrain bien gras qu'on aura préparé en y ajoutant du fumier et ensuite les recouvrir avec de la terre »).
[39] « Il faut briser à la pioche toutes les mottes de terre. »

attestation est dans le *Libro dell'arte*, rédigé par Cennino Cennini à la fin du XIVe siècle.

La forme verbale réflexive *incapparsi*, énième hapax sémantique, passe, quant à elle, du sens général de « tomber sur quelqu'un » ou même « faire une mauvaise rencontre » (sens que le verbe possède toujours en italien moderne) à celui d' « être blessé, coupé par un outil en fer » : « *Cum vitis bidente laeditur aut ferro, plagam, si terra iuncta est, adline stercore ovillo* » (*Op. Agr.* IV 7, 5), « Quando le viti s'incappano con alcuno ferro, se est la calteritura lungo la terra, impiastrala con sterco pecorino[40]» (*Ashb. 524* f° 62r). La forme réflexive du verbe « incappare » est une tentative, peut-être maladroite, du traducteur anonyme, de rendre la forme passive du texte source (*laeditur*). Dans ce même passage, *cacherello* (littéralement « petite crotte », « crottin ») assume le sens très spécifique de « crottin de bique », alors que ce diminutif de *ca(c)ca* (« crotte ») a dû vraisemblablement indiquer, à l'origine, les excréments d'une quelconque bête de petite taille. Cette nouvelle acception jouira d'une immense fortune dans la littérature agronomique jusqu'au XIXe siècle (*GDLI, s.v. ; DEI s.v.*).

D'autres néologismes trahissent une certaine maladresse de la part d'un représentant d'une école de traduction encore à ses débuts : ainsi en est-il du syntagme *canale da pigiare*, traduction du latin *calcatorium* (« fouloir, pressoir »), littéralement le « conduit qui sert pour presser (les raisins) », c'est-à-dire le « conduit relié au pressoir » :

> *Cellam vinariam septemtrioni habere debemus oppositam, frigidam vel obscurae proximam… ita instructam necessariis, ut non vincatur a fructu, sic autem dispositam, ut basilicae ipsius forma calcatorium loco habeat altiore constructum, ad quod, inter duos lacus qui ad excipienda vina hinc inde depressi sint…* (*Op. Agr.* I 18, 1) = « La cella da vino dobbiamo havere verso septentrione fredda, quasi scura... et non si piccola che sia vinta dal fructo ma sia disposta si che la forma di questa cella habbia lo canale da pigiare ordinato illuogo più alto al quale intra due vasi e' quali a ricevere el vino stieno l'uno di qua e l'altro di là... [41]» (*Ashb. 524* f° 14r)

Un autre cas identique est représenté par la lourde paraphrase utilisée pour rendre le latin *deductoria liquoris* (« canal d'évacuation » des eaux souterraines), « là ove discorre umore » (littéralement « l'endroit où l'eau coule, s'évacue[42] »),

---

[40] « En cas de plaie provoquée par un outil en fer sur le cep de la vigne, si la blessure est au ras du sol, il faudra la soigner en l'enduisant de crotte de brebis. »

[41] « Le cellier doit être exposé au nord, frais et presque obscur… suffisamment équipé pour permettre de faire face à une récolte abondante, enfin conçu de telle sorte qu'il soit muni d'un fouloir surélevé comme la tribune d'une basilique, auquel on accédera par trois ou quatre marches, et de chaque côté duquel deux réservoirs seront ménagés… »

[42] *Locus tamen deductoria liquoris accipiat, ne humor insidens limo germen extinguat* (*Op. Agr.* XII 7, 21) = « El luogo sia volto verso là ove discorre omore, così che l'omore covandovi con suo letame non spenga la pianta » (*Ashb. 524* f° 121r) (« Le terrain où l'on plantera

ou encore par les noms *suoni* et *panni*, qui correspondent respectivement aux « crécelles » (*crepitacula*) et aux « épouvantails » (*panni*) du texte source :

> *Ab apium castris longe sint omnia odoris horrendi, balneae, stabulae, coquinae fusoria. Fugemus praeterea animalia quae sunt apibus inimica, lacertos, et blattas et his similia; aves etiam pannis et crepitaculis terreamus.* (*Op. agr.* I 37, 4) - « Ma dagli habituri di queste api sieno di lungi tucte le cose d'odore fetido et da fuggire si come bagni, stalle, cucine, sedii, acquai, fogne et simili cose. Et ancora fuggiremo tucti gli animali nimichevoli alle api, si come sono lucertole et blate et simiglianti a queste. Spaventiamo ancora gli uccelli con suoni et con panni[43] » (Ashb. 524, f° 24r)

Ce sens de *panni* constitue un hapax sémantique, l'énième de la version florentine de l'*Opus agriculturae*[44], et illustre bien la technique traductive des premiers humanistes toscans, basée sur une fidélité extrême au texte source. L'autre néologisme qu'on retrouve dans ce même passage, *suoni* (« crécelle »), s'avérera par contre particulièrement productif en italien[45].

## 3. Calques

Les calques sont aussi très nombreux : parmi ces derniers, particulièrement intéressant est le cas de l'hapax *coprimento*, du mot à mot « couverture », « manteau », traduction du latin *amictus*, qui désigne ici la « roue du paon ». Il s'agit de toute évidence d'une traduction littérale du vocable du texte source qui désigne un manteau à capuche qui couvrait les épaules et une partie du dos, utilisé ici dans un sens figuré. Ce sémantème, partiellement modifié[46], s'avère productif dans la littérature italienne technique des siècles suivants puisqu'on le retrouve dans l'*Histoire Naturelle* de Pline traduite par Lodovico Domenichi (1515-1564) parue à Venise en 1562 et très vraisemblablement influencée par la version florentine de l'*Opus* de 1464[47]. Les exemples de calques peuvent être aisément multipliés. Je me limiterai aux verbes *spesseggiare* et *scortecciare*,

---

les châtaigniers devra être muni d'un canal d'évacuation des eaux souterraines, pour éviter que celles-ci, en s'accumulant dans le sol, ne fassent pourrir les graines »).

[43] « Mais il faut que soient éloignés des ruches tous les endroits malodorants : bains, étables, fosses d'écoulement des eaux de la cuisine. Il faut aussi en chasser les animaux qui sont nuisibles aux abeilles : lézards, blattes et autres semblables ; les oiseaux également doivent être effrayés par des épouvantails et des crécelles. »

[44] Il n'y a apparemment aucune attestation dans les sources écrites du mot *panni* ayant le sens dans lequel l'utilise notre traducteur (*GDLI* XII : 490-493).

[45] *GDLI* XX, p. 524 : la première attestation après celle de la version vulgaire de l'*Opus* est représentée par le *Novellino* (1525). Le premier témoignage de l'it. *crepitacolo* (« crécelle ») date de la fin du XVII^e^ siècle (*GDLI* III : 954). Pour d'autres cas de figure similaires voir Campetella, 2009 : 91.

[46] Chez Domenichi, le lexème indique les différents épidermes de certains animaux aquatiques, comme le pelage des veaux et des hippopotames, la carapace des tortues ou encore les écailles des poissons (*GDLI* III : 755).

[47] Voir *GDLI s.v.*

respectivement hapax lexical et sémantique. Le premier, qui signifie « passer plusieurs couches de peinture sur un mur », formé sur l'adverbe *spesso* (« souvent », « fréquemment ») est calqué sur le latin *frequentare*, lui-même dérivé de l'adverbe *frequenter*, « souvent », « à plusieurs reprises[48] ». *Scortecciare* est un calque du latin *ex-codicare*. Ce dernier désigne l'opération d'« enlever la terre autour du pied du cep de la vigne », à l'endroit où l'écorce – la *corteccia*, *codex* en latin – ou plutôt les résidus ligneux qui se sont accumulés au cours de l'année précédente sont plus épais. Le passage en question constitue le seul témoignage écrit du sémantème, puisque le terme technique intégrera officiellement le vocabulaire agronomique italien des siècles suivants mais dans le sens, bien différent, d' « enlever l'écorce », « nettoyer le tronc » (de n'importe quel arbre ou arbuste).

Intéressant est aussi le cas de spécialisation technique constitué par le mot *acrescimenti*, littéralement « augmentations, développements » (*incrementa*), dans un passage consacré à la meilleure façon de tailler la vigne : « *Longius adminiculum vitis incrementa producit* » (*Op. Agr.* I 6, 9), « Più lungamente l'aiutorio facto alle viti produce e' loro acrescimenti [49] » (*Ashb.* 524, f° 7v). Le vocable *acrescimenti*, indique ici le « jet » (« gettata » en italien moderne), c'est-à-dire les nouvelles pousses engendrées à partir des yeux, les bourgeons de la vigne d'où se développeront plus tard les « coursonnes », les branches fruitières[50].

## Conclusion

La traduction florentine de l'*Opus agriculturae* de 1464 symbolise parfaitement l'attitude d'une *intelligentsia* intellectuelle dont la langue littéraire avait déjà atteint son apogée depuis au moins deux siècles, mais dont le lexique technique était encore en formation. La maladresse du traducteur, à laquelle nous avons fait plusieurs fois allusion au cours de cette étude, fait pendant à la richesse débordante de l'acte créateur, presque démiurgique, de la toute nouvelle prose scientifique. Au siècle suivant il en ira tout autrement : les néologismes seront beaucoup moins abondants dans les traductions de textes de la même nature. Pour s'en convaincre il suffit de lire la version du même *Opus Agriculturae*

---

[48] *Parietum vero tectura sic fiet fortis et nitida : prima trullis frequentetur inductio* ; *cum siccari coeperit, iterum inducatur ac tertio* (*Op. Agr.* I 15) = « La intonicatura della parete si farà così forte et necta : prima con la cazzuola si spesseggi et quando lo intonicato si comincerà a seccare, l'altra lo intonica et poi terza volta » (Ashb 524 f° 12v) (« Voici comment on rendra solide et brillant le revêtement des murs : il faut passer et repasser la truelle sur la première couche ; quand celle-ci commencera à sécher, en appliquer une deuxième, puis une troisième »).

[49] « Un support élevé fait croître la vigne en hauteur. »

[50] Nous avons déjà examiné plus haut deux autres cas de spécialisation : *cacherello*, qui désigne exclusivement les petits crottins de chèvre (*Op. agr.* II 14, 2 = Ashb. 524 f° 32r) et *spiccare* (*Op. agr.* III 18, 6 = *Ashb. 524* f° 47r).

rédigée par Francesco Sansovino en 1561[51] pour un aristocrate vénitien. Le peaufinage stylistique y est parfois poussé à l'extrême, en raison de la volonté de ces écrivains du *Cinquecento* de se démarquer de Cicéron et de créer une langue meilleure, plus élégante. Par contre, la vivacité lexicale du texte ainsi que la faculté possédée par ces auteurs de façonnage sémantique des vocables techniques y sera nettement inférieure.

## Abréviations

DEI = BATTISTI C., ALESSIO G. (1950-1957), *Dizionario Etimologico Italiano*, Firenze, Barbèra.
DELI = CORTELAZZO M., ZOLLI P. (1980-1985), *Dizionario Etimologico della Lingua Italiana*, 5 voll., Bologna, Zanichelli.
FEW = WARTBURG (VON) W., (1928-1966), *Französisches Etymologisches Wörterbuch*, Bonn-Leipzig-Berlin.
GDLI = Battaglia S. (dir.) (1961-2000), *Grande Dizionario della lingua italiana*, Torino, UTET.
REW = MEYER-LÜBKE W. (1911), *Romanisches Etymologisches Wörterbuch*, Heidelberg, C. Winter.

## Références

BALLARD M. (1992), *De Cicéron à Benjamin. Traducteurs, traductions, réflexions*, Lille, Presses Universitaires du Septentrion.
CAMPETELLA M., (2009), Les traductions du latin en italien des XVe-XVIe siècles, *Traduire*, Vol. 221, 86-103.
DIONISOTTI C. (1967), Tradizione classica e volgarizzamenti, In Dionisotti C. (dir.), *Geografia e storia della letteratura italiana*, 103-144, Torino, Einaudi.
FOLENA G. (1973), Volgarizzare e tradurre, In Malmberg B. (dir.), *La traduzione. Saggi e studi*, 59-109, Trieste, Lint.
GIOVANARDI C. (1994), Il bilinguismo italiano-latino del Medioevo e del Rinascimento, In Serianni L., Trifone P. (dirs.) *Storia della Lingua italiana*, II, *Scritto e parlato*, 435-467, Torino, Einaudi.
MARTIN R. (1976), *Palladius*, *Traité d'agriculture*, Paris, Belles Lettres.
ORVIETO P. (1996), Lorenzo de' Medici e l'Umanesimo toscano del secondo Quattrocento, In Malato E.(dir.), *Storia della letteratura italiana*, III, *Il Quattrocento*, 295-403, Roma, Salerno.
SVENNUNG J. (1928), De Columella per Palladium emendato, *Eranos*, Vol. 32, 145-208.
VAN HOOF H. (1991), *Histoire de la traduction en Occident : France, Grande-Bretagne, Allemagne, Russie, Pays-Bas*, Paris, Duculot.

[51] *La villa di Palladio Rutilio Tauro Emiliano tradotta nuovamente per Francesco Sansovino, nella quale si contiene il modo di coltivar la terra di mese in mese, di inserire gli arbori, di governar gli horti & i giardini, con la proprietà dei frutti, delle herbe & degli animali...*, in Venetia, appresso Francesco Sansovino, 1561

# Neologismos de ida y vuelta. De la *Doctrina de la neurona* de Cajal al *Principles of Neural Sciences* de Kandel[1]

Joaquín García Palacios[2], Miguel Sánchez Ibáñez[3]

## Resumen

Los mecanismos de actualización de las lenguas de especialidad reflejan las relaciones que, a través de los intercambios de conocimiento y del establecimiento de lenguas que se convierten en referentes en el contexto internacional, van configurando el panorama científico de las diferentes disciplinas. Con frecuencia, estos vínculos demuestran dinámicas de jerarquización y dependencia terminológica, pero también pueden dejar entrever rasgos de dinamismo neológico en lenguas aparentemente dependientes. El análisis de la neología en español en el campo de las neurociencias permite observar esa dilogía. La perspectiva diacrónica de este estudio, que analiza la creación de términos especializados en ese ámbito durante dos momentos concretos de la historia de la lengua española, separados por casi un siglo, nos habla del camino recorrido por una lengua de especialidad, desde su estatus de referencia mundial en un ámbito del conocimiento (de la mano de un especialista de importancia capital en el nacimiento de la disciplina, Santiago Ramón y Cajal) hasta la categoría actual de código dependiente e importador de unidades.

## Palabras clave

Neología terminológica, dependencia terminológica, neurociencias, lengua española

## 1. Introducción

El estudio sobre la neología terminológica en una lengua ha de abarcar necesariamente las producciones lingüísticas en diversas áreas de conocimiento si lo que se pretende es tener una visión lo más completa posible acerca del

---

[1] La investigación para realizar este estudio ha sido subvencionada por la Junta de Castilla y León con cargo al proyecto *Mecanismos de generación neológica y actualización léxica del español a través del estudio de la prensa castellanoleonesa* (SA030A10-1). Este proyecto está posibilitando, además de estudios sobre neología general, abordar otros trabajos de neología comparada como el que se presenta en estas páginas.

[2] Grupo NEOusal, Universidad de Salamanca, gpalacios@usal.es

[3] Grupo NEOusal, Universidad de Salamanca, miguelsanchez@usal.es

funcionamiento de los mecanismos de generación léxica en los ámbitos especializados en una lengua en un momento determinado de su historia.

Así el grupo NEOusal de la Universidad de Salamanca ha ido acometiendo el estudio de la neología terminológica a partir de análisis parciales de distintos ámbitos[4]. Entre estos adquiere un lugar fundamental el estudio sobre el ámbito de las neurociencias por diversas razones.

En primer lugar porque se trata de un ámbito de especialidad que se ha desarrollado a lo largo de todo el siglo XX y está experimentando un avance notorio en las últimas décadas, con un desarrollo muy significativo en las investigaciones llevadas a cabo en todas las subáreas que lo integran. En el contexto actual, de un envejecimiento progresivo de la población, las disciplinas que se engloban dentro de esa etiqueta general de neurociencias, han adquirido un protagonismo particular dentro de la medicina.

Porque además, al tratarse de un ámbito con una intensa actividad científica, el movimiento constante del conocimiento que se va acumulando y transformando en ella propicia los cambios para conseguir adecuar el lenguaje usado por los especialistas. El progreso del propio dominio obliga a actualizar el código empleado, lo que favorece la aparición de nuevas formas léxicas, o la transformación de las ya existentes, para denominar las realidades novedosas.

En tercer lugar porque el de las neurociencias es un campo que puede tomarse como modelo significativo para la caracterización de la comunicación científica en una lengua románica como el español en el momento actual.

Solo por esas tres razones podrían estudiarse los procesos de creación de léxico especializado poniendo la atención en un campo del conocimiento como este. Pero se trata además de un ámbito pujante en el que se ven reflejados fielmente los movimientos, las jerarquías y las relaciones de fuerza que se producen entre las distintas comunidades lingüísticas. Estas comunidades, así como los distintos grupos de hablantes que las integran, con sus actitudes y prejuicios lingüísticos, condicionan, como veremos más adelante, dichos procesos neológicos.

Por otro lado, y debido al interés que tienen para nosotros tanto la dimensión sincrónica como la diacrónica de los términos, hemos querido partir de los textos en los que la neurociencia se va conformando como disciplina científica, máxime cuando ese punto de partida se había gestado en la lengua que constituye el objeto principal de nuestro estudio. Es en español, y debido a la pluma de Santiago Ramón y Cajal, en la lengua en que se redactan los artículos fundacionales de la neurociencia. Como dice J. de Felipe (2007a: 83), "sus estudios iniciados en 1887 sobre la microorganización del sistema nervioso…

[4] Estos estudios se han centrado por el momento en los ámbitos de la archivística, la teledetección espacial y las enfermedades infecciosas (sida).

han proporcionado el esqueleto intelectual de nuestras investigaciones actuales sobre la estructura y función del cerebro en condiciones normales y patológicas".

## 2. Objetivos del estudio

Los estudios anteriores en que hemos abordado la caracterización de distintos procesos neológicos en los que se pone de manifiesto la dependencia terminológica de las lenguas románicas respecto del inglés en el momento actual (García Palacios, 2010; García Palacios y Sanz, 2012; Humbley y García Palacios, 2012), nos llevan a intentar dar un paso más, y a que en este artículo se nos plantee como fundamental la comparación entre dos momentos en los que la comunicación en el ámbito de las neurociencias en español transcurría por derroteros completamente diferentes. Por un lado la época en la que este ámbito estaba asentando sus cimientos, y lo hacía en los artículos científicos que Ramón y Cajal publicaba en español. Por otro, los últimos años, en los que apenas se publican textos científicos de primer nivel editados en esa lengua. Nuestro estudio debería mostrar cómo los neologismos van generándose en español en el primer caso, y cómo, en el segundo, los textos que se producen en español incorporan neologismos que se han adaptado a partir de las denominaciones primeras generadas en inglés.

Por tanto el interés de este estudio debería residir en mostrar en el primer caso una terminología en pleno desarrollo en una lengua, y cómo esa terminología, en la misma lengua, se convierte en deudora de otra en los últimos tiempos.

Al mismo tiempo que perseguíamos ese objetivo principal de estudiar la neología terminológica en español en el campo de las neurociencias, atendiendo a dos momentos relevantes (y bien diferenciados) de su recorrido disciplinar, hemos pretendido con nuestro trabajo:

- Conseguir una muestra que nos permitiese extraer conclusiones más amplias sobre el papel desempeñado por los procesos de formación primaria frente a los de formación secundaria en las primeras fases de creación léxica.
- Reflexionar sobre la implicación del científico en el establecimiento de denominaciones especializadas eficientes y productivas.
- Reparar en el papel de la traducción a lo largo del proceso, teniendo en cuenta el peso del multilingüismo y de las relaciones jerárquicas que se establecen entre las distintas comunidades de especialistas.
- Incidir en la importancia de la dimensión diacrónica de los lenguajes de especialidad para los estudios de neología.
- Contribuir con los resultados de este estudio a la configuración general de nuestras investigaciones en torno a la neología en el campo de las neurociencias.

## 3. Cuestiones metodológicas

La orientación lingüística de nuestro estudio, coincidente en los puntos básicos con los postulados de la Teoría Comunicativa de la Terminología, conformada por T. Cabré y sus colaboradores en distintos artículos (Cabré, 1999, 2008) nos lleva a asumir también otras cuestiones de carácter metodológico que no se han tratado suficientemente al enunciar una teoría que se centra en la terminología, y no tanto en la neología terminológica:

- La necesidad de no olvidar la inscripción de los términos neológicos en un eje diacrónico al tiempo que se procede al estudio sincrónico de los mismos[5].
- El enfoque del hecho neológico desde una perspectiva global que atienda a las cuestiones formales de los términos así como a los aspectos de tipo pragmático que los determinan.
- Las particularidades de la actuación de los distintos agentes implicados en las situaciones comunicativas en que se generan los términos neológicos. Por un lado los científicos, creadores de la terminología del campo de especialidad, y por otro los intermediarios lingüísticos, colaboradores necesarios para que se lleve a cabo la difusión de los neologismos terminológicos.

Para la realización del estudio se procedió a constituir dos corpus, ambos integrados por textos reales, producidos directamente en español y pertenecientes al ámbito de las neurociencias. Sin embargo, la distribución cronológica de estos corpus es muy diversa: mientras que el primero de ellos abarca desde finales del siglo XIX hasta 1933, el segundo recoge textos de los últimos años (2010 y 2011). Por otro lado, y en cuanto a lo relacionado con la autoría de los textos, el primer corpus pertenece en su totalidad a un único científico, pero el segundo recoge artículos de revista pertenecientes a varios autores.

Tampoco desde el punto de vista temático nos encontramos con la homogeneidad de los dos corpus. En los textos que integran el primero de ellos se sientan las bases de la doctrina neuronal de Santiago Ramón y Cajal, es decir, los fundamentos de lo que hoy entendemos por neurociencias. En el segundo, en cambio, se aborda un único tema relacionado con las neurociencias: la enfermedad de Alzheimer.

---

[5] No se trata únicamente de actuar desde una perspectiva filológica que muestre la filiación de las voces neológicas, sino de conseguir la integración entre el análisis diacrónico y el sincrónico en el estudio de las unidades terminológicas de carácter neológico. Los trabajos de P. Dury y A. Picton (2009), o de A. Picton (2009) son una buena muestra de esa orientación. Puede consultarse también el libro de D. Candel y F. Gaudin (2006).

Sí es verdad que en uno de sus aspectos caracterizadores el peso de los corpus es semejante, pues ambos tienen una cantidad similar de palabras: el primer corpus unas noventa mil palabras y el segundo alrededor de las ochenta y cinco mil.

| | | | |
|---|---|---|---|
| CORPUS 1 "Cajal" | Todos los artículos de la *Revista Trimestral de Histología* | Entre 1888 y 1889 | Más de 50.000 palabras |
| | Artículo *¿"Neuronismo" o "reticularismo"?* | 1933 | Más de 40.000 palabras |
| CORPUS 2 "Revista Española de Neurología" | Artículos con "enfermedad de Alzheimer" como palabra-clave | 2010 - 2011 | Unas 85.000 palabras |

La comparabilidad de estos dos corpus no se sustenta en este caso en los factores que normalmente determinan que dos corpus especializados se consideren como comparables.[6]

De hecho nuestros corpus podrían entrar dentro de la categoría de lo que Pearson (1998: 48) denomina "special purpose corpora", es decir, un corpus cuya composición responde a un objetivo determinado y cuyas características, en cuanto a extensión, tipología textual o lengua estarán marcadas, en consecuencia, por ese objetivo. Nos interesaba comparar una muestra representativa de textos de una disciplina científica en el momento en que esta se estaba comenzando a desarrollar, con otro conjunto de textos de la misma materia en el momento actual, para poder observar así lo que ha cambiado y lo que ha permanecido igual con el paso del tiempo en lo que se refiere a los procesos de generación terminológica en ese lenguaje de especialidad.

El primer corpus (al que denominaremos *Cajal* o *corpus1*) está integrado por dos subcorpus:

[6] Estamos, pues, ante dos conjuntos de textos que no responden a la definición de corpus comparables, si los entendemos como "collections of monlingual corpora which use the same or similar sampling procedures and categories for each language" (Pearson, 1998: 57), ni que tampoco se adscriben a la categoría de los que normalmente se denominan paralelos, puesto que no cumplen las premisas de Teubert (1996: 245), quien especifica que para que dos corpus sean comparables han de tener una cantidad similar de palabras y coincidir en cuanto a la tipología de las estructuras textuales (*ibid*).

*Cajal1*, en el que hemos incorporado todos los artículos aparecidos en la *Revista Trimestral de Histología Normal y Patológica*, que el propio Ramón y Cajal costeó con el fin de publicar y difundir rápidamente los avances que estaba realizando en su laboratorio de la Universidad de Barcelona[7]. Son once artículos altamente especializados, distribuidos en dos números (el primero de 1888 y el segundo de 1889) en los que el científico español, en un momento de gran creatividad, va exponiendo los descubrimientos que está haciendo y que constituyen la base de toda la teoría neuronal. Unos textos que, en estudios muy significativos sobre la obra de Cajal (Tello, 1935; Laín, 1946; López Piñero, 2006; De Felipe, 2007a y 2007b), se señalan como fundamentales para entender toda la concepción actual del sistema nervioso, la que el propio Cajal empezó a llamar la "nueva verdad". Artículos que rompían con planteamientos anteriores y que supusieron el reconocimiento internacional del investigador español.

*Cajal2* está constituido por una pequeña monografía publicada en 1933, un año antes de la muerte de Cajal, y titulada *¿Neuronismo o reticularismo?*, que a continuación lleva el subtítulo "Las pruebas anatómicas de la estructura nerviosa". En ella el autor hace un buen resumen de su "doctrina de la neurona" o neuronismo, la nueva teoría que él contribuyó como nadie a fundamentar, y que fue consiguiendo el apoyo de la mayoría de los grandes histólogos de su época, como His, Forel y Kölliker, entre otros. Aprovecha al mismo tiempo esa publicación para poner un poco de orden en toda la controversia que se había producido entre quienes defendían "la vieja tradición de las redes de Gerlach" (Cajal, 1933: 1), es decir, la unión por contigüidad de las células nerviosas, como formando una especie de red, y quienes, como él, habían demostrado con pruebas objetivas la discontinuidad neuronal y la relación por contacto entre unas neuronas y otras.

Para elaborar el segundo corpus (denominado *corpus2*) nos centramos en el ámbito de las publicaciones especializadas en español y comprobamos que había una que respondía a nuestras necesidades de especificidad temática, actualización cronológica y accesibilidad: la *Revista Española de Neurología*, que tiene como objetivo tratar los principales progresos disciplinares en el ámbito y que es la publicación oficial de la Sociedad Española de Neurología[8] y de más de quince sociedades similares de carácter regional. En la actualidad se han publicado 56 volúmenes, de carácter bianual. Los volúmenes están compuestos por un número variable de números, con una periodicidad bimestral en unos

---

[7] En la *Revista Trimestral de Histología Normal y Patológica* aparecen algunos artículos fundamentales en la concepción científica de Cajal, como "Estructura de la retina de las aves", "Sobre las fibras nerviosas de la capa granulosa del cerebelo", o "Contribución al estudio de la estructura de la médula espinal".

[8] Más información en www.sen.com. Se trata de una sociedad científica española fundada en 1949. En el año 2010 tenía en torno a 1850 miembros, que responden al perfil de médico español especializado en esta rama de la medicina. La sociedad se encuentra adscrita a organizaciones internacionales del ámbito, como la World Federation of Neurology o la European Federation of Neurological Societies.

casos, mensual y quincenal en otros, en función de los años. Su página web (www.revneurol.com) pone a disposición de los lectores todos los artículos de los números publicados desde el volumen 26 (enero-junio 1998)[9] por separado y en formatos fácilmente descargables y manejables para su análisis textual.

Todos los artículos están indexados en función de una serie de áreas temáticas a las que se adscriben al ser incluidos en la publicación. Cada una de estas áreas responde a un subdominio dentro del ámbito neurológico, y un mismo artículo puede estar clasificado en varias de ellas. Configuramos el corpus con todos los artículos aparecidos en la Revista Española de Neurología entre los años 2010 y 2011 que contienen "enfermedad de Alzheimer" como palabra clave.

Además de ese conjunto de textos, tomamos otras dos obras como textos de apoyo. Se trata de dos manuales de referencia para los especialistas en español, uno redactado originalmente en castellano (*Manual de Neurociencia*, de J.M. Delgado *et al.*) y otro traducido del inglés (*Principios de Neurociencia*, de E. Kandel *et al.*)

Las divergencias que han impedido la constitución de dos corpus realmente comparables surgen de un hecho fundamental, como es la evolución de la propia disciplina y de la comunicación científica en general a lo largo del siglo XX. Como iremos viendo en nuestro estudio, a lo largo del último siglo se fue pasando de una situación en la que era normal la convivencia entre unas cuantas lenguas, y la ayuda que prestaba la traducción cuando el entendimiento entre los científicos presentaba dificultades, a otra situación muy distinta en la que una sola lengua se ha ido imponiendo como lengua central, y única, en ese tipo de comunicación.

> A partir de los años 30 comienza una fuga de cerebros desde Europa hacia América que contribuye a que Europa pierda el protagonismo (…): inicia así el inglés una carrera imparable hacia la universalidad de su condición, siendo una peculiaridad de la ciencia americana de este siglo el ser cultivada en gran medida por no americanos y desarrollada en inglés por no anglófonos. (Gutiérrez Rodilla, 1998: 81)

El caso de las neurociencias no fue una excepción al planteamiento general que se hace en el párrafo anterior. Mientras que Ramón y Cajal escribe en su lengua, aunque muchos de sus textos son traducidos con bastante rapidez al francés, al alemán, y en algunos casos al inglés[10], varios de sus principales discípulos, como Lorente de Nó o Fernando de Castro, han de abandonar sus investigaciones en España con motivo de la guerra civil, y las continúan en países de habla inglesa.

---

[9] Última consulta: julio de 2012.

[10] Dice Ramón y Cajal en *Textura del sistema nervioso del hombre y de los vertebrados*, tomo I, prólogo, p. VI: "…nuestro modesto trabajo de 1894, *Las nuevas ideas sobre la estructura del sistema nervioso*, del cual se agotaron rápidamente las ediciones española y alemana, y dos tiradas sucesivas de la excelente versión francesa del Dr. Azoulay".

Lorente de Nó marcha a los Estados Unidos en 1931, y permanece en ese país hasta su jubilación, "porque España no le ofrece los medios para sus ambiciones de investigación", según le cuenta a Cajal en una carta en 1930[11]. Otros neurocientíficos españoles como Pío del Río-Hortega (quien no siguió directamente el magisterio de Cajal, pero sí su influencia), tuvieron que exiliarse también a países anglófonos en los años 30 del siglo XX, con lo que comienzan a producir sus textos en inglés, lo que va a contribuir sin duda a marcar el cambio de tendencia lingüística en el ámbito; es a su vez un fiel reflejo de la progresiva pujanza que adoptó el inglés como lengua para el intercambio lingüístico en el transcurso del pasado siglo[12]. La simple lectura y comparación de los títulos de los artículos escritos por Cajal y por sus discípulos directos nos revela el progresivo cambio hacia el "monolingüismo anglófono" del que hablan autores como Martínez-Almagro (2007: 62).

## 4. Análisis de datos: dos pilares de la neurociencia

Los análisis realizados en nuestro estudio sobre los dos corpus anteriormente descritos han girado en torno a dos núcleos temáticos que constituyen además los pilares sobre los que se asienta la teoría neuronal de Cajal: el concepto de 'neurona' y el de 'sinapsis'. Dos conceptos fundamentales para entender la neurociencia, a los que el mismo Cajal se refería así ya en 1889:

> (…) las células nerviosas son elementos independientes jamás anastomosados ni por sus expansiones protoplasmáticas [dendritas] ni por las ramas de su prolongación de Deiters [axón], y que la propagación de la acción nerviosa se verifica por contactos al nivel de ciertos aparatos o disposiciones de engranaje, cuyo objeto es fijar la conexión, multiplicando considerablemente las superficies de influencia.[13]

### 4.1. La neurona

El concepto 'neurona' entendido como una 'célula del tejido nervioso formada por un cuerpo celular del que se ramifican diversas dendritas y un axón, y que constituye la unidad funcional del sistema nervioso'[14], aparece configurándose

---

[11] R. González Santander *La escuela histológica española*, tomo IV, Alcalá de Henares, 2000, p. 159.

[12] Puede consultarse a este respecto la obra de Rafael González Santander *La escuela histológica española* (sobre todo los tomos IV y VI, Alcalá de henares, 2000 y 2003 respectivamente).

[13] *Apud* De Felipe (2007b: 59). Las aclaraciones entre corchetes son de J. de Felipe.

[14] Definición traducida de la que proporciona el *Diccionari de Neurociència* de A. Valero-Cabré *et al.*, editado por el TERMCAT en 2012. Edición en línea: http://www.termcat.cat/ca/Diccionaris_En_Linia/140/. Se trata de una definición mayoritariamente aceptada, que refleja también el *Trésor de la Langue Française* (s.v. *neurone*): 'Cellule nerveuse qui est l'unité fondamentale morphologique et fonctionnelle du tissu nerveux, formée d'un corps cellulaire, d'un prolongement constant, l'axone, et de prolongements inconstants, les dendrites'.

como tal en las investigaciones de Cajal de finales de los años ochenta del s. XIX. De hecho, una de las características más interesantes de nuestro corpus *cajal1,* en lo que se refiere al concepto 'neurona', es que en los artículos que lo integran va afianzándose ese concepto de unidad funcional del sistema nervioso, representado en la mayor parte de las ocasiones por el sintagma "célula nerviosa". La que posteriormente sería denominada "Doctrina de la neurona" (o neuronal) de Cajal, supuso, como ya hemos comentado, la ruptura con el reticularismo y la asunción de una nueva concepción de la célula nerviosa, su forma y sus funciones. La teoría imperante en el estudio del sistema nervioso en el momento en que Cajal comienza sus investigaciones es una teoría continuista con la tradición que señalaba la unión entre los elementos del sistema nervioso[15]. Gerlach y Golgi (quien compartió el Nobel con Cajal) son los puntos fundamentales en la defensa del reticularismo, es decir, la concepción de una estructura que forma una red, y entre cuyas unidades se da una continuidad. Cajal descubre con sus investigaciones, y precisamente aprovechando el método desarrollado por Golgi, que se puede demostrar la individualidad de unas células nerviosas que se comunican con otras células nerviosas por contigüidad; o, lo que es lo mismo, que hay una proximidad entre unas partes de unas células y las de otras, y que entre ellas se transfiere algún tipo de impulsos. Pero no hay unión entre ellas, sino contacto.

Los resultados que hemos encontrado en *cajal1* son los siguientes:

- No aparece el término neurona ni tampoco sus partes con las denominaciones que tienen en la actualidad (axón o neurita, dendritas, …)
- Las únicas formas que encontramos con el formante neuro- son neuroglia (11 apariciones), y sus derivados (neuróglico, -ca, con 67 apariciones en sus distintas formas; y neurogliforme, que aparece en 4 contextos).
- Los textos permiten ver cómo Cajal tenía una clara conciencia del concepto 'neurona' y de sus partes, pues como hemos dicho está descubriendo su carácter de "unidad" morfológica y funcional, y de cómo esa unidad se relaciona con otras de su misma clase. Pero ese concepto está expresado siempre con el sintagma célula nerviosa, que aparece en 23 ocasiones.

De hecho, en la documentación que hemos manejado, tanto de textos originales de Cajal pertenecientes a esos años (convenientemente digitalizados para realizar búsquedas fiables), como en los textos de referencia que analizan tanto su obra como sus descubrimientos principales, no se indica en cuál de sus

[15] Laín Entralgo (1946:84) enlaza esa tradición investigadora con los postulados continuistas de la sustancia nerviosa, cuyo inicio está en Galeno. Unos años antes de Cajal esa tendencia estaba representada por el "primitivo reticularismo protoplasmático de Gerlach (1871) y la más compleja red axónica de Golgi (1886)" (*ibid.,* 86).

artículos aparece por primera vez el término *neurona*. Al no poder disponer de todas las obras de Cajal en soporte informático no podemos aventurar una primera datación de esa forma, pero sí estamos en condiciones de señalar que Cajal se hizo eco muy pronto de la "feliz idea" que, según sus propias palabras, tuvo el neurólogo alemán Heinrich Wilhelm Gottfried Waldeyer (1836-1921) al proponer en 1891 el término *neuron*, formado a partir del griego, para referirse a las células nerviosas.

En los textos que recogen las conferencias que dictó en la Clark University (Worcester) entre 1889 y 1899 con el título "Comparative study of the sensory areas of the human cortex", Cajal escribe en la primera conferencia dos veces el término *neuron*, en inglés.

En otras conferencias que pronunció en la Academia y laboratorio de ciencias médicas de Cataluña en marzo de 1892, y que fueron publicadas en la *Revista de Ciencias Médicas de Barcelona* (números. 16, 20, 22 y 23, de 1892, tomo XVIII) con el título *Nuevo concepto de la Histología de los centros nerviosos*, podemos leer la que para nosotros es la primera referencia que Cajal hace a *neurona* en español: "Los elementos nerviosos representan verdaderas unidades celulares, o *neuronas*, según la expresión de Waldeyer"[16].

En 1906 debía de ser ya una denominación tan habitual para Cajal, que este titula su conferencia para la recepción del Premio Nobel *Estructure et conexions des neurones*, y en el texto de la conferencia aparece constantemente la forma *neurone* y alguno de sus derivados como *neuronal*.

En el subcorpus *cajal2* el uso de *neurona* se convierte en algo completamente normal, lo que se ve confirmado por el hecho de que este término se repita en 99 ocasiones. Una normalidad reafirmada por el comportamiento productivo de ese término, que genera una buena cantidad de unidades neológicas, unas veces mediante el mecanismo de la derivación, y otras con la formación de términos compuestos a partir de la base de composición *neuro-*. Formas como *antineuronista* (3), *neurita* (31), *neurobiología* (1), *neurofibrilla* (66), *neurofibrillar* (78), *neurogénesis* (2), *neuro-génesis* (1), *neurogenética* (5), *neurogenia* (1), *neurogénicas* (2), *neurogenistas* (1), *neurología* (3), *neurológico* (3), *neurólogo* (16), *neuronal* (51), *neuronismo* (12), *neuronista* (2), *neuroplasma* (9), *neuroplasmático* (2), *neurosomas* (1), *neurotrópico* (1) y *neurotropismo* (1). Y junto a ellas otras que ya habíamos visto en textos anteriores de Cajal como *neuroblasto* (3), *neuroglia* (4) y *neuróglico* (9).

En este corpus aparece además *axón* en 91 ocasiones, y también distintos derivados de esta forma, mientras que solo hemos encontrado 7 casos de la que

[16] Citamos por una edición del año siguiente: *Nuevo concepto de la Histología de los centros nerviosos*, Barcelona, Imprenta de Henrich y cía en comandita, sucesores de N. Ramírez, 1893, p. 62.

era su denominación anterior en *cajal1*, *cilindro-eje*.[17] Aparecen también otras partes de la neurona, como *dendrita* (53 casos), y algunos derivados suyos.

La concepción de la neurona en el *corpus2* como unidad mínima del sistema nervioso apenas ha sufrido variaciones desde los textos de Cajal, que fue quien, como ya se ha visto, reinterpretó su naturaleza y su funcionamiento en el organismo.

> El continuo flujo de información de los receptores se organiza en el cerebro mediante percepciones (algunas de las cuales se almacenan en la memoria para futuras consultas) y después mediante respuestas conductuales apropiadas. Todo esto se logra en el cerebro empleando las neuronas y las conexiones entre ellas. Cada neurona, *la unidad básica del cerebro*, es relativamente simple en cuanto a su morfología. (Kandel, 1991: 19; la cursiva es nuestra)

Para observar el tratamiento lingüístico que se hacía del núcleo temático 'neurona' en el corpus de textos actuales, seleccionamos todas las unidades que aparecían en él que contuvieran el formante *neuro-* en su estructura formal. En total, encontramos 183 unidades que compartían esta característica, de las que 110 estaban en español, 69 en inglés, 3 en francés y 1 en catalán. La alta proporción de unidades en inglés en un conjunto de textos redactado en español, y cuyos destinatarios son hablantes de esa misma lengua, parecía confirmar *a priori*, y de una manera bastante evidente, el peso que la producción científica redactada en lengua inglesa tiene sobre los demás, en consonancia con el cambio de dinámica en los intercambios lingüísticos entre científicos que, como ya hemos visto, tuvo lugar a mediados del s. XX: del plurilingüismo propio de los tiempos de Cajal se pasaba, pues, al predominio de una sola lengua.

Sin embargo, al rastrear de manera aleatoria algunas unidades comprobamos que, contrariamente a lo que podía parecer en un principio, una gran proporción de las que estaban escritas en inglés se encontraban en los resúmenes de los artículos que constituían el corpus o en citas puntuales insertas en el cuerpo del texto. La distribución de esas unidades era, pues, muy uniforme y respondía a los cánones actuales para la redacción de artículos científicos, en los que es muy habitual que el resumen que se incluye antes del cuerpo del texto esté redactado en inglés, independientemente de que el trabajo

---

[17] Como bien sabemos, la variación terminológica es constante en los períodos de asentamiento de las unidades terminológicas nuevas. Todavía en 1916 en *A practical medical dictionary* (by Thomas Lathrop Stedman, 4ª ed., New York, William Wood and company) s.v. *neuron* dice que entre sus partes está el "*axis-cilinder* process o *neuraxon*". En el *Larousse Médical Illustré* (ed. de 1921 del Dr. Galtier-Boissière) las únicas denominaciones que aparecen para el 'axón' son las cajalianas "*prolongements*" (prolongaciones), a veces acompañadas de un adjetivo: "*prolongement cylindraxile*" (s.v. *neurone*). En *cajal1* aparecen referencias constantes tanto a *cilinder* como a *cilindro* y *cilindro-eje*.

vaya a ser publicado en una revista destinada para comunidades de especialistas no anglohablantes. Esta realidad, que tiene como claro objetivo atraer la atención de los lectores, recuerda a lo afirmado por Ciapuscio (2003: 54), cuando decía que "la relevancia que determina la inclusión o no de determinadas informaciones se vincula con la funcionalidad del texto, pero también con factores relativos a la situación comunicativa". Comprobábamos, en definitiva, que una tipología textual determinada, marcada por factores que no tenían por qué ser lingüísticos, condicionaba significativamente nuestro proceso de detección de unidades.

Por otro lado, las unidades que aparecían en inglés tenían en la gran mayoría de los casos un equivalente en español dentro del corpus. Además, dejando a un lado los vocablos formados con *neuro-* y atendiendo al cómputo global de unidades que constituían el corpus, el peso del inglés se veía relativizado también por el hecho de que, de las 150 formas con una mayor frecuencia, sólo 19 estaban en inglés. A su vez, de esas 19, sólo 3 aparecían en contextos en español (*Alzheimer Disease* y dos siglas: *MMSE* y *CADASIL*).

En los 150 primeros puestos del listado generado en nuestros análisis aparecen las siguientes unidades escritas en una lengua diferente del español:

| Orden (por frecuencia) | Unidad | Apariciones | Textos | Contexto en español? |
|---|---|---|---|---|
| 2 | *Alzheimer* | 952 | 35 | Sí |
| 9 | *Disease* | 596 | 35 | Sí |
| 17 | *Dementia* | 368 | 31 | No |
| 21 | *Cognitive* | 338 | 31 | No |
| 34 | *MMSE* | 225 | 18 | Sí |
| 40 | *Study* | 205 | 33 | No |
| 41 | *Impairment* | 204 | 27 | No |
| 46 | *Choline* | 197 | 1 | No |
| 48 | *Patients* | 196 | 31 | No |
| 54 | *Brain* | 186 | 23 | No |
| 63 | *Mild* | 166 | 23 | No |
| 70 | *Clinical* | 159 | 30 | No |
| 86 | *Treatment* | 134 | 20 | No |
| 92 | *Psychiatry* | 123 | 30 | No |
| 97 | *Neurology* | 119 | 34 | No |
| 99 | *Citicoline* | 117 | 1 | No |
| 109 | *Health* | 112 | 27 | No |
| 110 | *Stroke* | 111 | 15 | No |
| 121 | *CADASIL* | 103 | 2 | Sí |

La hegemonía del inglés resultaba visible si interpretábamos esa inclusión de subtextos ajenos al idioma de los especialistas que producían y recibían los trabajos como un medio de legitimación de sus investigaciones: si el inglés se

considera como la *lingua franca* para los intercambios de información científica, la inclusión de partes del trabajo redactados en ese idioma facilitaría, en principio, la difusión del artículo y su aceptación en el contexto internacional como un trabajo relevante.

Sin embargo, y a pesar de lo evidente que resultaba el uso de subtextos en inglés como medio legitimador de las publicaciones escritas en español, los datos confirmaban un comportamiento recurrente a este respecto en lo que concernía a la distribución de información en una estructura textual determinada, pero no corroboraban el peso del inglés en la creación de términos en el español actual en el ámbito de las neurociencias. ¿Dónde encontrar ese poso del inglés que la realidad de la dinámica científica actual nos indicaba que existía? En la traducción.

El siguiente paso fue, por tanto, rastrear hasta qué punto las unidades en las que se encontraba el formante *neuro-* en español tenían un equivalente en inglés cuya existencia fuera anterior en el tiempo, como indicio de estar frente a traducciones de términos existentes en inglés, lo que también podía ser un factor que confirmara la hipótesis de dependencia lingüística. Con ese objetivo, rastreamos las 110 unidades en español en los dos manuales de apoyo con los que contábamos. En primer lugar, las buscamos en el manual de Kandel, (escrito en 1991, en su cuarta edición traducida al español de 2001), y comprobamos que 77 de ellas constaban en el volumen. A continuación buscamos en el manual de Delgado, escrito originalmente en español en 1998, las 33 que no aparecían en la traducción del manual de Kandel, y encontramos tan sólo 10 en este segundo manual. Es decir, a pesar de lo limitado de los resultados y de estar relativizados por el factor cronológico (entre la publicación del manual de Delgado y la de los textos que constituían el corpus median doce años), constatamos la baja proporción de términos con el formante *neuro-* que aparecían en la obra publicada originalmente en español pero que no constaban en la obra traducida, lo que nos pareció un claro indicio de la importancia de la traducción del inglés como recurso de formación de las unidades con el formante *neuro-* en el español actual. Las unidades que aparecían en el manual de Delgado pero que no constaban en el de Kandel traducido al español eran las siguientes: *neurofisiológico, -ca; neuroinmunológico, -ca; neurotóxico, -ca; reurorreceptor; neurovegetativo; neurita; neurofarmacología*.

Vemos, en conclusión, el predominio de la importación de unidades con el formante *neuro-,* aun teniendo en cuenta que estas se incorporan traducidas, como procedimiento para la inclusión de neologismos terminológicos en español dentro del ámbito que nos ocupa. Una importación que deja entrever una relación de dependencia denominativa con respecto del inglés que poco tiene que ver con la manera de enriquecer la lengua científica que llevó a cabo Ramón y Cajal.

### 4.2. La sinapsis

El segundo núcleo temático en el que centramos nuestro análisis comparativo fue el de 'sinapsis', nodo conceptual que engloba a todo lo que atañe a la relación funcional de contacto entre las terminaciones de las células nerviosas, según la definición de este término que proporciona *Dicciomed*.[18]

El punto de partida de esta denominación en el griego σύναψις, con ese significado de 'punto de contacto', 'contacto', se encuentra en los orígenes del concepto, el cual, como en el caso de 'neurona', va tomando forma en los textos que integran nuestros dos corpus de Ramón y Cajal.

En el primero de ellos (*cajal1*) podemos observar cómo, junto a la descripción de las unidades funcionales del sistema nervioso (las *neuronas*), aparecen referencias constantes a la relación funcional de contacto entre las terminaciones de las células nerviosas. Para acometer esta tarea Cajal recurre, como tantas otras veces, a las posibilidades que le proporciona la lengua que conoce bien. Esta le permite tomar distintas voces patrimoniales y moldearlas convenientemente para disponer de los recursos con los que comunicar sus descubrimientos de una manera eficiente. En unos casos lo hará con procedimientos formales, como el añadido de adjetivos, en otros extendiendo las posibilidades significativas de esas unidades que toma como punto de partida:

> (…) las ramitas descendentes de las prolongaciones nerviosas de las células pequeñas de la zona molecular, tienen *contacto íntimo*, no sólo con el cuerpo del elemento de Purkinje sino también con la porción más alta, desnuda de mielina de las espansiones de Deiters. Ahora bien; este fenómeno de relación tan singular […], ¿no parece abonar la hipótesis de la trasmisión por contigüidad de las acciones nerviosas?

Tanto *contacto* (7 apariciones) como *conexión* (46) son dos de las palabras que Cajal carga de significado especializado para aludir a ese concepto nuevo. Su inserción dentro de unidades sintagmáticas como *íntimo contacto*, *contacto íntimo*, indicios de un *contacto celular*, *conexión celular*, *conexión intercelular*, o *conexiones de sus espansiones*, le proporciona los recursos que necesita. Y se trata en ambos casos de voces con unos significados asentados desde el s. XV que difieren muy poco de los actuales, como puede deducirse de los datos aportados por el corpus histórico NDHE.[19]

---

[18] Este diccionario etimológico en su edición de 2011 (consulta en diciembre de 2012) señala su origen en el griego del s. IV a.C. sýnapsis σύναψις , cuyo significado era 'punto de contacto', 'contacto'. Por su parte el *Diccionari de Neurociència* del TERMCAT define este término como "Zona d'intercanvi de senyals nerviosos entre dues neurones o entre una neurona i un múscul o una glàndula de secreció endocrina o exocrina".

[19] La consulta de la edición provisional del corpus del *Nuevo Diccionario Histórico del Español*, dirigido por José A. Pascual, se realiza en la página web de la Fundación Rafael Lapesa (http://web.frl.es/CNDHE/view/inicioExterno.view)

En *cajal2* el término *sinapsis* aparece en 21 ocasiones, utilizado unas veces de manera independiente, y otras como núcleo de diferentes sintagmas terminológicos: *sinapsis neuronales*, *sinapsis interneuronales*, *sinapsis por engranaje*, *sinapsis corticales*, o *sinapsis dendro-nerviosa*.

El concepto está plenamente asumido por Cajal desde principios de siglo, o incluso antes, como bien reflejan sus textos. En investigaciones anteriores ya había desarrollado y explicado ese concepto, y probablemente aportó con sus explicaciones las claves para el desarrollo de la denominación. Sin embargo, a otro le correspondió de nuevo ver asociado su nombre al nuevo avance científico. Fue Ch. S. Sherrington en 1897, en un artículo publicado en la obra de M. Foster *A Text-Book of Physiology*[20], quien utilizó por primera vez esa denominación: "Such a special connection of one nerve cell with another *might be called a synapse*".

A pesar del comportamiento completamente normal que ese término tiene en *cajal2*, se trata de una unidad que no tuvo en Cajal una repercusión y asimilación rápidas, como en el caso de *neurona*. De hecho todavía en 1906, cuando recibe el premio Nobel, no incorpora la nueva denominación a su conferencia, aunque desde las primeras páginas de su discurso puede entreverse que está aludiendo directamente a ese concepto:

> (…) les éléments nerveux possèdent des relations réciproques de contigüité et non de continuité, et que ces rapports de contact plus ou moins intime s'établissent toujours, non entre les arborisations nerveuses seules, mais entre ces ramifications d'une part et le corps et les prolongements protoplasmatiques d'autre part.

Por otro lado, como sucede en el texto anterior, Cajal sigue confiriendo una gran parte de la sustancia significativa a unidades terminológicas que le han ayudado a transmitir ese concepto en otras ocasiones. Así hemos encontrado 35 contextos de *contacto* (*transmisión por contacto*, *intimidad del contacto*, *contacto íntimo*,…) y 71 de *conexión* (*conexión íntima*, *conexión por contacto*, *conexión interneuronal*, además de la forma simple).

En el *corpus2*, el formado por textos actuales, no ha sido frecuente el hallazgo del término simple *sinapsis* (aparece solo en 3 ocasiones), por lo que hemos dirigido nuestra búsqueda hacia compuestos sintagmáticos en los que aparece el adjetivo *sináptico, ca* (*hendidura sináptica*, *liberación sináptica*) y hacia unidades que incluyen el formante *sinapto-*, especialmente productivo en estos textos (*sinaptogénesis*, *sinaptosomal*), como puede verse en la siguiente tabla.

[20] Aparece en la Part III, London: Macmillan.

| Término | Frecuencia Absoluta | Textos |
|---|---|---|
| *Sinapsis* | 3 | 3 |
| *Sináptica*<br>*Plasticidad sináptica*<br>*Transmisión sináptica*<br>*Liberación sináptica*<br>*Hendidura sináptica*<br>*Función sináptica*<br>*Disfunción sináptica*<br>*Reparación sináptica*<br>*Reactividad sináptica* | 14 | 6 |
| *Sinápticas*<br>*Conexiones sinápticas* | 4 | 3 |
| *sináptico* | 1 | 1 |
| *Sinápticos* | 1 | 1 |
| *Sinaptogénesis* | 3 | 2 |
| *Sinaptosomal* | 1 | 1 |
| *Sinaptosomas* | 3 | 1 |
| *Sinaptosómica* | 1 | 1 |
| *Synapses* | 1 | 1 |
| *Synaptic plasticity* | 1 | 1 |
| *Synaptosomal*<br>*Synaptosomal protein*<br>*Synaptosomal membranes*<br>*Synaptosomal transport* | 3 | 1 |
| *Synaptosomes* | 2 | 1 |

Observamos, como sucedía con las unidades formadas con *neuro-*, que la presencia de denominaciones en inglés se reduce también en este caso a los resúmenes de cada artículo y a las citas insertas en estos. Y hemos podido comprobar también que la productividad de los formantes cultos, como es el caso de *sinapto-*, resulta bastante elevada. La tendencia a usar unidades basadas en este formante es clara, y confirma la generalización de un procedimiento que vimos claro en *cajal2* para el caso de *neuro-*. Sin embargo su introducción debió de ser posterior a 1933, pues en Cajal no hemos encontrado ninguna unidad formada a partir de *sinapto-*.

Por otro lado realizamos un segundo rastreo basado esta vez en la búsqueda de las denominaciones más utilizadas por Cajal para hacer referencia a las conexiones neuronales. Como ya hemos visto, estas giraban en torno a la dimensión del 'contacto' y de la 'conexión'. Pues bien, esa diversidad denominativa presente en Cajal se pierde y, por tanto, frente a esa pluralidad de soluciones terminológicas recogida por el científico español a principios del siglo XX para explicar las conexiones interneuronales se pasa a una simplicidad denominativa en la que la referencia a las nociones de 'contacto' y 'conexión' se difumina casi por completo. Frente a la profusión de sintagmas en *cajal1*, como ya hemos visto en los apartados anteriores (*conexión axo-dendrítica*, *conexión axo-somática*, *conexión interneuronal*, *conexión por engranaje*, etc.), en el corpus de la *Revista*

*Española de Neurología* encontramos los sintagmas *conexión numérica*, *conexión episódica* y *conectividad cerebral*, además de los únicos que son coincidentes: *conexiones neuronales* y *conexiones sinápticas*.

## 5. Algunas apreciaciones

Los análisis realizados sobre los dos corpus nos han permitido recordar la situación que era más habitual en la comunicación científica a finales del s. XIX, y que perduró hasta no hace más de cincuenta años, aunque su evolución pudiese verse algo antes: un espacio para el plurilingüismo en el que se producía la convivencia de varias lenguas en los intercambios especializados entre científicos. En esa situación, que vemos perfectamente reflejada en los textos de Cajal, era normal que las citas de artículos apareciesen en distintas lenguas, de la misma manera que lo era que los neologismos terminológicos se construyesen sobre unas bases comunes que propiciaban un acercamiento entre las lenguas de especialidad. Denominaciones en una lengua en las que puede observarse su cercanía respecto a las de las otras lenguas internacionales, pero que, no obstante, a veces chocan con las dificultades propias de códigos lingüísticos diferentes. Y en esos casos lo que hemos observado es el interés de los científicos por llegar a la comprensión de algo en lo que están realmente interesados. Unas veces lo consiguen, aunque sea con cierto esfuerzo, pero en otras ocasiones no, y en esos casos se produce el recurso a la traducción de los textos científicos.

Los escritos de Cajal tuvieron una repercusión importantísima en la ciencia de su tiempo, a pesar de estar escritos en español, y él a su vez pudo evaluar los avances de colegas de otros países, aunque estos escribiesen en francés, inglés o alemán. A. Kölliker, uno de los histólogos más importantes de su tiempo, está dispuesto primero a aprender español, y después a traducir del español al alemán el trabajo cajaliano sobre el asta de Ammon. Y lo mismo le sucede al científico sueco G. Retzius, quien manifiesta su deseo de aprender español al sentirse desilusionado por no poder leer correctamente los artículos de Cajal[21]. En un sentido diferente observamos cómo los textos de Cajal se encuentran repletos de alusiones a los avances de científicos que escriben en otras lenguas. Los cita, escribe las referencias de sus artículos en la lengua en que están escritos, y los incorpora a su discurso de una manera natural. En aquellos momentos se produjeron, como es lógico, problemas en la comunicación, diferentes de los actuales, pero estos no fueron mayores de los que pueden acaecer en una situación de monolingüismo científico como la que impera hoy en día. El verdadero problema para nosotros radica sobre todo en quienes quieren ver con los planteamientos globalizadores actuales hechos de un pasado

[21] En el artículo de Sánchez Ron (2005) se aportan las pruebas textuales de estos dos testimonios, ambos de científicos de primera línea a finales del s. XIX.

en el que los flujos de información eran mucho menores, y en los que la rapidez en la comunicación no era comparable a la de ahora.[22]

La evolución de las denominaciones en Cajal supone una buena muestra de uno de los procesos habituales en la generación neológica en una lengua de especialidad. Se produce en primer lugar la indeterminación inicial en la que conviven varias formas, normalmente sintagmáticas, de tipo descriptivo. Más adelante se tiende a un proceso de asentamiento de una unidad terminológica (o varias, pero en número más reducido), en el que una de ellas se va imponiendo (el sintagma se fosiliza por la repetición), o se genera una forma nueva utilizando formantes grecolatinos, forma simple que tiene muchas más posibilidades de éxito si está bien formada, como ocurrió en los casos de *neurona* y de *sinapsis*. La simplificación posterior de las formas (especialmente de las sintagmáticas) va concediendo más posibilidades a su internacionalización, puede que porque, en un principio, tengan una mayor garantía de éxito las unidades monoléxicas que las unidades sintagmáticas, como pudiera pensarse por el hecho de que los sintagmas propuestos por Cajal se hayan perdido con más facilidad que las unidades con el formante *neuro-,* que son las que finalmente han pervivido.

Uno de los problemas que tienen los científicos actuales hablantes de lenguas románicas es el de encontrarse en clara desventaja con los que tienen el inglés como lengua materna[23]. En el caso de Cajal la situación era muy distinta, y el uso de la lengua en que era plenamente competente le permitió ir explicando y denominando convenientemente cada uno de los avances de sus investigaciones. La posibilidad de establecer con garantías el puente necesario para las denominaciones metafóricas, de caracterizar terminológicamente a una palabra de uso frecuente para extraerle todas sus posibilidades significativas, y, finalmente, de usar las nuevas denominaciones internacionales ya adaptadas, pues le aportan precisión, sin duda contribuyeron a ello. Máxime si además se produjeron en un espacio de multilingüismo en el que la lengua del otro no era minusvalorada.

El cambio en la situación interlingüística[24] entre aquella época y la actual en lo que a comunicación científica se refiere lo hemos podido observar no solo por el contraste entre el corpus antiguo y el actual, sino también por los cambios

---

[22] Un trabajo excelente como el de Gordon M. Shepherd de 1991 (*Foundations of the Neuron Doctrine*) tiene sin embargo algunos errores de apreciación de ese tipo.

[23] En la comunicación de F.A. Navarro "La traducción médica ante el siglo XXI: tres retos para el lenguaje científico en español" se explica perfectamente el entramado actual de la comunicación científica, donde se produce una clara "discriminación lingüística" hacia quienes no tienen el inglés como lengua materna.

[24] Cuando hablamos de situación interlingüística queremos hacerlo con todas las implicaciones que tiene este término, entre las cuales figuran las relaciones de poder o subordinación entre lenguas, y también las actitudes de aprecio lingüístico, prestigio y desprestigio. Véase Loubier (2008) y García Palacios (en prensa).

que se van produciendo entre los dos subcorpus de Cajal, con textos que solo tienen treinta años de diferencia en su producción. En ambos se pone de manifiesto la generación de neologismos con los recursos de la propia lengua, y también la gran cercanía que existe entre los términos que usa Cajal y científicos que escriben en otras lenguas (de., fr., en.)[25], pero en el primer subcorpus el uso de las unidades terminológicas sintagmáticas es mayor, mientras que en el segundo la balanza se inclina hacia la formación de compuestos a la manera culta. En los dos subcorpus se muestra el poder denominativo de Cajal, pero en el primero aparecen con mayor frecuencia las denominaciones que podemos denominar más "cajalianas" (*expansiones*, *prolongaciones*, *cilindro*), mientras que en el segundo van predominando las que conocemos en la actualidad, aunque no se deban a su pluma (*dendritas*, *axón, neurona, sinapsis*). Por otro lado en *cajal1* se observan muchas referencias a textos alemanes, franceses o italianos, y a alguno en inglés, una proporción que se resuelve en *cajal2* a favor de estos últimos, cuyo número aumenta considerablemente.

## 6. Conclusiones

El mantenimiento incuestionado de determinados tópicos en torno a la comunicación científica internacional en nuestros días nos impide con cierta frecuencia detenernos a reflexionar sobre algunas cuestiones especialmente importantes. Las traemos en estos momentos a colación por tratarse también de conclusiones de nuestro estudio.

a) Es posible llevar a cabo la generación neológica en un ámbito científico en una lengua como el español, según lo que nos han mostrado los textos fundadores de las modernas neurociencias, obras de S. Ramón y Cajal. Y puede producirse a pesar de una realidad como la actual, que nos muestra una lengua que solo parece ser neológicamente activa en la formación secundaria de términos, pero no en la primaria (Sager, 1990).

b) La neología terminológica puede producirse y desarrollarse perfectamente en un espacio en el que convivan las denominaciones en varias lenguas. Si era posible el multilingüismo en la comunicación científica y el ámbito de las neurociencias se desarrolló en profundidad en otros momentos, no nos parece insignificante cuestionarnos por qué la situación actual no puede sufrir modificaciones con el fin de facilitar las posibilidades denominativas en lenguas diferentes de la inglesa.

c) La confluencia de lenguas en la neología no tiene por qué ser un factor distorsionador que favorezca la variación terminológica. O al

[25] Por poner un solo ejemplo, en *cajal2* encontramos tanto *fibras musgosas* como *mossy fibres* y *fibres mousseuses*.

menos no tiene por qué serlo en mayor medida que en una situación de monolingüismo para la lengua científica como la actual.

d) El monolingüismo en la lengua científica no responde únicamente al hecho de que una comunidad sea científica y técnicamente dependiente de otra que impone su lengua. De hecho los cauces de esa hegemonía en lo que respecta a la neología terminológica responden a una realidad muy compleja en la que entran en juego conceptos como el de globalización, asimetría lingüística, representación, dependencia terminológica, y prestigio o desprestigio lingüísticos (Humbley y García Palacios, 2012; Loubier, 2008).

A la hora de plantear opciones denominativas en una lengua determinada, la internacionalización de la ciencia y el monolingüismo en torno a ella aconsejan tener en cuenta las características y particularidades de las unidades terminológicas correspondientes en inglés. Es la forma de establecer similitudes y diferencias, poder calibrar las relaciones de dependencia terminológica entre ambas y actuar en consecuencia.

El estudio que hemos realizado nos ha permitido contemplar cómo determinados hallazgos científicos en el ámbito de las neurociencias le correspondieron a Cajal. Y con ellos esos titubeos denominativos que resultan tan expresivos. Sin embargo a otros les correspondió la fortuna de la denominación acertada con visos de internacionalidad. Así, los nombres de Waldeyer y de Sherrington, a causa de ese "power of the word in science", que crea "a symbol for a new concept" (Shepherd, 1991: 211), permanecieron indisolublemente unidos a *neurona* y *sinapsis*, esos dos grandes descubrimientos de Santiago Ramón y Cajal.

No es, sin embargo, esa adscripción de términos lo que en estos momentos puede resultar más preocupante.

Las conclusiones que hemos querido plantear desde una manifiesta ingenuidad nos han permitido constatar de nuevo la realidad lingüística actual, y los peligros que entraña para lenguas como las románicas con carácter internacional que, a pesar de su extensión demográfica, se ven envueltas en una situación de asimetría lingüística en la lengua científica. No hay muchos pasos de ahí a una verdadera situación diglósica, con lo que eso significa de empobrecimiento, y de peligro para la supervivencia lingüística de esas lenguas en todas las situaciones de comunicación.

## Referencias bibliográficas

CABRÉ M.T. (1999), *La terminología. Representación y comunicación. Una teoría de base comunicativa y otros artículos,* Barcelona, IULA, UPF.

CABRÉ M.T (2008), "El principio de poliedricidad: la articulación de lo discursivo, lo cognitivo y lo lingüístico en Terminología", *Ibérica*, 16, 9-36.

CANDEL D., GAUDIN F. (dir. 2006), *Aspects diachroniques du vocabulaire*, Mont-Saint-Aignan, Publications des Universités de Rouen et du Havre.

CIAPUSCIO G. (2003), *Textos especializados y terminología*, Barcelona, IULA, UPF.

DE FELIPE J. (2007a), "Historia de la neurona: Influencia de los estudios de Santiago Ramón y Cajal en la neurociencia moderna", En *Histología del sistema nervioso del hombre y de los vertebrados. Santiago Ramón y Cajal,* Tomo I, Ministerio de Sanidad y Consumo, CSIC, BOE, Madrid, 19-99.

DE FELIPE J. (2007b), "Cajal y la neurociencia del siglo XXI", En DE FELIPE J. *et al. Paisajes neuronales*, Madrid, CSIC-Instituto Cervantes, 41-97.

DELGADO J.M. *et al.* (1998), *Manual de neurociencia*, Madrid, Síntesis.

*Dicciomed. Diccionario médico-biológico, histórico y etimológico*, Dir. CORTÉS GABAUDÁN F., Edición en línea: http://dicciomed.eusal.es

DURY P., PICTON A. (2009), « Terminologie et diachronie : vers une réconciliation théorique et méthodologique », *Revue Française de Linguistique Appliquée*, Vol. XIV, 2009-2, 31-41.

GARCÍA PALACIOS J. (2009a), "El trabajo sobre la neología terminológica en español". En CABRÉ M.T. *et al.* (eds.), *Actes del I Congrés Internacional de Neologia de les Llengües Romàniques*, Barcelona, Institut Universitari de Lingüística Aplicada, 555-566.

GARCÍA PALACIOS J. (2009b), "La competencia neológica especializada en el estudio y la actuación sobre la neología terminológica", *Revue Française de Linguistique Appliquée*, Vol. XIV, 2009-2, 17-30.

GARCÍA PALACIOS J. (2010), "Ruptura y recuperación del equilibrio terminológico en el ámbito de la archivística: entre la innovación, la regularización y el diccionario de especialidad", En J. GARCÍA PALACIOS (dir.) *La terminología de la archivística*, Gijón, TREA, 2010, 13-40.

GARCÍA PALACIOS J. (2013), "Neología y traducción especializada: ¿un mismo espacio de actuación?", En ALVES, I.M. (dir.), *Neologia nas línguas románicas*, São Paulo: Humanitas, 2013 (en prensa).

GARCÍA PALACIOS J., SANZ VICENTE M.L. (2012), "El papel de la traducción en la formación secundaria de términos sintagmáticos", *Neologica*, 6, 95-109.

GUTIÉRREZ RODILLA B. (1998), *La ciencia empieza en la palabra*, Barcelona, Península.

HUMBLEY J., GARCÍA PALACIOS J. (2012), "Neology and Terminological Dependency", *Terminology*, 2012, 18:1, 59-85.

KANDEL E., SCHWARTZ J., JESSELL T. (1991), *Principles of Neural Science,* New York, Elsevier.

KANDEL E., SCHWARTZ J., JESSELL T. (2001), *Principios de Neurociencia*, Trad. de AGUD J. *et al.*, Madrid, McGraw-Hill Interamericana de España.

LAÍN ENTRALGO P. (1946), "La obra de Cajal", En edición de PUERTA J.L., *Escritos sobre Cajal*, Madrid, Triacastela, 2008, 65-94.
LÓPEZ PIÑERO J.M. (2006), *Santiago Ramón y Cajal*, Valencia, Publicacions de la Universitat de València.
LOUBIER CH. (2008), *Langues au pouvoir*, Paris, L'Harmattan.
MARTÍN-ALMAGRO A. (2007), *Terminología, método científico y estadística aplicada en ciencias de la salud,* Morphos.
MORENO FERNÁNDEZ A. M. *et al.* (2009), *Cajal y su obra*, Sevilla, Universidad de Sevilla.
NAVARRO F.A. (2001), "La traducción médica ante el siglo XXI: tres retos para el lenguaje científico en español", En Congresos de la lengua española. Edición en línea en: http://congresosdelalengua.es/valladolid/ponencias/nuevas_fronteras_del_espanol/1_la_traduccion_en_espanol/navarro_f.htm (consulta, diciembre 2012).
PEARSON J. (1998), *Terms in context,* Amsterdam, John Benjamins Publishing Company.
PICTON A. (2009), *Diachronie en langue de spécialité. Définition d'une méthode linguistique outillée pour repérer l'évolution des connaissances en corpus. Un exemple appliqué au domaine spatial,* Toulouse, Universidad de Toulouse.
RAMÓN Y CAJAL S. (1889), "Conexión general de los elementos nerviosos", *La Medicina Práctica*, 88, 341-346.
RAMÓN Y CAJAL S. (1888-1889), *Revista trimestral de Histología Normal y Patológica*, Barcelona, Tipografía de la Casa provincial de caridad, 1-2.
RAMÓN Y CAJAL S. (1899), *Textura del sistema nervioso del hombre y de los vertebrados*, Madrid, Imprenta y Librería de Nicolás Moya.
RAMÓN Y CAJAL S. (1933), *¿Neuronismo o reticularismo? Las pruebas objetivas de la unidad anatómica de las células nerviosas*, Archivos de Neurobiología, t. XIII (2, 4, 5, 6), Madrid, 1933.
SAGER J. C. (1997), "Term formation", En WRIGHT S. E., y BUDIN G. (eds.) *Handbook of Terminology Management*, Amsterdam, John Benjamins, vol. 1, 25-41.
SÁNCHEZ RON J.M. (2005), "El español y la ciencia: el ejemplo de Ramón y Cajal". *Panace@,* vol. VI, nº 21-22, 2005, 201-203.
SHEPHERD G. M. (1991), *Foundations of the Neuron Doctrine*, New York/Oxford, Oxford University Press.
TELLO J. F. (1935), *Cajal y su labor histológica*, Madrid, Tipografía artística.
TEUBERT W. (1996), "Comparable or parallel Corpora?", *International Journal of Lexicography*, 9 (3), 238–264.

# Le traitement de la néologie technique dans la neuvième édition du *Dictionnaire de l'Académie Française*

Christophe Rey[1], Amandine Delacroix[2]

## Résumé

Cet article aborde la question du traitement de la néologie technique au sein du répertoire bien particulier que constitue le Dictionnaire de l'Académie Française, ouvrage caractérisé par un rythme de publication particulièrement prudent, mais ayant aussi historiquement fait le choix de ne pas traiter les lexiques spécialisés.

## Mots clés

Néologie, Dictionnaires, Académie Française.

## Introduction

Dans le cadre de cette communication, nous avons choisi de nous intéresser à la question de la néologie en langue de spécialité au sein d'un répertoire tout particulier : le *Dictionnaire de l'Académie Française* (désormais *DAF*).

L'originalité de notre analyse réside dans la nature même du dictionnaire observé dans la mesure où celui-ci, depuis sa première édition (1694), prétend opérer une séparation entre la « langue commune » et les lexiques de spécialités[3].

Pourtant, une consultation même rapide des différentes éditions du répertoire des Académiciens suffit pour mesurer l'existence massive de termes proprement techniques, d'ailleurs le plus souvent rattachés explicitement à leur domaine d'origine (ex : « Armurier », « Astrophysique », « Automobile », « Banque », etc.) (Cf. BOUVEROT D. (1998)).

À travers une analyse articulée autour de l'examen diachronique du discours des Académiciens, nous montrerons dans un premier temps que le *DAF* a très tôt été contraint de s'ouvrir aux termes spécialisés pour recenser avec fidélité

[1] LESCLAP (CERCLL-EA 4283), Université de Picardie Jules Verne, christophe.rey@u-picardie.fr

[2] LESCLAP (CERCLL-EA 4283), Université de Picardie Jules Verne, amandine.delacroix@etud.u-picardie.fr

[3] Les mots relevant des langues de spécialités feront l'objet d'une publication la même année (1694) dans le *Dictionnaire des Arts et des Sciences* de Thomas Corneille.

l' « usage ». Conditionnée par le rythme même d'élaboration du dictionnaire – la huitième édition du *DAF* a été achevée en 1935 et la neuvième (toujours en cours) a commencé à paraître en 1992 –, cette contrainte d'ouverture, nous le verrons en ne nous focalisant dans un second temps que sur la dernière édition, est encore renforcée par les contraintes grandissantes de la dictionnairique, dimension poussant les lexicographes modernes à se tourner - en plus de ces lexiques spécialisés - vers l'intégration de régionalismes linguistiques de France et de mots de la Francophonie.

Ainsi que nous nous attacherons à le montrer enfin, le concept même de néologismes spécialisés dans les dictionnaires de langue, qu'on les considère comme de simples « usages techniques », des « termes »[4] ou encore des « technolectes »[5], mérite d'être remis en question. En effet, en dépit de leur caractère obsolescent[6] et parfois éphémère[7], les lexiques spécialisés sont la plupart du temps enregistrés par les dictionnaires généraux[8], y compris le *DAF*, mais avec un décalage temporel plus ou moins important.

## 1. De 1986 à aujourd'hui, une neuvième édition en voie de finalisation et dotée d'innovations significatives

Avant de nous lancer dans l'examen que nous nous sommes proposé de livrer ici, il nous semble important de nous attarder quelque peu sur une présentation succincte du *DAF*.

Seul dictionnaire de l'institution avec le *Trésor de la Langue Française*, le *DAF* est avant tout caractérisé par sa longévité exceptionnelle. À la suite de sa première version parue en 1694, quatre éditions vont en effet en être publiées au XVIII^e^ (1718, 1740, 1762, 1798), deux au XIX^e^ (1835, 1878) et une au XX^e^ (1932-35). Le *DAF* existe donc depuis cinq siècles et en est aujourd'hui à sa neuvième édition.

En choisissant d'agencer les entrées de leur répertoire par racines, en prenant le parti de ne pas illustrer leurs définitions par des citations d'auteurs ou en décidant enfin de ne pas inclure les termes des arts et des sciences mais uniquement ceux de la langue commune, les Académiciens vont conférer une véritable identité à leur dictionnaire. De fait, cet ouvrage, tout en évoluant au fil des siècles, va occuper une place privilégiée au sein du paysage lexicographique français.

---

[4] Au sens de la discipline Terminologie.

[5] Cf. MESSAOUDI L. (2002).

[6] Cf. REY C. (À paraître).

[7] Préface de la huitième édition du *DAF* (1932-35).

[8] Pour plus de détails concernant les raisons pour lesquelles les grands dictionnaires de langue modernes s'ouvrent massivement aux termes spécialisés, nous renvoyons à REY C. (À paraître).

Ayant commencé à paraître en 1992, l'édition que nous avons choisi d'évoquer pour notre étude est encore à ce jour en cours de rédaction et apporte un lot de changements importants dont les plus significatifs se trouvent listés ci-dessous[9] :

1) Introduction du dispositif des remarques normatives
2) Ouverture à la variation linguistique régionale
3) Ouverture plus grande à la francophonie
4) Introduction de l'étymologie
5) Introduction de titres d'œuvres célèbres
6) Adaptation à la modernité par sa mise à disposition sur Internet

Ajoutons à ces différentes innovations le retour à deux bureaux du dictionnaire, à savoir le « service du dictionnaire » et la « commission du dictionnaire », travaillant conjointement à ce chantier lexicographique sur des volumes différents, notamment pour traiter le flux lexical important découlant des révolutions techniques et scientifiques modernes[10].

Pour revenir à la thématique que nous nous sommes proposé d'examiner dans cette communication, mentionnons également le fait que cette dernière édition du *DAF* se caractérise par l'introduction de nombreux mots nouveaux. Si nous reprenons les chiffres communiqués par l'Académie elle-même, le premier tome de cette neuvième édition (de A à Enzyme) renfermerait 14 024 mots, dont 5 500 mots nouveaux, le deuxième tome (de Éocène à Mappemonde) environ 11 500 mots, dont 4 000 mots nouveaux et enfin le troisième tome (de Maquereau à Quotité) 9 860 mots, dont 3 828 mots nouveaux.

La quantité de mots introduits dans cette dernière édition du répertoire est assez importante, mais doit néanmoins être relativisée en considérant le rythme d'élaboration de ce dictionnaire et donc la période généralement assez importante qui sépare deux éditions.

Si nous examinons à présent la nature des entrées nouvellement introduites, l'Académie communique explicitement sur ce point et indique une ouverture vers de nombreux registres de langue et domaines. C'est en tout cas ce qu'atteste la liste reproduite ci-après pour le troisième tome du *DAF* : 1) Usage courant, 2) Sciences/Techniques, 3) Société/Politique, 4) Emplois familiers, 5) Emplois populaires, 6) Termes vulgaires ou triviaux, 7) Termes argotiques.

---

[9] Pour la conception du *Dictionnaire de l'Académie Française* en tant que répertoire lexicographique en évolution, nous renvoyons à REY C. (2012a) et REY C. (2012b).

[10] Sur ce point, rappelons que l'Académie française valide le travail de commissions de terminologie (adaptation lexicale des termes technologiques étrangers) créées dans différents ministères (20 environ).

La présence du domaine des « sciences/techniques » dans cette liste confirme ainsi que le programme d'accroissement de l'usage défendu par l'Académie passe par le recours aux technolectes.

## 2. Le discours de l'Académie Française sur les lexiques spécialisés : examen diachronique

Après cette rapide présentation de la dernière édition du *DAF*, nous avons choisi de débuter notre étude proprement dite en examinant le ou les discours de l'institution sur ce réservoir lexical particulier que sont les technolectes. Nous avons pour cela examiné les différentes préfaces du *DAF* à travers les siècles et pu dégager deux grands positionnements adoptés à ce propos.

### 2.1. Le Dictionnaire de l'Académie Française : recenser l'usage

Ainsi que nous l'avons souligné en introduction, historiquement, les Académiciens ont choisi d'adopter au sein de leur répertoire une description de la « langue commune »[11], une langue qui, comme l'illustre l'extrait reproduit ci-dessous, exclut les termes des arts et des sciences :

> C'est dans cet estat où la Langue Françoise se trouve aujourd'huy qu'a esté composé ce Dictionnaire ; et pour la representer dans ce mesme estat, l'Académie a jugé qu'elle ne devoit pas y mettre les vieux mots qui sont entierement hors d'usage, ni les termes des Arts et des Sciences qui entrent rarement dans le Discours ; Elle s'est retranchée à la Langue commune, telle qu'elle est dans le commerce ordinaire des honnestes gens, et telle que les Orateurs et les Poëtes l'employent ; Ce qui comprend tout ce qui peut servir à la noblesse et à l'Elegance du discours. Elle a donné la Definition de tous les mots communs de la Langue dont les Idées sont fort simples ; et cela est beaucoup plus mal-aisé que de definir les mots des Arts et des Sciences dont les Idées sont fort composées. (*DAF*, 1694)

Ce positionnement lexicographique sera conservé mais adapté au fil des éditions, notamment en raison de la pénétration de termes techniques dans cette langue commune que le *DAF* entend décrire :

> Pour ce qui est des termes d'Art, l'Académie a cru ne devoir admettre dans son Dictionnaire que ceux qui sont extremement connus et d'un grand usage, à moins qu'ils ne soient amenez par le mesme mot de la langue, qui a dans la langue une signification differente ; par exemple, à la suite du mot travail, qui signifie labeur, peine, etc. on trouve travail, qui signifie, une machine qui sert aux Maréchaux pour contenir les chevaux difficiles à ferrer. (*DAF*, 1718)

[11] BRANCA-ROSOFF S., FOURNIER J.-M., GRINSHPUN Y., REGENT-SUSINI A. (eds). (2011)

Les troisième et quatrième éditions du *DAF* marquent pour leur part une ouverture de plus en plus large de la langue commune aux termes issus des arts et des sciences :

> L'Académie a jugé encore à propos de n'y faire entrer que ceux des termes d'art et de science que l'usage a introduits dans la Langue commune, ou ceux qui sont amenez par quelque mot de cette même Langue. Ainsi à la suite de Parabole, qui signifie une Allégorie sous laquelle on cache quelque vérité importante, on trouvera Parabole, terme de Géométrie et qui signifie une certaine ligne courbe.
>
> […] Le Public ne manquera pas de remarquer qu'il se trouve dans la nouvelle Edition, un bien plus grand nombre de termes d'art et de science, que dans les deux précédentes. Nous ne nous sommes pas écartez néanmoins de la règle que nos Prédécesseurs s'étoient prescrite, de n'admettre que ceux de ces termes qui sont d'un usage si général, qu'ils peuvent être regardez comme faisant partie de la Langue commune, ou qui sont amenez par un mot de cette Langue. Mais depuis environ soixante ans qu'il est ordinaire d'écrire en François sur les arts et sur les sciences, plusieurs termes qui leur sont propres, et qui n'étoient connus autrefois que d'un petit nombre de personnes, ont passé dans la Langue commune. Auroit-il été raisonnable de refuser place dans notre Dictionnaire, à des mots qui sont aujourd'hui dans la bouche de tout le monde ? (*DAF*, 1740)
>
> Les sciences et les arts ayant été plus cultivés et plus répandus depuis un siècle qu'ils ne l'étoient auparavant, il est ordinaire d'écrire en François sur ces matières. En conséquence plusieurs termes qui leur sont propres, et qui n'étoient autrefois connus que d'un petit nombre de personnes, ont passé dans la Langue commune. Auroit-il été raisonnable de refuser place dans notre Dictionnaire à des mots qui sont aujourd'hui d'un usage presque général ? Nous avons donc cru devoir admettre dans cette nouvelle Edition, les termes élémentaires des sciences, des arts, et même ceux des métiers, qu'un homme de lettres est dans le cas de trouver dans des ouvrages où l'on ne traite pas expressément des matières auxquelles ces termes appartiennent. (*DAF*, 1762)

La première tendance que nous relevons au sein du *DAF* est donc celle qui consiste historiquement à séparer lexique commun et lexiques spécialisés. Sans pour autant l'obliger à déroger à ce principe, l'évolution des arts et des sciences a néanmoins ouvert l'Académie à ouvrir progressivement son répertoire aux lexiques spécialisés.

### 2.2. L'Académie et l'obsolescence des vocabulaires spéciaux

La seconde tendance observée au sein du discours préfacier de l'Académie Française renforce dans une certaine mesure le positionnement de principe qu'elle avait historiquement adopté en choisissant de séparer langue commune et langues de spécialités.

Au XIXe siècle, siècle de grand foisonnement lexical[12], tout en ouvrant son dictionnaire aux technolectes, l'Académie Française ne manque pas de mettre en évidence le phénomène d'obsolescence des termes techniques. Mise en évidence par Danielle Bouverot (1998), cette obsolescence est évoquée dans la préface de la huitième édition du *DAF* :

> Aux dernières années du XIXe siècle, quand l'Académie s'occupa de préparer une nouvelle édition de son Dictionnaire, elle se trouva en présence d'une brusque pénétration des vocabulaires des Sciences et des Arts dans le parler de tous [...] De quel artisan, de quel paysan de France restent ignorés des termes tels que microbe, sanatorium, otite, diphtérie, hydravion, commutateur, carburateur, court-circuit ?
>
> Mais, dans cet afflux de vocables nouveaux, il en est beaucoup dont l'existence ne peut être qu'éphémère. Les uns disparaîtront avec les objets, eux-mêmes éphémères, qu'ils représentent ; d'autres, qui se sentent de l'improvisation, seront remplacés par des dénominations plus exactes ; d'autres enfin ne dépasseront pas le domaine où ils sont nés et, n'étant compris et employés que par des initiés, n'ont point chance de pénétrer dans l'usage commun. C'est ce départ qu'a essayé de faire l'Académie dans la préparation de cette nouvelle édition. Travail minutieux, qui ne pouvait être exécuté à la hâte, et qui exigeait un double effort d'adaptation au mouvement moderne et de prudence avisée. (*DAF*, 1932-35)

Ce constat se trouve encore renforcé dans la neuvième et dernière édition du *DAF* :

> Or, l'usage demande du temps à s'établir, et du temps encore à se constater. Le langage subit des modes saisonnières. Des expressions nées de la dernière pluie s'en iront avec la sécheresse suivante. Des vocables inventés une année seront désuets l'an d'après. Il faut attendre pour reconnaître ceux qui continuent d'avoir «cours public» parce que répondant à un besoin véritable, de même qu'il faut être attentif à ce que les termes apparus soient de formation correcte, afin d'empêcher que la mauvaise monnaie ne chasse la bonne. C'est à quoi s'emploient ces «gens éclairés» ou supposés tels qui composent, aujourd'hui comme jadis, l'Académie. [...]
>
> Jamais l'humanité n'a eu, en si peu de temps, autant de choses nouvelles à nommer !
>
> [...] Et comment, d'autre part, n'aurions-nous pas considéré la nécessité immédiate où se trouve notre langue de répondre, par la création de mots nouveaux, aux besoins toujours croissants des sciences et des techniques ? Il importe au plus haut point que cette néologie se fonde, comme il en fut dans le passé, sur les racines grecques et latines qui ont fourni la plupart de nos mots "savants". Le français, autrement, y perdrait sa physionomie. (*DAF*, 1992)

---

[12] Cf. PRUVOST J. (2006).

> Nous ne donnons entrée, parmi les termes techniques, qu'à ceux qui, du langage du spécialiste, sont passés par nécessité dans le langage courant, et peuvent donc être tenus pour réellement usuels. (*DAF*, 1992)

Nous le voyons à travers les deux extraits de préface reproduits, l'équation que l'Académie Française s'est proposé de résoudre est assez difficile. Elle doit en effet alimenter son réservoir de mots de la langue commune avec des termes techniques s'imposant comme « réellement usuels » et faire le pari que les termes spécialisés qu'elle choisit d'introduire auront une véritable pérennité. Cette équation est d'autant plus compliquée à résoudre compte tenu du rythme d'élaboration particulièrement lent du *DAF* et contrastant avec des initiatives émanant du privé telles que la millésimation du *Petit Larousse* ou du *Petit Robert*.

Nous avons donc cherché à éprouver ici l'argument académique selon lequel « [...] grâce à la prudente réserve de l'Académie, son dictionnaire, avec bien peu de changements, a pu suivre les progrès incessants de la science, et rester ouvert aux termes nouveaux qu'une science, qui ne s'arrête jamais, enfante et popularise tous les jours. » (*DAF*, 1878).

Un examen des mots sortis du dictionnaire prouve-t-il l'obsolescence des technolectes ? En nous appuyant sur les travaux de veille lexicale assurés par Camille Martinez[13], nous nous sommes ainsi intéressés à la liste des termes présents dans la huitième édition du *DAF* et non retenus dans la neuvième édition de ce même dictionnaire. Les exemples reproduits ci-dessous sont particulièrement évocateurs :

> FLOSCULEUX, EUSE. adj. T. de Botanique. Qui est composé uniquement de fleurons. Les fleurs de la centaurée sont flosculeuses. Fleur semi-flosculeuse ou demi-flosculeuse, Fleur composée qui n'est formée que de demi-fleurons, comme celles du pissenlit. (*DAF*, 1932-1935)

> ABSTERSION. n. f. T. de Chirurgie. Action d'absterger. (*DAF,* 1932-1935)

> EMMANNEQUINER. v. tr. T. de Jardinage. Mettre des arbustes ou des plantes dans des paniers, dans des mannequins. (*DAF,* 1932-1935)

> LANGUEYEUR. n. m. Celui qui fait le métier de langueyer les porcs. (*DAF*, 1932-1935)

> MADÉFACTION. n. f. T. de Pharmacie. Action de madéfier. (*DAF,* 1932-1935)

> CRIBRATION. n. f. T. de Pharmacie. Séparation que l'on fait des parties les plus délices des médicaments, tant secs qu'humides ou oléagineux, d'avec celles qui sont les plus grossières. (*DAF,* 1932-1935)

---

[13] Cf. MARTINEZ, C. (2012).

> DÉCIARE. n. m. T. d'Arithmétique. Mesure de superficie qui vaut la dixième partie de l'are. Il est peu usité. (*DAF,* 1932-1935)

Ne constituant qu'un échantillon restreint, cette liste est néanmoins représentative d'une tendance forte qui réside en effet dans la suppression d'entrées lexicales appartenant à des domaines scientifiques ou techniques, confirmant ainsi les propos de l'Académie, elle-même sujette à cette obsolescence des termes techniques.

## 3. Hypothèse : la place grandissante de la dictionnairique neutralise quelque peu les divergences typologiques

Au-delà du discours qu'elle adopte à propos des lexiques spécialisés, l'Académie Française reste toutefois, d'après-nous, également contrainte d'intégrer des termes spécialisés dans son répertoire en raison d'une dimension de plus en plus influente et présente dans le domaine de la lexicographie : la « dictionnairique » :

> À la dictionnairique correspond tout ce qui a trait à l'élaboration que définit le dictionnaire lorsqu'il fait l'objet d'un commerce. Ainsi, déterminer le nombre de pages, le nombre de signes, choisir la hiérarchie des caractères en fonction de la lisibilité, prévoir le public auquel il sera destiné, adapter le contenu à ce public, programmer la vente de l'ouvrage, sa date de lancement, la publicité dont il fera l'objet, tout cela relève de la dictionnairique. Elle n'est pas moins importante que la lexicographie, elle est simplement chronologiquement seconde tout en ayant parfois des impératifs qui s'exercent dès l'élaboration du projet. (PRUVOST J., 2003 : 23)

Pour quiconque travaille sur les dictionnaires, cette dimension est devenue de plus en plus incontournable, illustrant ainsi la place grandissante qu'elle occupe dans les processus lexicographiques que la millésimation de nos dictionnaires impose.

L'hypothèse que nous formulons ici est que la dictionnairique amplifie l'ouverture du lexique général au champ des lexiques spécialisés. Au sein du réservoir lexical à la disposition des lexicographes du français pour alimenter leurs répertoires, trois grands types de catégories de mots semblent se dégager, à savoir les régionalismes lexicaux, les mots de la Francophonie et enfin, les termes scientifiques et techniques.

Marie-Claude L'homme et Jean-Claude Boulanger, déjà en 1991, avaient mis en évidence la grande proximité entre lexique général et technolectes :

> Les dictionnaires ne font que répondre aux exigences et aux besoins de leur public respectif. Ils sont contraints de récupérer des termes spécialisés et d'en dessiner le profil dans leurs répertoires. Si, en principe, le lexicographe ne peut pas laisser échapper un mot, il doit, en revanche, effectuer une sélection très rigoureuse dans l'ensemble des sous-lexiques technolectaux qui se répartissent

> sur une palette très large (alimentation, art culinaire, banque, intelligence artificielle, bioéthique, conchyliologie, etc.). Ce n'est pas le degré de spécialisation qui permet de cataloguer les termes dans un dictionnaire grand public. Tous les technolectalismes sont susceptibles de rejoindre un jour ou l'autre une macrostructure générale si le besoin est manifesté par les consommateurs (ex. *micro-onde*). (BOULANGER C., L'HOMME M.-C., 1991 : 25)

Constituant en quelque sorte une « résidence secondaire » pour les technolectes, les dictionnaires généraux sont néanmoins contraints par leur espace même. L'exhaustivité conceptuelle des terminographes étant impossible pour les lexicographes, des choix lexicaux s'imposent donc. Les principaux dictionnaires de langue divergent ainsi sur l'intégration des termes techniques dans leur nomenclature. L'article Agammaglobulinémie en est un exemple :

> agammaglobulinémie, ▸ nom féminin
> IMMUNOLOGIE Déficit profond ou absence de gammaglobulines, et donc d'anticorps, dans le plasma sanguin. (*Petit Larousse 2009*)

> agammaglobulinémie [agamaglɔbylinemi] nom féminin, étym. v. 1970 ◊ de 2. a-, gammaglobuline et -émie
> ■ Méd. Absence ou insuffisance de gammaglobulines dans le plasma sanguin. » (*Petit Robert 2009*)

> « *AGAMMAGLOBULINÉMIE n. f. XX^e^ siècle. Composé du préfixe privatif a-, de gammaglobuline, et de l'élément suffixal -émie, du grec haima, « sang ». PATHOL. Absence de gammaglobulines dans le plasma sanguin, entraînant un trouble profond des réactions de défense de l'organisme. (*DAF* 1992)

Le recensement de ce terme est unanime mais des marquages différents sont adoptés par les trois répertoires, le rattachant successivement au domaine de l' « immunologie », de la « médecine » et de la « pathologie ». Cette absence d'homogénéité en termes de marquage peut être interprétée comme un flottement illustrant en quelque sorte l'urgence avec laquelle les équipes de lexicographes doivent traiter ce fond lexical nouveau.

Le domaine de la « pathologie » adopté par le *DAF* pour cet exemple nous donne l'occasion de proposer une autre illustration de la pénétration significative des termes techniques dans la lexicographie moderne, y compris dans le répertoire des Académiciens.

La différents termes listés ci-après appartiennent tous au domaine de la « pathologie » et reçoivent donc chacun le marquage « PATHO » : *Allergie, Arthrose, Boulimique, Cancérigène, Colite, Elongation, Fibrose, Grippal(e), Hématome, Névrosé(e), Nodule, Pneumopathie, Psychopathe, Psychotique, Pubalgie*.

Absents de la huitième édition du *DAF*, ces termes sont tous introduits dans la dernière édition de ce dictionnaire, illustrant ainsi l'ouverture massive de

l'ouvrage aux lexies de ce domaine. Notons toutefois que cette ouverture mérite – une fois de plus – d'être relativisée, compte tenu de la distance temporelle entre la 8e et la 9e édition du *DAF*, la publication du dernier tome de *DAF8* datant de 1935 et la publication du premier tome de *DAF9* datant de 1992.

Si les différents termes listés plus haut peuvent chacun être considérés comme en effet inscrits dans l'usage commun de la langue, nous pouvons néanmoins nous interroger sur la pénétration dans l'usage de chacun des mots suivants : *Entérocolite, Entérocoque, Entérokinase, Entéropathie, Entomophage, Entomophile, Entomostracés, Entoptique*.

Également marqués comme appartenant au domaine de la « pathologie », ces termes semblent encore empreints d'une technicité difficilement contestable. Peuvent-ils ainsi illustrer l'introduction d'une vision « élargie » de l'usage chez les Académiciens ?

## 4. De vrais néologismes dans le *DAF* ?

L'intérêt que nous portons au *DAF* dans cette étude nous amène par ailleurs à mettre encore davantage en perspective l'introduction des termes techniques dans cet ouvrage et à nous demander plus largement si ce dernier est en mesure d'introduire de vrais néologismes ?

Le rythme d'élaboration du *DAF*, même s'il s'appuie à nouveau sur le concours de 2 bureaux, confère à cet ouvrage un aspect plus « daté », à la fois pour les termes généraux et pour les termes techniques. Ainsi, tout en manifestant une extrême prudence quant aux technolectes qu'ils introduisent dans leur répertoire, les Académiciens sont non seulement exposés à l'obsolescence des termes retenus, mais aussi confrontés aux « vides » et aux « absences » qu'engendre l'impossibilité d'accroissement de la nomenclature de leur ouvrage pour des tranches alphabétiques publiées dès 1992. Comment l'Académie peut-elle ainsi garantir une version de son ouvrage véritablement calquée sur l'usage, sur un usage « actualisé » ? Du rythme même d'élaboration du *DAF* découle donc un retard dans la complétude de sa nomenclature.

Au même titre que les autres dictionnaires généraux, le *DAF* se caractérise par ailleurs par un recensement « tardif » des néologismes. Dans son article intitulé « Alain Rey, le linguiste et les mots nouveaux »[14], Jean-François Sablayrolles montre en effet que ce que nous considérons comme des néologismes dans les dictionnaires ne sont en réalité que des mots dont l'existence dans l'usage est ancienne, voire très ancienne comme l'illustre le terme *oligarque* enregistré dans le millésime 2007 du *Petit Robert* et pourtant daté de 1568.

---

[14] SABLAYROLLES J.-F. (2011), « Alain Rey, le linguiste et les mots nouveaux », in *Alain Rey, vocabuliste français*, GAUDIN F. (éd.). (2011), Limoges, Éditions Lambert-Lucas (collection « La Lexicothèque »), p. 87-98.

Sur cette question de l'existence d'une véritable néologie technique dans le *DAF*, il nous semble enfin important de rappeler que les Académiciens ont explicitement pris le parti – c'est en tout cas ce qui est explicitement affirmé dans la préface de leur ouvrage – de n'intégrer que les termes techniques s'étant réellement installés dans l'usage. La philosophie même de cet ouvrage semble donc interdire la présence de véritables néologismes techniques en son sein.

## Conclusion

Même si la comparaison entre le *DAF* et les dictionnaires millésimés peut parfois s'avérer assez peu significative tant ces ouvrages divergent du point de vue de leur philosophie et de leurs méthodes d'élaboration, une dichotomie semble pouvoir être établie au sein du paysage lexicographique français sur la question du traitement des technolectes. Celle-ci s'établit ainsi entre un dictionnaire qui prétend « aller à l'essentiel » et donc n'intégrer qu'avec parcimonie les technolectes (le *DAF*), et des dictionnaires « de l'éphémère », prenant le parti d'alimenter leur nomenclature avec des termes induisant certes une certaine modernité mais restant particulièrement sujets à l'obsolescence (les *Petit Larousse* et *Petit Robert* par exemple).

Tout en ayant historiquement choisi de séparer lexique général et lexiques spécialisés, le *DAF* a su/dû intégrer progressivement de plus en plus de technolectes, illustrant ainsi le fait que ce réservoir lexical a toujours constitué et constituera toujours une source privilégiée pour le renouvellement du lexique général.

La dimension dictionnairique que nous avons évoquée dans cette contribution joue selon nous un rôle majeur dans le renforcement de cette tendance au sein des répertoires modernes, et pas seulement pour ceux qui se trouvent distribués sous la forme de millésimes et qui sont donc davantage soumis à la nécessité d'afficher un renouvellement lexical d'une année sur l'autre.

Nous souhaitons achever notre propos en nous demandant si l'aspect « éphémère » et l' « obsolescence » des technolectes contre lesquels les Académiciens semblent lutter ne constituent pas au final des critères essentiels pour la dimension culturelle et vivante des langues.

## Références bibliographiques

BOULANGER J.-C., L'HOMME M.-C. (1991), « Les technolectes dans la pratique dictionnairique générale. Quelques fragments d'une culture », *Meta : journal des traducteurs / Meta: Translators' Journal*, vol. 36, n° 1, 1991, p. 23-40.

BOUVEROT D. (1998), « Mots nouveaux et vocabulaires spéciaux dans le Dictionnaire de l'Académie », *Actes du Colloque international Le Dictionnaire de l'Académie française et la lexicographie institutionnelle européenne*, Institut de France,

novembre 1994; (ed. B. Quemada & J. Pruvost), Paris, Champion : 237-252.
BRANCA-ROSOFF S., FOURNIER J.-M., GRINSHPUN Y., REGENT-SUSINI A. (eds). (2011), *Langue commune et changements de normes*, Paris, Champion, 2011, pp. 516.
*Dictionnaire de l'Académie Française*, 9e édition. Version informatisée : http://atilf.atilf.fr/academie9.htm
MARTINEZ C. (2012), *L'Orthographe des dictionnaires français. La construction de la norme graphique par les lexicographes,* Paris : Honoré Champion, coll. "Lexica, mots et Dictionnaires, 22".
MESSAOUDI L. (2002), « Le technolecte et les ressources linguistiques. » L'exemple du code de la route au Maroc, *Langage et société*, 2002/1 n° 99, p. 53-75.
PRUVOST J. (2006), *Les dictionnaires français : outils d'une langue et d'une culture*. L'Essentiel français Ophrys, Paris.
PRUVOST J. (2003), « R. Galisson : un pionnier explorateur des mots et de contrées lexicologiques et lexiculturelles », *Mots et lexiculture*, H. Champion, Collection Études de lexicologie, lexicographie et dictionnairique, pp. 11-17.
REY C. (À paraître), « Lexicographie et technolectes: le dictionnaire de langue comme dictionnaire de spécialité », *Actes des journées marocaines du dictionnaire*, 09 décembre 2011, Kénitra.
REY C. (2012a), « Le Dictionnaire de l'Académie Française : un modèle lexicographique en évolution », *Actes du Troisième Congrès mondial de linguistique française*, 4-7 juillet 2012, Université Lumière Lyon 2, pp. 1031-1041.
REY C. (dir.). (2012b), *Études de Linguistique Appliquée (ÉLA)* n° 163, « L'Académie Française : un modèle qui traverse les siècles », Éditions Klincksieck.
SABLAYROLLES J.-F. (2011), « Alain Rey, le linguiste et les mots nouveaux », in *Alain Rey, vocabuliste français*, GAUDIN, F. (éd.), Limoges, Éditions Lambert-Lucas (collection « La Lexicothèque »), p. 87-98.

# La Tulipe « royne » : prolifération et circulation néologiques aux XVIe-XVIIe siècles

Philippe Selosse[1]

« *La Tulipe Royne des fleurs* » (Colin, 1619, V : 165)
« *Les Tulipes* […] *Reines des fleurs* » (Morin, 1658 : 181)

## Résumé

L'Occident, après avoir découvert la Tulipe à la Renaissance, a commencé à élaborer une très vaste nomenclature de cette plante, qui comprend plus de mille dénominations pour les seules années 1560-1660. La néologie se réalise différemment selon le rang taxinomique envisagé. Le nom générique (*Tulipa, Tulipe, Tulpe*…) se forme dans toutes les langues par emprunt au turc, selon des motivations épistémiques particulières. Les noms spécifiques, qui prennent la forme de noms-phrases en langue véhiculaire (néo-latin), sont construits principalement au XVIe siècle par les botanistes, dans un cadre essentialiste où le nom est la définition logique de la plante. Les noms variétaux, enfin, de forme principalement binomiale en langue vernaculaire (français, flamand, anglais), sont construits au XVIIe siècle par les fleuristes, selon des modalités et des structures syntaxiques propres, qui anticipent la nomenclature linnéenne où les noms fonctionnent plutôt comme des noms propres.

## Mots clés

Nomenclature botanique, Nomenclature binomiale, Nom-phrase, Nom propre, Emprunt, Ellipse, Conversion lexicale, Tulipe, Genre, Espèce, Variété

## Introduction

Pour étudier, en diachronie, la formation des néologismes, leur circulation et leur évolution dans un discours de spécialité, il faut un poste d'observation qui permette à la fois de cerner avec certitude l'apparition des néologismes et de disposer d'un corpus quantitativement représentatif du phénomène de création, diffusion et amplification de ceux-ci. Il semble que la Tulipe soit un exemple qui répond parfaitement à ces exigences :

(i) on peut dater avec précision tant sa découverte par les voyageurs européens en Turquie au milieu du XVIe siècle (autour de 1550) que son introduction en Occident (autour de 1555) et, surtout, sa première description

[1] UMR 5037 (Institut d'Histoire de la Pensée Classique), Université Lyon 2, selosse.philippe@wanadoo.fr

et sa première mention nomenclaturale (« Tulipe ») dans la littérature écrite européenne (Gessner, 1561) ;

(ii) c'est une plante qui a la particularité, unique, de présenter des panachages très variés de ses pétales, lesquels constituent des centaines et jusqu'à des milliers de formes nouvelles et particulières[2]. C'est cette particularité morphologique qui va susciter aux XVI-XVIIe siècles une admiration religieuse, esthétique et même économique pour la Tulipe, donnant lieu à des mouvements de « tulipomanie » dont l'un débouchera sur la première « bulle spéculative » chez les Hollandais dans les années 1630 (Roman, 1637a, 1637b, 1640) ;

(iii) en un siècle, ces formes de tulipes vont faire l'objet de très nombreux échanges entre plusieurs pays, occasionnant une prolifération néologique, d'abord en néo-latin, langue des botanistes, puis en divers vernaculaires (français, néerlandais…), langues des jardiniers, fleuristes et amateurs ou commerçants de ces plantes.

Pour appréhender progressivement l'ampleur et la complexité du phénomène néologique, je suivrai le développement chronologique de la connaissance de cette fleur, autrement dit, je partirai de la découverte du genre de la Tulipe, pour en venir aux nouvelles espèces que distinguent très rapidement les botanistes, avant d'en arriver aux milliers de formes ou variétés que les fleuristes vont tenter d'isoler, par la culture comme par la nomenclature.

## 1. La dénomination du genre : un emprunt au « vulgaire »

Au milieu du XVIe siècle, la botanique n'est pas encore constituée comme discipline. Le savoir qui porte sur les plantes porte le nom de *Res Herbaria* (disciplines des herbes) et est développé principalement par les médecins et herboristes, les philosophes et historiens de la nature, qu'on désignera ci-dessous, par commodité[3], sous le nom générique de « botanistes ». C'est relativement à l'*épistémè* de ces humanistes que cette fleur, en tant que nouveau genre de plante, va se trouver dénommée. Ce cadre épistémique peut être schématiquement décrit de la sorte :

– il est fixiste, au sens où tous les genres de plantes sont supposés avoir été créés lors de la Genèse et avoir donc pu être décrits par les auteurs de l'Antiquité (Théophraste, Dioscoride, Pline). Cette conception construit un monde fini qui refuse toute possibilité de néologismes, lesquels reviendraient,

---

[2] Panachage dont la cause ne sera découverte qu'en 1927 : il s'agit d'un virus transmis par des pucerons, qui infecte bulbes et bulbilles – mais n'est pas transmissible par les graines – et qui affaiblit la plante, produisant entre autres la décoloration partielle des pétales et les panachages consécutifs appréciés des amateurs.

[3] Les premiers « botanistes », se désignant comme tels, apparaissent dans les années 1580.

par définition, à supposer possible la création de nouveaux genres depuis la Genèse ;

– il n'est cependant pas cognitivement fermé, en ce qu'il admet la possibilité de plantes nouvelles en tant que celles-ci auraient été inconnues de l'homme jusqu'à la Renaissance (Cesalpino, 1583) : cette nouveauté concerne surtout des espèces mais aussi, à la fin de la Renaissance et sous le poids de l'apport des plantes du Nouveau Monde, des genres de plantes exotiques ;

– il assimile l'acte dénominatif d'une « nouvelle » plante à l'acte religieux du « baptême » (Lobel, 1576), ce qui implique que la juste dénomination ne peut se faire que dans les langues sacrées des autorités gréco-latines mais surtout pas dans les langues dites « exotiques ».

Ce qui précède explique, d'une part les dénominations que les « botanistes » du XVIe siècle vont utiliser pour désigner initialement le genre de la tulipe, d'autre part, le recours à leur langue véhiculaire, le néo-latin. On discernera trois cas de figure.

(i) **Absence de néologisme** : les premières *références* à cette fleur se font sans aucune dénomination qui lui soit propre. C'est ainsi que les premiers auteurs emploient souvent le nom *Lys* (*Lilium*) [1], genre de fleur auquel ceux qui vont utiliser le néologisme de *tulipe* continueront de rapporter volontiers cette nouvelle plante [2] :

> 1. [Les Turcs] portent plusieurs [fleurs] seule à seule dedans le reply de leurs turbans […] Les Lils rouges sont si communs, qu'il n'y a celuy qui n'en ait des plantes en son iardin (Belon, 1553, III, LI : 206 v°).
> 2. C'est une autre herbe que les hommes savants désignent en Italie du nom turc de Tulipe, dont Johann Kentmann nous a envoyé une gravure : à trois feuilles, à fleur comme celle du petit Lis, jaune, à bulbe roux, semblable à l'oignon[4] […] (Gessner, 1561 : 263 v°).

(ii) **Emploi d'un néologisme partiellement motivé** : de nombreux auteurs ont recours à la dénomination *Lilionarcissus* ou (plus rarement) à son allomorphe *Narcissolilium*. Cette dénomination est choisie pour signifier que la plante participe (en termes néo-platoniciens) du Lys et du Narcisse et ce, indifféremment, d'où la permutation possible des interfixes : il s'agit ainsi, pour les botanistes, de marquer le fait que cette plante relève de deux formes végétales créées et déjà connues des botanistes dans le monde fini de la Nature [3]. Cette nomenclature est abondamment utilisée [4]-[5] :

> 3. Lys-narcisse (*Lilionarcissus*) Narbonnais jaune de montagne […]. Nous avons pris soin que ce narcisse, déterré il y a déjà de nombreuses années des sommets

[4] Dans tous les exemples, c'est moi qui souligne ; pour les textes latins, c'est moi qui traduis.

dits du Jardin de Dieu, dans les Cévennes Narbonnaises, soit apporté en Belgique par des amis Anversois, qui l'ont d'abord appelé *Tulipe*, tandis que nous, nous lui avons conféré le nom de *Narcisselys* (*Narcissolilium*) *jaune*, avec peut-être beaucoup de hardiesse mais selon notre habitude visant à faire connaître plus distinctement les choses (Lobel & Pena 1570 : 51).
4. *Lilionarcissus purpureus. Lilionarcissus Chalcedonicus totus luteus, sanguineo puniceo par & similis* (Lobel, 1576 : 65).
5. *Lilio-Narcissus niveus totus. Lilio-Narcissus candore et rubore confusus* […] (Gobel, 1581 : 166).

(iii) **Emploi d'un néologisme totalement immotivé, par emprunt** : dès la première *description* de la plante, la forme néo-latinisée *Tulipa* a été requise :

6. Quelques-uns l'appellent du mot turc de *Tulipe* (Gessner, 1561 : 263 r°).

Cependant, d'une part, comme on l'a vu, le sens opaque de cet emprunt est systématiquement compensé par référence au genre *Lilium* pensé comme similaire ; d'autre part, le néologisme est habituellement pris avec distance. En effet, on note en premier lieu une réticence marquée à l'usage du néologisme dans le cadre fixiste – réticence fréquente pour les plantes possiblement « nouvelles » – de sorte que, si les auteurs utilisent le nom *Tulipa*, ce n'est souvent que temporairement, par commodité, en attendant que soient réidentifiés la plante et son nom anciens, dont la connaissance est supposée s'être perdue depuis l'Antiquité. On relève ainsi la formulation fréquente d'un doute ou d'une interrogation dans les textes, la Tulipe n'étant jugée être qu'une espèce d'un genre déjà existant, le Satyrion (sorte d'orchidée) – je souligne ci-dessous les marques de modalisation :

7. Et peut-être notre Tulipe aussi est-elle une espèce de Satyrion, notamment du Satyrion rouge, puisqu'il est à fleur rouge (Gessner, 1561 : 263 v°).
8. Conrad Gessner a jugé que la Tulipe était le Satyrion […] ou tout au moins une certaine espèce de Satyrion (Dodoens, 1568 : 199).
9. Nombreux sont ceux à se demander par quel nom les Anciens appellaient cette plante. Il y en a pourtant quelques-uns de très experts pour juger qu'elle doit être rapportée au Satyrion de Dioscoride, nommé « à trois feuilles » par certains ; de fait, elle touche à la description de ce dernier par de nombreuses notes (L'Escluse, 1576 : 514).

En second lieu, les botanistes ont aussi de fortes réticences religieuses et savantes à utiliser un néologisme vulgaire, *i.e.* en langue vernaculaire, qui ne soit pas formé dans une des langues sacrées de l'Antiquité :

10. There is lately a *Flower* […] a *Tulip* […] As for the use thereof in *Physic*, no *Physician* hath honoured it yet with the mention, nor with a *Greek* or *Latin* name (Révérend Fuller, [1655], 1867 : 279-280).

Ce rejet est accru par le fait que l'emprunt *Tulipa* repose sur une erreur linguistique que relève très tôt le polyglotte Charles de L'Escluse, à savoir que pour désigner cette fleur, la langue turque n'utilise pas le nom *Tulipe* (*tülbent*) mais le nom *lâlé* :

> 11. J'ai observé deux genres de cette plante étrangère, appelée vulgairement Tulipe – non pas qu'ils diffèrent entre eux par la forme mais parce que les fleurs de l'un ont l'habitude d'être précoces, tandis que celles de l'autre sont tardives. Je note que les Turcs, en leur langue, appellent le premier genre *Cavalá lalé* [Tulipe de Kaválla, en Macédoine], tandis qu'ils appellent l'autre *Café lalé* [Tulipe de Kefe = Feodosia en Ukraine], noms tirés des lieux d'où ils furent d'abord apportés à Constantinople – en effet, *lalé* signifie « fleur » chez eux (L'Escluse, 1576 : 510).
> 12. Turc *tülbend* (persan *dul(i)-band* « lien ») > fr. *turban* et *tulipe* vs turc *lalé* « tulipe ».

L'explication de cette confusion (Pavord, 2001) pourrait être une méprise du premier voyageur qui, dans les explications que lui donnait son guide-interprète, aura pris pour le nom de la plante elle-même le nom de la forme en turban de la fleur ou du lieu où elle était portée – le turban dans lequel les Turcs arboraient leurs fleurs (voir citation [1]).

De la dénomination du genre *Tulipa*, nous pouvons donc tirer deux conclusions à ce stade :

(i) quant à l'évolution et au repérage des néologismes. Jusqu'à la fin du XVIe siècle, les réticences des locuteurs attestent nettement le statut de néologisme du nom *Tulipa*. En revanche, à partir du XVIIe siècle, on note une modification dans l'attitude des locuteurs, qui ne pointent plus l'erreur de la dénomination initiale mais justifient son étymologie comme celui de n'importe quelle autre unité lexicale, ce qui atteste son intégration de mot dans le lexique véhiculaire ou vernaculaire et la perte de son statut néologique :

> 13. Ceux qui les premiers nommerent les Tulipes, ont curieusement recherché de luy donner vn nom conuenable à sa beauté : ce qui a fait qu'en la considerant dans sa forme, ils l'ont connuë de celle des Tulbans. Et pour sa beauté, comme les Turcs s'étudient particulierement à l'enrichissement de leurs Tulbans, pour leur donner de l'eclat & du lustre, ils ne purent trouuer vn nom qui eust plus de rapport à la Tulipe que celuy de Tulban (Monstereul, 1658 : 30-31).

Sans doute ce passage du néologisme *Tulipa* dans l'usage, sans plus de réticences, est-il dû au fait que, dès la fin du XVIe siècle, la culture de cette fleur a révélé une telle diversité de variétés que celle-ci n'a pu qu'imposer l'idée d'une plante vraiment nouvelle et absolument distincte du Lys ou du Satyrion, nécessitant un nom propre ;

(ii) quant à la circulation et à la diffusion des néologismes. La forme *Tulipa* se retrouve dans toutes les langues vernaculaires, moyennant des adaptations morphologiques au système de la langue d'accueil :

> 14. All. *Tulpe*, angl. *tulip*, esp. *tulipán*, fr. *tulipe*, it. *tulipano*, port. *tulipa*, rus. *tioul'pan*, suéd. *tulpan*…

Or, il est impossible que la même erreur originelle de désignation, pointée par L'Escluse, se soit répétée dans chaque langue : il faut donc supposer une diffusion dans les différents vernaculaires à partir de la forme erronée néo-latine (*Tulipa*), au gré des échanges botaniques et économiques.

## 2. La dénomination des espèces : les néologismes néo-latins des botanistes

Les botanistes qui échangent et cultivent les premiers bulbes de Tulipes constatent très vite une grande diversité de formes, qu'ils considèrent comme des espèces. La dénomination de celles-ci se fait dans le cadre épistémique précédemment décrit, d'un hommage obligé au Créateur à travers la diversité de sa Création. Les multiples espèces de Tulipes sont donc scrupuleusement observées et dénommées selon le modèle nomenclatural botanique qui s'élabore alors (Selosse, 2004, 2007) : il s'agit d'un modèle essentialiste, définitoire et logique, où chaque nom (i) vise à définir l'essence de l'espèce qu'il dénomme et (ii) récapitule les différentes étapes de la division logique qui constitue la méthode du savoir alors privilégiée.

C'est ainsi (voir figure n° 1) que le « genre suprême » de la Tulipe se divise en plusieurs « espèces », la *Tulipe précoce*, la *Tulipe tardive* et la *Tulipe* [de floraison] *incertaine*[5]. Dans la mesure où chacune de ces espèces est à son tour divisible, elle constitue simultanément, au sens logique, un « genre » ; dans le cas du genre *Tulipe précoce*, celui-ci se divise en plusieurs « espèces », la *Tulipe précoce monocolore* et la *Tulipe précoce bigarrée*. Toutes deux sont à leur tour divisibles et forment conséquemment des « genres ». Le genre *Tulipe précoce bigarrée* se divise en plusieurs « espèces », dites « ultimes » dans la mesure où elles ne sont plus divisibles à leur tour – parmi lesquelles la *Tulipe précoce bigarrée jaune*, la *Tulipe précoce bigarrée rouge*, la *Tulipe précoce bigarrée blanche*, etc.

[5] Pour simplifier la démonstration, j'illustre la méthode nomenclaturale directement en français, mais il est à noter que les exemples d'origine sont bien en véhiculaire néo-latin (*Tulipa praecox*, *Tulipa serotina*, *Tulipa dubia*, etc.). Les crochets de la figure n° 1 indiquent des étapes de division logique qui ne sont pas détaillées, pour des raisons de simplification également.

*Figure 1. La division logique dans le genre* Tulipe *chez Bauhin (1623)*

La nomenclature néologique ainsi constituée se caractérise tout d'abord par le fait que ce sont les épithètes, placées en dernière position dans le syntagme, qui sont porteuses de l'essence des espèces dénommées – en l'occurrence, « jaune », « rouge » et « blanche » :

15. *Tulipa praecox varia lutea. Tulipa praecox varia rubra. Tulipa praecox varia candida.* Tulipe précoce bigarrée jaune. Tulipe précoce bigarrée rouge. Tulipe précoce bigarrée blanche.

Par ailleurs, cette nomenclature a pour particularité d'être ouverte, en ce qu'elle est formée commodément par addition, au fur et à mesure des étapes de division logique, et en ce qu'elle peut produire de nouveaux noms, en fonction des découvertes. Ainsi Caspar Bauhin, auteur d'une synthèse du savoir sur les plantes à la Renaissance et d'une nomenclature très utilisée par la majorité des botanistes jusqu'au XVIIIe siècle, peut-il créer autant de néologismes que nécessaire pour accompagner l'augmentation du nombre de Tulipes recensées, de 26 espèces en 1596 à 37 en 1623. Cet accroissement se fait moyennant la réfection des dénominations initiales, construites sans patron très défini, de sorte que la nouvelle nomenclature reflète la structuration logique et accepte en théorie comme en pratique l'insertion d'autres nouvelles espèces et dénominations (Bauhin, s. d.). L'exemple de l'espèce suivante le montre bien :

16. *Tulipa rubra* (Bauhin, 1596) > *Tulipa praecox varia rubra* (Bauhin, 1623). Tulipe rouge > Tulipe précoce bigarrée rouge.

Il est évident que si Bauhin en était resté à un néologisme reposant sur la couleur et réduit au binôme N + Adj. (*Tulipe rouge*), la création néologique aurait été, d'une part, bien moins logique que lorque Bauhin utilise des polynômes faisant apparaître les divisions intermédiaires (précoce *vs* tardive, bigarrée *vs* monocolore) ; et d'autre part, bien moins facilement extensible : le nom *Tulipe précoce bigarrée rouge* laisse entre autres la possibilité d'insertion de nouvelles espèces telles que la *Tulipe tardive rouge*, la *Tulipe précoce monocolore rouge*, etc.

Toutefois, lorsque les botanistes observent que chacune de ces formes « ultimes » peut elle-même se diversifier extrêmement, ils ne poussent pas plus loin la division logique et arrêtent plus ou moins leur recensement et leur nomenclature scientifique au stade bauhinien, *i.e.* à une quarantaine d'espèces. Ce sont les « fleuristes » (horticulteurs) et les « curieux » (riches amateurs de plantes d'ornement) qui vont prendre le relais et amplifier le nombre de formes recensées et durablement dénommées.

## 3. La dénomination des variétés : les néologismes vernaculaires des fleuristes

La diversité des tulipes consécutive au panachage des pétales est conçue dans le cadre religieux des XVI-XVII^e^ siècles comme une démonstration de la « puissance créatrice infinie de Dieu » : et les variétés apparaissent alors comme les simples réalisations en acte d'une forme créée dès la Genèse. Les conditions sont ainsi assurées pour que, en toute compatibilité avec le cadre épistémique, toute nouvelle variété puisse faire l'objet d'un néologisme, d'autant plus que des intérêts diversement motivés contribuent à rendre nécessaire cette dénomination des variétés : motivations esthétique [17], sociale – améliorer le « commerce » (les relations) entre « honnêtes gens » [18] – et surtout mercantile [19] :

> 17. Diuersité : est ce que nous appelons beauté en la Tulipe (Monstereul, 1658 : 73).
> 18. Si les Tulipes estoient renduës communes, ce seroit encor oster le plus loüable commerce qui se pratique entre les Hommes, & les priuer de la plus douce societé qu'il y aye entre les gens d'honneur. Combien leur rareté donne-t'elle de connoissance aux Esprits curieux ? combien d'agreables visites ? combien de douces conuersations ? & combien de solides entretiens ? (Monstereul, 1658 : 180-181).
> 19. S'il se produit un changement dans une tulipe, on va le signaler à un fleuriste, et très vite tout le monde en parle. Tout le monde veut la voir. S'il s'agit d'une fleur nouvelle, chacun donne son opinion : l'un la compare à telle fleur, l'autre à telle autre. Si elle ressemble à une 'Admiral', vous l'appelez 'General' ou vous lui donnez n'importe quel autre nom à votre goût, et vous offrez une bouteille de vin à vos amis pour qu'ils n'oublient pas d'en parler (Roman 1637 ; traduction in Pavord, 2001 : 173).

Le virus qui contamine diversement les bulbes (et qui ne sera découvert qu'au XX^e^ siècle) engendre un nombre infini de variétés, qui s'accompagnent d'autant de dénominations. Les exemples ci-dessous témoignent de l'essor considérable de la quantité de néologismes pour désigner les différentes variétés, sachant que les fleuristes et curieux ne citent pourtant même pas l'intégralité de la nomenclature alors en usage et que nous n'en avons donc qu'un aperçu :

> 20. Sweert (fleuriste, flamand) 1612 : 31 dénominations.
> Besler (fleuriste, allemand) 1613 : 44 dénominations.
> Cos (peintre de « livre de Tulipes », flamand) 1637 : 54 dénominations.

Roman (écrivain satirique, flamand) 1637a : 131 dénominations.

Morin (fleuriste, français) 1651 : 100 dénominations (parmi plusieurs centaines, précise-t-il).

Monstereul (« curieux », français) 1658 : 451 dénominations – la plupart se retrouveront encore chez Anonyme (1756) et dans l'*Encyclopédie* de Diderot et d'Alembert (1778).

Eslholz (botaniste, allemand) 1663 : 214 dénominations.

Après une brève présentation du mode de formation de ces néologismes de fleuristes, je m'attacherai, à partir de la critique que font les botanistes de cette nomenclature, à en mettre à jour les principales caractéristiques structurales.

### 3.1. Modes de formation néologique : du « nom-phrase » descriptif au nom étiquette

Bien qu'elles soient surtout dues à des fleuristes, les variétés font d'abord l'objet de dénominations en néo-latin, sous forme de noms-phrases qui sont de véritables descriptions des principales couleurs et panachages des pétales. C'est ainsi que Basilius Besler, quoique apothicaire et « amateur éclairé de la *Res Herbaria* », propose un « florilège », autrement dit un ouvrage de fleuriste, qui illustre « des collections végétales vivantes ornant des jardins » (Aymonin, 1987), en l'occurrence celles du jardin d'Eichstätt, et qui contient entre autres les Tulipes suivantes, auxquelles il attribue ses dénominations, très différentes de celles d'un Bauhin ou d'un L'Escluse :

21. *Tulipa Coccinea albescentibus oris* = Tulipe écarlate à bords blancs.
*Tulipa lutea margine roseo* = Tulipe jaune à bordure rosée.
*Tulipa lutea maculis aspersa minutis* = Tulipe jaune parsemée de petites taches.
*Tulipa candida lineis viridibus* = Tulipe blanche à lignes vertes.
*Tulipa miniata linea viridi* = Tulipe rouge minium à ligne verte.
*Tulipa floribus ex sulphureo colore pallescentibus lituris prope apices roseis flammiformibus* = Tulipe à fleurs d'une couleur soufre tirant sur le pâle à rayures en forme de flamme, roses près des pointes (Besler, 1613 : 81 et 85).

De tels noms sont d'abord utilisés par les fleuristes et même cités par les botanistes (Bauhin, 1623). Mais lorsque commence la vogue des « livres Tulipes », qui ont pour fonction de donner à voir aux acheteurs ce que seront les bulbes proposés à la vente, les fleuristes délaissent la nomenclature néo-latine pour une nomenclature vernaculaire. On voit ainsi Sweert utiliser dès 1612 des nomenclatures vernaculaires en trois langues (français, allemand, néerlandais), bien qu'il les rattache encore nettement à la nomenclature véhiculaire néo-latine, placée en tête :

22. Tulipa lutea, oris sulphureis, annulo rubro punctis rubris intus pleno, fundo vario.
Tulipe jeaune a bords sulphures portant anneau rouge auec points rouges interieurement plein d'un fond varié.

Tulipa gelb / die ecken schwebelgelb ausswendig ein rothe ringel / darinnen viel rothe sprenckeln / der boden gelbgruen und schwarz.

Tulipa gheel dij boirden solpher gheel van bunten eenen roothen rinck / binnen voll roote sprenckelen / den gront geel / groen / ende swarth (Sweert, 1612 : *passim*).

La langue néo-latine demeure en effet de manière sensible le pivot linguistique, au point que le néologisme prend en partie la forme d'un xénisme (permanence du nom *Tulipa* dans les dénominations en allemand et en néerlandais) et que, d'une manière générale, les traductions vernaculaires se présentent comme un mot à mot de l'original latin, à la limite de la correction syntaxique et sémantique, comme on le voit nettement dans le texte français[6] : la nouvelle nomenclature de Sweert constitue ainsi un stade intermédiaire, sous forme de concordancier quadrilingue. Vient très rapidement le temps où la nomenclature ne se présente plus qu'en vernaculaire, français ou néerlandais. Se manifeste alors une rupture totale dans le mode de dénomination, qui instaure une distinction fondamentale entre le nom d'une part, et la description d'autre part – distinction que Morin (1651, 1655, 1658) ou Monstereul (1658) soulignent par la ponctuation (virgule) et le type de caractères utilisés (romains *vs* italiques) pour l'une et l'autre catégorie :

23. Fr. Agate Robin, *a du pourpre, rouge et blanc* (Monstereul, 1658 : 209).
Fr. *Agate Morin*, est blanche, gris sale, et pourpre (Morin, 1658 : 185).
Néerl. *Admirael Gouda* (Swanenburch, 1628 : 59).

Une rapide recension, non exhaustive, des différents types sémantiques de néologismes dus aux fleuristes et particuliers montre combien ces néologismes s'apparentent à présent à des noms propres et sont détachés de toute description originelle, à l'inverse des noms-phrases :

24. **A. Noms intensifs**
(i) Titres aristocratiques en français : *Duc, Marquise, Prince, Reyne, Roi*… – ou grades militaires en flamand : *Admirael, General*…
(ii) Hommes illustres :
– dieux et héros mythologiques : *Jupiter, Mercure, Pallas, Venus, Argus, Hercule, Hector*…
– personnages historiques de l'Antiquité latine : *Auguste, Caesar, Scipion, Marcus Aurelius*…
– rois et hommes politiques contemporains : *Cardinal Richelieu* ;
– imprimeurs (*Elzevier*), peintres (*Latour, Poussin*), « botanistes » et fleuristes : *Columelle, Esclusius* [sic], *Richeual, Morine, Robine, Bransion, Monstereulle*…
(iii) Superlatifs : *Eminente, Geante, Generalissime, Glorieuse, Impériale, Monstre, Parangon, Triomphante*… – entre autres par conversion de proposition ou syntagme (souvent adjectival) : *Cedanulle* ('qui ne cède à nulle autre'), *Tousiours belle, Merveille de…, Vnique de*…

[6] Quoi qu'il en soit du système souple du français pré-classique, ce sont de véritables anacoluthes par rapport à la syntaxe qui ressort de l'usage majoritaire.

**B. Noms géographiques**

(i) Pays : *Angloise, Chinoise, Polonoise…*

(ii) Villes : *Florentine, Iaspée de Harlan, Belle d'Anuers, Pourpre de Roüen, Vnique de Caen…*

**C. Noms descriptifs** : *Amarantine, Bruijnpurpur, Canelée, Coquille marbrée, Drap d'argent, Estoilée, Fleurdelisée, Flamboyante, Frangée, Incarnat Meerschijn, Ourlée, Plumage, Plumerolle, Quatricolor, Rosée, Reguliere, Satinée, Sulphurgeel met rood…*

## 3.2. Interprétation des néologismes : défauts et ambiguïtés

C'est d'ailleurs en partie pour cela que les botanistes vont faire une critique en règle des néologismes vernaculaires, en leur opposant la nomenclature botanique logique :

> 25. Mais en ce qui concerne les nomenclatures françaises et belges[7] des Tulipes, bien que la manière d'imposer ces noms ne doive pas tout à fait être rejetée, elle est cependant moyennement approuvée par les plus savants. Je ne me souviens vraiment pas qu'aucun auteur classique en botanique ait suivi une telle voie, laquelle est en outre glissante et inconstante, eu égard au fait qu'elle change au gré des songes du premier jardinier ignorant venu. Ajoutez que cela devient difficile et sans la moindre utilité, puisqu'à ce jour, cette manière de faire est dépourvue de méthode écrite et que la transmission de ces appellations, dans le cadre d'un enseignement marginal, fait qu'on ne les apprend que par un simple passage en revue, tout comme les caractères de l'alphabet (Eslholz, 1663 : 210).

Comme le montrent les termes soulignés (« manière », « voie », « méthode »), ce qui manque principalement aux dénominations vernaculaires, c'est une régularité méthodique, car leur principe est tout à fait inverse, gouverné par la fantaisie, c'est-à-dire le changement individuel, imprévisible et irrégulier en tant que tel (« glissant et inconstant »). Par ailleurs, pour un botaniste, les nomenclatures vernaculaires présentent trois graves défauts :

(i) elles substituent des caractéristiques humaines, métonymiques, aux caractéristiques végétales, le plus souvent morphologiques (couleur des fleurs, type de panachage des feuilles, saison de floraison), de la nomenclature botanique ;

(ii) elles détachent ces caractéristiques adjectivales du genre nominal *Tulipe*, en raison de leur nominalisation (par conversion, type *La Cedanulle*, ou par ellipse, type *La Venitienne*). Or, du point de vue botanique, une différence essentielle ne peut s'entendre sans son genre, si bien qu'un botaniste n'occultera jamais les différents niveaux de division logique, autrement dit les différences constitutives des genres intermédiaires (*précoce* / *bigarrée*), et *a fortiori*, il ne pratiquera pas l'ellipse du nom du genre suprême (*Tulipe*) :

---

[7] Comprendre : en néerlandais.

26. Fleuristes : *Tulipe blanche printanière* > *La Blanche Printanière*.
27. Botanistes : *Tulipe précoce bigarrée blanche* ⇒ **La blanche*.

(iii) elles présentent des néologismes qui ne sont compris que par les individus qui les ont créés, de sorte qu'ils deviennent rapidement incompréhensibles par la collectivité (voir fin de la citation [20] sur la compréhension et la transmission). La critique est assez fondée, car dès le XVIIe siècle, même un spécialiste des Tulipes comme Monstereul ne sait plus correctement interpréter la nomenclature vernaculaire héritée des Hollandais :

28. Les Hollandois ont vne autre façon de nommer les Tulipes ; ils leur donnent le nom de Generaux d'Armée (Monstereul, 1658 : 43).

C'est un véritable contre-sens par interprétation à la lettre, car les Tulipes en question ne renvoient pas à un Genéral ou à un Amiral particulier mais à une ville ou un horticulteur, le terme *Admiral* n'étant qu'un intensif sans référence concrète – voir *supra*, [24], classe A. (i) :

29. *Admirael Gouda* = une Tulipe obtenue et cultivée dans la ville de Gouda.
*Admirael Catlein, General Catlein* = des Tulipes obtenues par Jan Casteleijn, pépiniériste de Haarlem.

### 3.3. Structuration de la nouvelle nomenclature

Ces problèmes d'interprétation, récurrents, rendent nécessaire la mise au jour de la structure profonde de la nomenclature vernaculaire et de ses principes de structuration, qui n'ont jamais été explicités ni formalisés par leurs concepteurs[8].

#### 3.3.1. Problèmes d'interprétation

Le premier problème qui se pose est celui de l'interprétation du constituant n° 2 dans des exemples comme ceux de [29] et [30], où le nom propre *Ca(s)telein* n'est pas incident syntaxiquement à *Admiral*. Les étapes de formation suivantes, attestées par le corpus néo-latin des fleuristes, permettent de comprendre l'origine de la structure :

30. *Tulipa Castelijn* = Tulipe Catelein.
31. *Tulipa Ioannis Castelijni* […] = Tulipe de Iean Catelin.
32. *Tulipa Castelijn nominata* […] = Tulipe nommée Castelijn.
33. En résumé : *Tulipe de X* > *Tulipe surnommée X* > *Tulipe X*.

[8] Pas plus que ceux de la nomenclature botanique logique, restitués par les travaux de Selosse (2004). Dans l'*épistémè* pré-classique (avant 1650), il est rare qu'une méthode scientifique fasse l'objet d'un exposé explicatif : ses principes, à caractère plus tendanciel que régulier, sont souvent à déduire d'éléments introductifs épars ou des données résultant de la méthode présupposée.

Une forme comme [30] est formée par suppression progressive des marques syndétiques (génitif [31], prédicat d'appellation [32]) entre constituants, au profit d'une pure parataxe déterminative du type *rue / président Carnot*. Cette parataxe est évidemment problématique quand le nom propre est ajouté à une Tulipe comme une *Admiral* ou une *General*, puisque le nom support est, en profondeur, le nom générique *Tulipe*, ici elliptique :

34. *Admiral X* = *Tulipe Admiral surnommée X / de X.*

Le deuxième problème est relatif à l'interprétation du constituant n° 3 dans une dénomination du type [35]. Dans ce cas, un adjectif comme *rectifié* n'est incident ni à *Admiral* ni à *Castelein* mais à l'ensemble de la dénomination, car c'est le type « Tulipe Admiral Castelein » qui est « rectifié » – autrement dit, qui constitue une variété stable, laquelle peut donc être nommée :

35. Admiral Castelein rectifié (Morin, 1658 : 185) =
(1) Tulipe « Admiral » (= d'un certain type de panachage)
(2) obtenue par Jan Casteleijn, pépiniériste de Haarlem
(3) présentant un panachage durable, stabilisé depuis au moins 2 ans (« rectifié = « [ayant] panaché par [le] travail [du Curieux] & ayant continué deux années sans changer » – Monstereul, 1658 : 32).

L'adjectif-participe *rectifié* est souvent omis [36], en tant que propriété impliquée par la possibilité même de la dénomination – le nom n'est en effet possible que pour une variété de Tulipe stable ; il va donc de soi qu'une Tulipe panachée est, *a priori*, rectifiée :

36. Admiral Castelein.

De même, le nom *Admiral* peut aussi être omis comme en [37], ce qui aboutit à une configuration très ambiguë, puisque coexistent les dénominations [38]-[40] et que, hors contexte, une hésitation devient possible quant à l'interprétation de [37], avec le choix entre quatre variétés, dont l'*Amiral* (le plus haut grade) et la *Paragon* (forme la plus parfaite et la plus prisée des Curieux) :

37. Castelijn[9] [= Admiral Catelein].
38. Paragon[10] Castelijn.
39. General Castelijn.
40. Tournay Katlein.

[9] Ce nom propre présente de nombreuses graphies (allomorphie classique dans une période où l'orthographe n'est pas encore fixée) : *Castelijn* = *Catelein* = *Catlein* = *Katlein*.
[10] *Para(n)gon* : « modèle, patron » (Académie, 1694, II : 177). *Se para(n)gonner* [pour une Tulipe] = « dev[enir] plus belle en se perfectionn[ant] » (Monstereul, 1658 : 33).

Le troisième problème, enfin, est celui des noms en général, formés par emprunt, qui s'accompagnent de nombreuses déformations et incompréhensions lors de leur adaptation dans la langue cible – ainsi des emprunts des Hollandais aux fleuristes français [41]-[43] et, après l'heure de gloire des fleuristes français, des Français aux fleuristes hollandais [44] :

41. Fr. *Merveille, Morillon, Paletot*[11]*, Paragon, Viceroy* > Néerl. *Marveilje / Morveille, Morlijon, Palto, Pargoen, Viseroij.*
42. Fr. *Trois couleurs van Nonville* > Néerl. *Troij cluer van Noviel.*
43. Fr. *Agate Robin*[12] > Néerl. *Achat rubin.*
44. Néerl. *Admirael van Engelhandt*[13]*, Admirael Liefkens* > Fr. *Amiral d'Angleterre, Amiral de Lief.*

Dans la nomenclature vernaculaire, peu importe, au fond, l'interprétation littérale et exacte des noms. L'essentiel est surtout de disposer de dénominations :

(i) qui établissent bien, dans et par le système de la langue, une distinction entre les créatures de Dieu que constituent les différentes variétés de tulipes (Monstereul, 1658) ;

(ii) qui soient bien identifiées comme « intensives » et interprétées comme désignant des variétés de grande valeur dans un système économique construit comme celui du marché des pierres précieuses, avec lesquelles la comparaison est d'ailleurs souvent faite – à cet égard, voir l'usage commun à la tulipe et aux pierres précieuses du mot « parangon[14] », ainsi que le recours à une unité de poids spécifique : l'as (0,048 gr.) pour les bulbes, le carat pour les pierres précieuses.

Si la signification importe peu, c'est qu'il revient en général aux illustrations des « livres de Tulipes », colorisées (Swanenburch, 1628 ; Cos, 1637), d'assurer le complément sémantique et sémiotique permettant l'identification visuelle des Tulipes par les acheteurs. Les dénominations vernaculaires ne seraient-elles alors que des noms propres au sens d'étiquettes simplement arbitraires et l'essor néologique introduirait-il un nouveau mode de conception de la nomenclature par rapport à la nomenclature essentialiste des botanistes ?

3.3.2. Principes de structuration : abrègement et binomialité

Au-delà de l'appréhension isolée de telle ou telle dénomination, la nomenclature des variétés s'inscrit en fait dans un système complexe de distinction de

[11] *Paletot* : « Tulippe bigarrée de diverses couleurs » (Furetière, 1690).
[12] Robin (Jean) : botaniste d'Henri IV.
[13] Engelant (François van) : fleuriste de Harlem.
[14] « Se dit des pierres precieuses excellentes » (Furetière, 1690).

« sortes », assez rigoureusement hiérarchisées – en trois niveaux pour ce qui est du système français :

> 45. A. 3 sortes : printanieres, medionnelles, tardives[15];
> B. 4 sortes : de couleur unique ; bordées ; diversement colorées (et obtenues par graines) ; panachées ;
> C. Panachées (p. ex.) de 7 sortes chez les Français : Paltots, Morillons, Agates, Marquetines, Monstres (ou Perroquettes), Jaspées, Fantasques (ou Bizarres).
> *Cf.* les sortes hiérarchisées des Hollandais : Admiral, General…

La nomenclature n'explicite presque jamais les deux premiers niveaux, ce qui constitue un défaut pour les botanistes. En revanche, elle rend presque toujours manifeste le troisième niveau [46], même si on observe quelques exceptions, réservées aux Tulipes très connues et dont la dénomination devient presque un surnom [47], au point qu'elle peut se passer du constituant signifiant le troisième niveau de classification :

> 46. Agates : *Agate Robin.*
> Jaspées : *Jaspée de Harlem.*
> 47. Amirales : *Amiral Gouda* > *(La) Gouda.*
> Fantasques : *Fantasque Genoise* > *(La) Genoise.*
> Parangons : *Parangon d'Acosta* > *(L') Acoste.*

Nous retrouvons ici un principe d'ellipse, déjà vu *supra* dans l'exemple [37]. Encore une telle possibilité de dénomination très elliptique ne peut-elle s'entendre et fonctionner que dans un cadre discursif déjà établi (le thème est celui des Tulipes) et entre connaisseurs spécialistes (curieux ou fleuristes) – sur le mode de l'usage nomenclatural populaire valant souvent pour les plantes uniques, prototypiques ou célèbres, qui s'oppose par sa forme monolexicale au caractère récapitulatif polylexical des nomenclatures savantes[16].

En dehors de ces rares exceptions contextuelles, la nomenclature est donc régulièrement structurée, ce qui infirme les affirmations du botaniste Eslholz sur la fantaisie individuelle des nomenclatures vernaculaires, étant simplement entendu que la structuration est abrégée et restreinte en surface aux deux derniers niveaux de la classification :

> 48. Il [le Curieux] peut luy donner tel nom qu'il luy plaira, pourueu qu'il soit de l'espece susdite, & se rapporte à l'vne des sortes. Par exemple, si cette Tulipe tient nature de Morillon, on la peut nommer <u>Morillon vn tel</u> ; ou si elle est

---

[15] *Medionnelles* : de floraison intermédiaire, entre les printanières (mars) et les tardives (mai-juin) – équivalent de la floraison « incertaine » des botanistes (voir *supra*, § 2).
[16] Cf. dans ces cas l'opposition entre monômes populaires d'espèces (p. ex. *Marjolaine*) et binômes savants linnéens Nom de genre + Adjectif d'espèce (p. ex. *Origanum vulgare* L.).

> Agate, dire Agate telle ; & ainsi du reste. […] Et ainsi chacun doit nommer les Tulipes par jugement, & non par fantaisie (Monstereul, 1658 : 33-34).

## Conclusions

La nomenclature néologique des variétés de tulipes se réalise ainsi en langue vernaculaire de manière économique, relativement à un niveau de base mais sans rappel de la hiérarchie super-ordonnée. Ce schéma abrégé conduit à un patron binomial très fréquent : ainsi peut-on relever que dans la dénomination néologique des variétés de tulipes, 85% des occurrences chez Monstereul (384 sur 451) et 56% chez Eslholz[17] (120 sur 214) ont la forme de binômes. Cette tendance, massive, est celle des nomenclatures populaires (hors cas d'ellipse vus ci-dessus) ; on la retrouve par exemple dans le système nomenclatural patronymique. Mais, surtout, ce schéma n'est pas sans rappeler la nomenclature binomiale que Linné a élaborée sous le qualificatif de « triviale » (*i.e.* vulgaire) et qu'il a développée en marge de la nomenclature scientifique à laquelle il travaillait principalement et qui, elle, était sur le patron récapitulatif de la nomenclature logique de la Renaissance (Selosse, 2008). Cette nomenclature linnéenne marginale, qui est devenue ultérieurement le cœur même de la nomenclature botanique universelle (McNeill, 2006), a été mise au point après un séjour en Hollande chez Georges Clifford, un riche amateur de plantes et de tulipes qui fut le protecteur de Linné : il n'est donc pas exclu que ce soit en Hollande que Linné ait pu voir à l'œuvre le système binomial fréquemment utilisé par les fleuristes et amateurs de tulipes, et qu'il s'en soit inspiré pour sa propre nomenclature.

Quoi qu'il en soit, la nomenclature néologique des Tulipes permet de voir le glissement d'une nomenclature purement scientifique, réaliste et récapitulative (nom du genre et noms des espèces en véhiculaire néo-latin) à une nomenclature plus populaire, pratique et abrégée (noms de variétés en vernaculaires). « Plus populaire » mais pas tout à fait, car si la structuration de cette nomenclature est populaire – ne renouvelant pas le système de la langue mais profitant d'un système de motivation et de cognition « classique » –, en revanche, elle reste un discours de spécialité dont les unités lexicales régulièrement créées ne s'échangent et ne sont comprises que dans un milieu social étroit, celui du marché de la Tulipe, formé des curieux et des fleuristes, autrement dit, de clients et de marchands spécialisés.

## Références bibliographiques

ACADÉMIE (1694), *Dictionnaire de l'Académie françoise dedié au Roy*, A Paris, Chez Jean Baptiste Coignard.

[17] Botaniste qui critique les dénominations populaires en usage mais les cite néanmoins avec une grande précision, à la suite de la nomenclature botanique, à laquelle il accorde le seul statut de véritable nomenclature.

ANONYME (1756), « Traité de la Culture des Fleurs », in J. de La Quintinye, *Instructions pour les Jardins*, t. 2, 359-554, A Paris, Par la Compagnie des Libraires associés.

AYMONIN G. G., éd. (1987), *Basilius Besler. L'Herbier des quatre saisons*, Paris, Citadelles & Mazenod.

BAUHIN C. (1596), *Phytopinax seu Enumeratio plantarum*, Basileae, Petri.

BAUHIN C. (1623), *Pinax Theatri Botanici*, Basileae, König.

BAUHIN C. (s. d.), *Manuscrit n° 4713 bis*, Basler Universitätsbibliothek [= annotations de C. Bauhin sur l'ouvrage de Gobel, 1581].

BELON P. (1553), *Les obseruations de plusieurs singularitez et choses memorables*, A Paris, En la boutique de Gilles Corrozet.

BESLER B. (1613), *Hortus Eystettensis*, Norinbergae, [s. n.].

COLIN A., t. f. (1619), *Histoire des Simples Medicamens apportés de l'Amerique, desquels on se sert en la Medecine*, 2e éd., A Lyon, Aux despens de Iean Pillehotte.

COS P. (1637), *Verzameling van een meenigte tulipaanen*, Haarlem, [s. n.]

DIDEROT D. et *al.* (1778), *Encyclopédie ou Dictionnaire raisonné des Sciences, des Arts et des Métiers*, Berne-Lausanne, Chez les Sociétés Typographiques.

ELSHOLZ J. S. (1663), *Flora Marchica, sive Catalogus plantarum quae partim in hortis electoralibus Marchiae Brandenburgicae… excoluntur*, Berolini, Ex officinia Rungiana, sumptibus Danielis Reichelii.

FULLER Th. ([1655], 1867), « Antheologia, or the Speech of Flowers », in *The Cause and Cure of a Wounded Conscience […]*, 271-321, [London, Printed for John Stafford], London, William Tegg.

FURETIÈRE A. (1690), *Dictionnaire universel, contenant generalement tous les mots françois*, La Haye-Rotterdam, [s. n.].

GESSNER C. (1561), « Casparo Collino Pharmocopoeo », in V. Cordus, *Annotationes in Pedacii Dioscoridis Anazarbei de Medica materia*, 213-216, Argentorati, Iosias Rihelius.

GOBEL S. (1581), *Plantarum seu Stirpium Icones*, Antverpiae, Ex officina Christophori Plantini.

L'ESCLUSE Ch. de (1576), *Rariorum aliquot stirpium per Hispanias obseruatarum Historia*, Antverpiae, Ex Officina Christophori Plantini.

LOBEL M. de, PENA P. (1570), *Stirpium Adversaria Nova*, Londini, T. Purfoetius.

LOBEL M. de (1576), *Plantarum seu Stirpium Historia*, Antverpiae, Ex officina Christophori Plantini.

McNEILL J. et *al.*, éds (2006), *International Code of Botanical Nomenclature (Vienna Code)*, Ruggell, A.R.G. Gantner Verlag KG (*Regnum Vegetabile*, vol. 146).

MONSTEREUL Ch., sieur de la Chesnée ([1654] 1658), *Le Floriste François Traittant de l'Origine des Tulipes […] Auec un Catalogue des noms des Tulipes, & distinctions de leurs couleurs*, 2e éd., A Rouen, Chez Louys du Mesnil.

MORIN P. le jeune (1651), *Catalogues de quelques plantes à fleurs, qui sont de présent au Iardin de P. Morin le jeune, dit Troisiéme, fleuriste*, Paris, De l'Imprimerie de François le Cointe.

MORIN P. le jeune (1655), *Catalogues de quelques plantes à fleurs, qui sont de présent au Iardin de P. Morin, fleuriste*, Paris, De l'Imprimerie de François le Cointe.

MORIN P. le jeune (1658), *Remarques necessaires pour la culture des fleurs. Diligemment obseruées par P. Morin. Auec un Catalogue des Plantes rares qui se trouuent à present dans son Iardin*, Paris, Chez Charles de Cercy.
PASSE C. van de (1614), *Hortus floridus*, Utrecht.
PAVORD A. (2001), *La Tulipe*, Arles, Actes Sud.
ROMAN A. (1637a), *Samen-spraeck*, Haarlem.
ROMAN A. (1637b), *Tweede Samen-spraeck*, Haarlem.
ROMAN A. (1643), *Register van de Prijsen der Bloemen*, Tot Amstelredam, Cornelis Danckaertsz.
SELOSSE Ph. (2004), « Prinzipien der botanischen Nomenklatur in der Renaissance », *The Translation of Domain Specific Languages and Multilingual Terminology Management*, Rita Temmerman & Uus Knops éds., *Linguistica Antwerpiensia*, *New Series*, vol. 3, 133-152.
SELOSSE Ph. (2007), « Die botanische Nomenklatur der Renaissance im Licht ihrer *Épistèmè* », *Zeitschrift für romanische Philologie*, vol. 123 / 2, 213-236.
SELOSSE Ph. (2008), « Les concepts de genre et d'espèce à travers l'évolution du modèle logique de la définition, de l'Antiquité à Linné », *Peut-on classer le vivant ? Linné et la systématique aujourd'hui*, D. Prat, A. Raynal-Roques, A. Roguenant, J.-Cl. Lacharpagne éds., 65-79, Paris, Belin.
SWANENBURCH J. I. van (1628), *Tulpenboek*, Bijzondere Collecties NEHA (Nederlandsch Economisch-Historisch Archief), 254.
SWEERT E. (1612), *Florilegium*, Francfort, [s. n.].

## Index des auteurs

BENGOECHEA Mercedes ..... 203

CALBERG-CHALLOT Marie ..... 121

CAMPETELLA Moreno ..... 231

CANDEL Danielle ..... 121

CONDAMINES Anne ..... 165

DELACROIX Amandine ..... 267

ERLOS Frédéric ..... 13

GARCÍA PALACIOS Joaquín ..... 245

KAGEURA Kyo ..... 37

LAZANA Irene ..... 75

LERSUNDI Mikel ..... 141

MANIEZ François ..... 189

PICTON Aurélie ..... 165

QUIRION Jean ..... 101

REY Christophe ..... 267

ROLDÁN VENDRELL Mercedes ..... 51

SANCHEZ IBAÑEZ Miguel ..... 245

SAN MARTIN Itziar ..... 141

SELOSSE Philippe ..... 279

XYDOPOULOS George J ..... 75

ZABALA Igone ..... 141