

Introduction

La lexicographie en tant que méthodologie de recherche en linguistique¹

Xavier BLANCO

Universitat Autònoma de Barcelona

La revue *Langue(s) et Parole* a décidé de consacrer son numéro 5 à la lexicographie considérée en tant que méthodologie de recherche en linguistique.

L'élaboration de dictionnaires a eu, de tout temps, une influence importante sur la réflexion linguistique. Il ne fait aucun doute, par ailleurs, que les vues linguistiques d'un certain « artisanat » lexicographique (d'Antoine Furetière à Paul Robert, en passant par Émile Littré) étaient beaucoup plus proches de la réalité du lexique que l'optimisme généralisateur de certaines théories linguistiques très répandues au siècle dernier et encore en vogue.

La deuxième moitié du XX^e siècle a vu paraître des dictionnaires importants (le *Dictionnaire du français contemporain* chez Larousse, les dictionnaires Robert, le *Trésor de la langue française*, entre autres) élaborés autant par des linguistes-lexicographes que par des lexicographes-linguistes. Au point que Bernard Quemada, peut-être face à une lexicographie de plus en plus conçue comme objet et moyen d'une étude scientifique de la langue, a créé le terme *dictionnairique* pour se référer aux conditions empiriques d'élaboration et de consultation des dictionnaires.

Les deux dernières décennies du XX^e siècle et les premières années du XXI^e siècle ont assisté, par ailleurs, à l'entrée en force de la *lexicomatique*. Ont été mises à disposition des chercheurs des bases de données lexicales, comme les dictionnaires électroniques DELAS et DELAC (LADL, Paris 7), les dictionnaires de classes d'objets (LLI, Paris 13), la base DiCo (basée sur le *Dictionnaire explicatif et combinatoire du français contemporain*, Université de Montréal), les dictionnaires

¹ Le numéro de revue que nous présentons a été partiellement financé par le *Ministerio de Ciencia e Innovación* espagnol dans le cadre du projet I+D+i COLINDANTE (PID2019-104741GB-I00).

électroniques de Jean Dubois et Françoise Dubois-Charlier, les dictionnaires orientés vers des études morphologiques de Pierre et Danielle Corbin et bien d'autres encore, sans oublier des banques de données terminologiques comme *Termium*, les bases lexicales à orientation diatopique (comme la *Base de données lexicographiques panfrancophone*), les descriptions du français dans le cadre de projets américains comme *WordNet* ou *FrameNet*, ou encore les produits qui ont été intégrés dans des logiciels de traduction automatique comme ceux de Systran ou de correction orthographique et grammaticale comme ceux de Druide informatique inc. Ces productions, qui émanaient de courants linguistiques fort élaborés et avec un haut degré de formalisation, constituaient, à la fois, des outils-clé pour le développement et la recherche dans ces mêmes courants.

La création de ces répertoires lexicaux a été marquée par un fort développement des études en phraséologie, syntaxe lexicale et sémantique formelle, sans que la recherche fondamentale exclue pour autant, bien au contraire, les recherches appliquées dans des domaines comme le traitement automatique des langues, la traduction ou l'enseignement-apprentissage des langues. Ainsi, le terme même de *dictionnaire* a connu une extension sémantique et en est venu à désigner tout un éventail de produits qui étaient, tout à la fois, des résultats et des moyens de recherche des sciences du langage.

D'une façon ou d'une autre, ces dictionnaires au sens large naissaient de la conviction, affichée de façon plus ou moins explicite, que le linguiste ne peut se contenter d'effectuer des observations concernant un nombre limité et aléatoire d'unités lexicales, mais qu'il doit tester ses hypothèses sur l'ensemble ou, du moins, une partie significative du lexique d'une langue. Cette conviction suppose la conjonction de deux activités qui étaient demeurées disjointes pendant des siècles : l'accumulation de données lexicales et la recherche des règles présidant à la combinaison desdites données. La formulation d'une « règle » semble désormais devoir reposer sur une description la plus exhaustive possible de son domaine d'application. La lexicologie (considérée au sens large, c'est-à-dire en incluant la phraséologie, les différents niveaux d'analyse de la langue et les variations diasystématiques) se voit ainsi attribuer une place centrale dans différentes théories linguistiques. Et la lexicographie, envisagée dans cette

perspective, devient une branche expérimentale de la lexicologie. Certains projets, abandonnant déjà le concept 'dictionnaire', s'inscrivent décidément dans des modèles formels innovateurs (comme le *Réseau lexical du français*, Université de Lorraine/ATILF CNRS).

Le présent numéro de *Langue(s) et Parole* « La lexicographie en tant que méthodologie de recherche en linguistique » réunit des études sur le lexique conçus soit comme partie intégrante de la description théorique d'une langue naturelle, soit comme champ d'observation et d'expérimentation dans le cadre d'une théorie ou d'une orientation linguistique concrètes.

Ont contribué à ce numéro des chercheurs ayant participé à la planification, réalisation ou exploitation de dictionnaires à large couverture, qu'il s'agisse de ressources disponibles, de descriptions en cours ou de productions en phase de prototype. Leurs descriptions incluent des réalisations lexicographiques diverses, comme des réseaux lexicaux multidimensionnels, des lexiques de compilation coopérative, des corpus étiquetés, des bases de données terminologiques, entre autres. Ces travaux ont en commun le fait de relever d'une théorisation préalable de la langue et de contribuer au développement des positions théoriques qui les sous-tendent.

Avec son article, *Élaboration d'un nouveau modèle de la langue et lexicographie*, **Denis Le Pesant** aborde de plein fouet le thème de ce numéro de *Langue(s) & Parole* en présentant les travaux des équipes de linguistes qui ont entrepris d'élaborer un lexique-grammaire de la langue française. L'auteur, à la fois chroniste et protagoniste de cette vaste entreprise de description linguistique, réussit le tour de force de faire ressortir en quelques pages les principaux fondements théoriques et les produits lexicographiques qui peuvent être considérés comme relevant de cet important courant linguistique ou compatibles avec cette approche, dans la mesure où ils peuvent être inscrits dans un modèle d'inspiration harrissienne.

En nous servant (sans doute de façon un peu forcée) d'une expression propre aux études historiques, nous oserions dire que l'auteur inscrit sa réflexion « sur la longue durée » (même si la période considérée

ne dépasserait pas une trentaine, peut-être une quarantaine d'années), ce qui lui permet de mettre en évidence les points communs et de regrouper, à très juste titre, dans un même mouvement de fond, des contributions que leurs auteurs auraient peut-être perçues comme fort distinctes les unes des autres. C'est là un mérite, et non le moindre, de la contribution de Le Pesant, qui souligne (et il s'agit là d'une de ses idées-force) que la théorie de la langue se doit d'associer la grammaire au lexique et vice-versa.

La lexicographie qui en résulte se veut ainsi description et du lexique et de la grammaire de sa langue-objet. Les tables de l'ouvrage fondateur de Maurice Gross (*Méthodes en syntaxe*) en constituent l'illustration archétypale. À la description des prédicats viendra s'ajouter la description des noms non prédicatifs, qui est l'un des apports importants (mais qui est loin d'être le seul) de chercheurs comme Gaston Gross, d'une part, et de Jean Dubois et Françoise-Dubois Charlier, d'autre part. Le premier, avec son modèle des classes d'objets, a joué un rôle-clé dans la prise en compte de la sémantique en traitement automatique des langues (dorénavant, TAL). Les deux derniers ont produit des descriptions lexicographiques aussi larges que précises.

Ce n'est pas par hasard que cette lexicographie rendant compte d'un nouveau modèle linguistique a également été pionnière dans l'utilisation d'une nouvelle forme : le dictionnaire électronique, qui ne se confond pas avec le dictionnaire sur support électronique, conçu comme version informatisée des dictionnaires-papier, mais qui est un produit pensé, structuré et formalisé en vue de son exploitation par des logiciels de TAL.

Max Silberztein, informaticien et linguiste, a été le concepteur et le développeur des successives plates-formes d'ingénierie linguistique (la plus récente, NooJ) qui ont permis d'implémenter les systèmes des dictionnaires électroniques de nombreuses langues. Dans sa contribution, *Linguistique et Traitement Automatique des Langues : une coopération est nécessaire*, il passe en revue différentes applications logicielles en TAL et montre, au-delà de tout doute raisonnable, qu'une informatique documentaire qui essaierait de se passer d'une description linguistique solide est vouée à l'échec ou, tout au moins, à l'approximation grossière. L'article est l'aboutissement de décennies de recherche, d'observations et d'expériences et vient confirmer

une thèse que l'auteur a défendue de façon convaincante dans un grand nombre de forums et, récemment, avec l'organisation de l'important workshop *Linguistic Resources for Natural Language Processing* (2018), intégré dans une des plus prestigieuses séries de congrès internationaux en linguistique informatique : l'*International Conference on Computational Linguistics* (COLING, dont l'édition de 2018 a eu lieu à Santa Fe, New Mexico, USA).

La thèse centrale de l'article de Max Silberztein est aussi simple que porteuse : les applications en TAL visent, en grande mesure, le traitement du sens. Or, le moyen d'accéder au sens est l'analyse de la forme. Les logiciels de TAL doivent, d'une façon ou d'une autre, établir donc un lien forme-sens.

La nature de la langue est telle qu'essayer d'établir un tel lien sans intégrer des connaissances linguistiques massives et détaillées est sans espoir et sans issue. Ces connaissances linguistiques concernent tous les niveaux d'analyse de la langue : orthographique (si l'on reste dans le domaine de l'écrit), morphologique, syntaxique et sémantico-pragmatique. L'unité lexicale est le lieu de rencontre de ces niveaux d'analyse, car elle joue le rôle d'unité de traitement possédant une (ou un ensemble de) forme(s) graphique(s), des paradigmes morphologiques (flexionnel et dérivationnel), une série de constructions syntaxiques (qu'elle instaure dans le discours ou au sein desquelles elle peut être intégrée, selon que l'on considère la valence active ou la valence passive) et un sens calculable uniquement à partir de la prise en compte de toutes ces données. Silberztein montre, de nombreux exemples à l'appui, que seule la prise en compte d'une information linguistique formalisée sous forme de dictionnaires et de grammaires électroniques permet d'identifier correctement les unités d'analyse.

Les systèmes fondés sur la manipulation statistique de formes graphiques conduisent trop souvent à des résultats inexacts ou inapplicables parce qu'ils se révèlent incapables d'identifier correctement les unités de base pertinentes. L'article met également en relief les principaux problèmes d'analyse liés à la phraséologie (locutions et collocations) et montre bien jusqu'à quel point le projet d'une phraséologie calculable à partir de la fréquence de co-occurrence n'est pas envisageable. Max Silberztein pose ainsi la nécessité même de l'existence de toute une discipline : la linguistique informatique conçue comme la

compilation et la formalisation d'une immense quantité de données hautement spécialisées. Il met en évidence, en même temps, les limitations d'une informatique qui essaierait de court-circuiter la linguistique.

Dans leur article, *Problèmes formels concernant la traduction des adverbes composés espagnol/portugais*, **Dolors Català**, **Jorge Baptista** et **Cristina Palma**, forts de leur longue et fructueuse expérience en description des langues espagnole et portugaise et experts en linguistique contrastive et en implémentations informatiques des données lexicales, nous offrent une analyse poussée des relations d'équivalence de traduction qui peuvent être établies entre deux dictionnaires électroniques d'adverbes composés.

Leur travail s'inscrit dans la même tradition de lexiques-grammaires à laquelle appartiennent les deux auteurs précédents et s'inspire de la description proposée par Maurice Gross dans le troisième volume de sa grammaire transformationnelle du français, sa *Syntaxe de l'adverbe* (1986), ce qui leur permet de créer, de façon indépendante, des descriptions linguistiques pleinement compatibles entre elles.

Le travail en description monolingue, tout en gardant pleinement son intérêt et ses possibilités d'application au sein d'un seul système linguistique, débouche ainsi naturellement sur des applications bilingues et multilingues robustes. La description moyennant des transducteurs présente, en plus, de nombreux avantages : elle offre un cadre d'analyse rigoureux et vérifiable à tout moment sur des *corpora* textuels, ainsi qu'une présentation élégante dotée de force explicative et pouvant être mise à l'instant à disposition de la communauté de chercheurs sur des plates-forme d'ingénierie linguistique comme NooJ (Cf. Silberztein ci-dessus). Cette méthodologie d'élaboration de ressources linguistiques détaillées, explicites, cumulables, comparables, réutilisables et partageables est une réponse efficace au problème endémique de la limitation de ressources humaines et financières et permet d'optimiser et de rentabiliser l'effort de description. Les auteurs nous présentent, donc, beaucoup plus qu'une description des adverbes composés, ils nous montrent la voie (ou, en tout cas une des voies les plus prometteuses) que devra suivre le labeur du linguiste du XXI^e siècle.

Rafael García Pérez, qui a fait d'importants apports à la description diachronique de la langue espagnole à travers le *Nuevo Diccionario Histórico del*

Español (NDHE), propose, dans son article, *Nominalisations de verbes supports dans un dictionnaire historique : les supports de la classe <infraction pénale>*, une description lexicographique adaptée au traitement d'unités « support », désémantisées ou relativement désémantisées.

L'exemple choisi est celui de la classe <infraction pénale> (nous devons la dénomination de cette classe au propre auteur), actualisée par des supports comme *hacer*, *cometer*, *perpetrar*, *incurrir* [en], *consumar* ou *ejecutar*. L'auteur se penche sur la question de la nominalisation des verbes supports (en partie, propre aux langues de spécialité) appartenant à la classe retenue comme exemple et à travers les différentes étapes de l'histoire de la langue espagnole. Il fait d'intéressantes propositions de structuration d'un article lexicographique que l'on pourrait désigner « à géométrie variable ». En effet, tout en mettant à profit les possibilités offertes par le support numérique, le lexicographe procéderait à un choix d'exemples permettant à l'utilisateur non spécialiste d'accéder facilement à l'évolution de l'usage et à la signification du mot-vedette. L'ensemble des données disponibles resterait cependant accessible à l'utilisateur spécialiste souhaitant de plus amples renseignements.

L'article de García Pérez présente, par surcroît, l'intérêt d'offrir au lecteur francophone la description des lemmes espagnols *perpetración* et *perpetrador*, en métalangue française, excellente illustration de la description lexicographique en œuvre dans le NDHE.

Finalement, il y a lieu de mettre en relief que l'auteur propose également une section de l'article lexicographique rendant compte des classes de prédicats avec lesquelles se combine le mot-vedette. Ce n'est pas une question mineure, puisqu'à force de focaliser sur les contraintes syntactico-sémantiques qu'un prédicat impose à ses arguments, on risque de négliger que le prédicat lui-même peut fonctionner (et fonctionne souvent) comme argument. C'est par là que la langue casse le carcan de la logique de prédicats du premier ordre et acquiert cette souplesse et cette capacité de condensation qui en fait le système informationnel par excellence. Décrire ces complexités reste un défi que García Pérez a accepté de relever.

Autorité reconnue en phraséologie espagnole, **Mario García-Page** nous donne, dans son article, *Locuciones españolas con nombre de instrumento*

musical, un excellent exemple de l'exploitation des dictionnaires pour la création de ressources en lexicologie. García-Page présente, classe et décrit les locutions espagnoles comportant un nom d'instrument de musique en suivant une méthodologie de recherche parfaitement explicite et exposée de façon impeccable, des sources clairement identifiées et une démarche d'extraction de données qui, pour l'exprimer dans les termes de la discipline de recherche d'information (RI), garantit aussi bien un rappel exhaustif qu'une précision maximale (toutes les unités pertinentes sont extraites et toutes les données extraites sont pertinentes).

Les trois classes basiques d'instruments sont représentées : instruments à corde, instruments à vent et instruments de percussion. L'auteur observe que la plupart des unités retenues correspondent à des locutions verbales à structure V+SN où le nom d'instrument constitue le noyau du syntagme nominal.

L'article comporte deux annexes où toutes les locutions relevées sont présentées et définies et où les sources d'où elles ont été tirées sont indiquées. Cette transparence absolue, la mise à disposition immédiate des données résultant de la recherche, la reproductibilité des procédures et des analyses sont des aspects que les nombreux apports de García-Page partagent avec les travaux du lexique-grammaire, courant dont l'auteur ne se réclame pourtant pas.

Dans son important article, *Apports de la lexicographie réticulaire : le cas de la polysémie relationnelle*, **Alain Polguère**, concepteur du Réseau Lexical du Français (RL-fr) à l'Université de Lorraine/ATILF-CNRS (Analyse et Traitement Informatique de la Langue Française), présente les caractéristiques formelles des Systèmes Lexicaux et montre comment on peut modéliser de façon relationnelle le traitement de la polysémie au sein du système lexical qu'est le RL-fr (projet d'une immense envergure et d'une formidable capacité descriptive).

À la différence de la lexicographie textuelle, qui offre un texte structuré en parties (avec toutefois une composante relationnelle : la médiestructure, ou système de renvois), la lexicographie réticulaire prônée par l'auteur construit des graphes qui brisent la barrière de l'organisation linéaire de l'information et permettent une navigation

multidimensionnelle à travers le lexique. Un Système Lexical est un graphe orienté dont les nœuds correspondent aux unités lexicales à décrire et les arcs à deux familles de relations lexicales : les relations paradigmatisées de dérivation sémantique et les relations syntagmatiques de type base-collocation (représentées à l'aide des fonctions lexicales). Ce type de modèle du lexique, non ontologique, présente des avantages descriptifs importants : il modélise des phénomènes linguistiques comme la proximité sémantique, permet une description relativiste (c'est-à-dire, l'introduction, sans risque, des données associées à un degré de confiance non maximale) et reflète le lexique mental mieux que les dictionnaires traditionnels.

Par ailleurs, la polysémie peut être modélisée au sein du RL-fr comme une relation entre lexies (copolysémie) représentée par des arcs étiquetés du Système Lexical. Les relations de sens entre les copolysèmes sont, donc, explicitement caractérisées et l'importance de l'intersection des sens est reflétée par une indication de poids (soit fort, soit faible). Les structures polysémiques sont ainsi réifiées et peuvent être visualisées et faire l'objet de calculs topologiques, ce qui permet de rendre compte de façon élégante de phénomènes comme la polysémie régulière. La « géométrie du lexique » est mise en évidence et permet d'envisager, d'ores et déjà, des progrès significatifs aussi bien descriptifs que théoriques. Pour terminer, il est important de souligner que les observations de Polguère se fondent sur l'analyse de 9.141 liens de copolysémie décrits dans le RL-fr. L'élaboration de la description du lexique va de pair, donc, avec la recherche en lexicologie, l'une permettant de rétroalimenter l'autre en permanence.

Marie-Sophie Pausé, rédactrice du *Réseau lexical du français*, a apporté des contributions particulièrement significatives à la description des phrasèmes non connexes, en commençant par la propre identification et définition de ce type de phrasèmes.

Elle nous présente dans son article, *Modélisation des locutions dans un réseau lexical*, un moyen formel de représenter la flexibilité syntaxique des locutions et leurs possibilités de défigement. Pour ce faire, elle identifie et décrit avec précision le patron syntaxique de la locution et ses constituants lexicaux, qui constituent ce que l'auteure appelle la *structure lexico-syntaxique* de la locution. Pausé montre ainsi que cette structure est

doublée d'une structure communicative, tout comme les syntagmes libres, ce qui s'avère particulièrement porteur. En effet, la structure syntaxique de la locution peut, à un moment donné, récupérer ses propriétés combinatoires. Pour décrire ce phénomène, Pausé propose également la notion de *projection structurale*, qui lui permet de décrire le parallélisme entre le sens d'une locution et sa structure lexico-syntaxique. On voit ainsi comment certaines composantes de certaines locutions deviennent soit des lexies néologiques (comme, p. ex., *page* de *tourner la page* dans des énoncés comme *la page qu'il a tournée hier* ou *je tournerais la page de ma vie de prof*), soit de nouvelles locutions (comme *main verte*, qui s'est petit à petit désolidarisé de la locution verbale *avoir la main verte* pour se constituer en vraie locution nominale dans, p. ex., *nous cherchons une personne avec une main verte*).

Pausé réussit de la sorte à décrire formellement le défigement qui relève, certes, de la créativité linguistique, mais d'une créativité linguistique qui est néanmoins soumise à de fortes contraintes syntactico-sémantiques. En même temps, cela lui permet de rendre compte du processus de changement linguistique.

Qu'il nous soit permis de souligner ici l'intérêt des travaux de Marie-Sophie Pausé, dont l'article publié dans le présent numéro est un aboutissement important. Pausé a été capable de proposer, dans le cadre de la Lexicologie Explicative et Combinatoire, une description formelle de ce qui a été observé (sans arriver pour autant à le décrire vraiment) dans de nombreuses études sur le défigement. Ce faisant, elle met en rapport synchronie et diachronie, parce que, d'une part, elle montre qu'il est possible de capturer le changement linguistique lorsqu'il est en train de se produire et, d'autre part, qu'on peut se doter de moyens de le prédire par analogie. Or, quand une description commence à avoir une valeur prédictive, on sait que l'on a dépassé la description pour atteindre la modélisation. Pour y arriver, il faut, comme l'a fait Pausé, postuler des notions théoriques solides et remettre en question la définition même de locution comme unité atomique de sens en la présentant désormais comme une unité douée d'une structure syntaxique « vivante », prête à engendrer de nouvelles réalités lexicales.

Dans son article, *Sport et culture : de la définition encyclopédique à la définition lexicographique systématique*, **Araceli Gómez Fernández** analyse

les définitions du vocable SPORT dans le *Dictionnaire culturel du sport* de Michaël Attali et Jean Saint-Martin en tenant compte d'autres dictionnaires français orientés au grand public.

Tout d'abord, l'auteure retrace l'évolution du sens des lexies *sport* à partir de quelques dictionnaires importants du XIX^e et du XX^e siècles. Ensuite, elle précise les similitudes et dissimilitudes entre différents types de dictionnaires : dictionnaire culturel, dictionnaire de langue, dictionnaire encyclopédique, dictionnaire spécialisé... Puis, dans la troisième partie de son article, Gómez Fernández fait une présentation du *Dictionnaire culturel du sport* et se penche, de façon particulière, sur les patrons de définition. Elle montre que les définitions des entrées relevant de l'hyperonyme *sport* (*athlétisme, tennis, water-polo...*) ne suivent pas un patron précis et, en se fondant sur la Lexicologie Explicative et Combinatoire, elle illustre l'importance et l'intérêt de disposer, d'une part, d'un patron de définition stable qui fasse ressortir le lien sémantique existant entre ces lexies et, d'autre part, d'une composante encyclopédique à visée culturelle.

Par ailleurs, outre les aspects mentionnés ci-dessus, l'article d'Araceli Gómez-Fernández a le mérite ajouté d'offrir tout un panorama, succinct, mais suggestif, des principaux monuments lexicographiques de la langue française.

L'article *Ontologie des marques de domaines appliquée aux dictionnaires de langue générale*, signé par une équipe de chercheurs formée par **Rute Costa, Sara Carvalho, Ana Salgado, Alberto Simões** et **Toma Tasovac**, rend compte d'une contribution remarquable à la numérisation (hiérarchisée, structurée et annotée) de l'information lexicale existant dans des dictionnaires papier, afin de réutiliser ladite information dans des applications liées au Web sémantique.

Après une analyse en profondeur de dictionnaires-papier, de classifications générales comme la CDEU, le thésaurus de l'UNESCO, les *WordNet Domains*, l'*EuroSciVoc*, la *ModernSCience Ontology* et de classifications spécifiques aux sciences de la santé comme SNOMED CT, MeSH, *Unified Medical Language System* ou *Disease Ontology*, les auteurs ont développé une ontologie regroupant tous les domaines qui relèvent des sciences médicales et de la santé : *OntoDomLab-Med*.

À l'aide de ladite ontologie, l'article montre comment relier, en suivant une méthodologie hybride, les marques d'*OntoDomLab-Med* et des entrées de dictionnaire balisées en TEI-Lex-0.

Bien entendu, une telle entreprise exige une réflexion poussée sur le rôle et la typologie des marques de domaines en lexicographie traditionnelle, qui est fournie (en référence au *Dicionário da Língua Portuguesa Contemporânea*, dorénavant DLPC) dans la section 2 de l'article. Ensuite est présentée et discutée la notion cruciale d'*interopérabilité sémantique*, comprise (nous simplifions ici la définition de cette notion) comme la capacité de deux ou plusieurs systèmes d'échanger de l'information et d'utiliser l'information échangée.

L'interopérabilité sémantique nécessite, donc, la capacité de comprendre aussi bien la structuration d'un message produit au sein d'un système donné que sa signification. Les ontologies (au pluriel), qui se distinguent nettement de l'ontologie (au singulier) en tant que branche de la philosophie qui étudie l'être, sont des artefacts computationnels qui permettent de structurer formellement des systèmes d'information moyennant une certaine conceptualisation. Elles sont devenues des entités cruciales pour le Web sémantique, dans la mesure où elles permettent de garantir et d'automatiser la communication entre différents systèmes d'information afin de répondre à des requêtes ou d'effectuer des tâches.

Une fois ces notions de base bien établies, l'article de Costa et al. donne un excellent exemple d'analyse et d'échange de données tout en mettant en évidence certaines inconsistances au sein de la description lexicographique du DLPC (qui se dote d'emblée d'une liste de 184 étiquettes de domaines mais ne se sert que de 173, alors qu'il utilise certaines marques qu'il n'identifie pas explicitement). Finalement, les auteurs expliquent comment il est possible de mettre en correspondance chacun des sens présents dans le DLPC (toujours dans le domaine de la santé) avec une ou plusieurs classes définies dans *OntoDomLab-Med*. Cela leur permet de montrer comment les données présentes dans le dictionnaire de langue peuvent être réutilisées dans le cadre du TAL tout en contribuant, en même temps, à renforcer la cohérence logique du dictionnaire.

Le numéro 5 de *Langue(s) et Parole* se clôt par une contribution toute spéciale. Teresa Lino (*Universidade Nova de Lisboa*), grande linguiste et amie très chère, avait accepté de participer à ce numéro avec un article portant sur un projet dans lequel elle était particulièrement active : le *Dicionário Português Europeu - Árabe Padrão*. Son décès, le 28 décembre 2019, l'a empêchée de terminer cet article.

Teresa Lino avait non seulement produit une œuvre de grande envergure dans le domaine de la lexicologie, la lexicographie et la terminologie, mais elle avait aussi contribué à former une quantité remarquable de brillants lexicographes-terminographes en langue portugaise. Quelques-uns de ses collègues ont entrepris de conclure l'article en lui donnant une orientation tout à fait pertinente à la thématique de la revue. Tout en restant dans le cadre du projet du *Dicionário*, elles ont choisi de cibler leur analyse sur la terminologie des termes des énergies renouvelables extraits de corpus bilingues portugais-arabe.

Leur contribution, *Terminologie vulgarisée des énergies renouvelables*, signée **Teresa Lino, Karima Fangour Robalo, Sílvia Barbosa, Olga Heitor, Fátima Ferreira et Catarina Lino**, a le mérite de regrouper les grands domaines d'activité de notre regrettée professeure (rappelons-le : lexicologie, lexicographie et terminologie) en y intégrant l'approche de la didactique et de la linguistique comparée, disciplines qu'elle avait également abondamment cultivées.

Après une introduction générale à la lexicographie, où une attention spéciale est portée à la lexicographie bilingue et à la définition (élément de microstructure qui est rarement associé à la lexicographie bilingue, mais dont les auteures montrent bien l'intérêt, à condition de considérer un type de définition adaptée à ce type de lexicographie), les auteures entrent dans le vif du sujet en présentant, d'une part, le projet du *Dicionário Português Europeu - Árabe Padrão* (qui est en grande mesure un dictionnaire d'apprentissage du portugais adressé à un public arabophone) et, d'autre part, la méthodologie d'extraction des termes vulgarisés relevant des énergies renouvelables (ce qui implique la présentation et discussion de la notion de vulgarisation scientifique). Cette méthodologie se sert, entre autres, de l'identification de paradigmes définitionnels et de paradigmes désignationnels ou désignatifs. Des

exemples de termes vulgarisés concernant l'énergie éolienne, l'énergie hydraulique, l'énergie géothermique et la biomasse sont proposés et discutés. La conclusion de l'article inclut une citation capitale de Teresa Lino, où elle souligne la nécessité de prendre en compte des phénomènes relevant de la socio-terminologie et la terminologie culturelle dans les processus d'extraction et de traitement de données linguistiques procédant de corpus.

Retenons, comme conclusion à cette présentation du numéro 5 de *Langue(s) & Parole*, que ce volume a une entité organique et que, tout en restant un numéro de revue, le résultat final n'est pas loin d'un volume collectif où s'entrecroisent les langues française, espagnole et portugaise avec des approches théoriques différentes (mais compatibles et même complémentaires) et des champs d'application comme le TAL, la lexicographie monolingue et bilingue, la métalexicographie ou la terminographie.

Le lecteur est, bien entendu, libre de ne choisir que le ou les articles pouvant l'intéresser de façon plus directe, mais il nous semble qu'une lecture suivie de tout le numéro pourra lui apporter une vision panoramique d'un champ de recherches prometteur. Un des problèmes cruciaux du XXI^e siècle sera, sans doute, l'accès au sens linguistique et son traitement automatique. Que cela se fasse par des approximations purement statistiques ou que cela soit accompli moyennant des algorithmes pouvant accéder à de grandes quantités de ressources linguistiques formalisées (héritage et témoignage aussi bien de la linguistique et de la sémiotique du XX^e siècle que des siècles de descriptions lexicographiques et grammaticales qui les ont précédées) est un enjeu de taille comparable à la récupération du savoir classique par l'Humanisme du XVI^e siècle. Ce ne sont probablement pas (hélas !) les étudiants de nos maltraitées facultés des Lettres et de Traduction qui recevront cet héritage. Il est fort probable qu'avant la fin du siècle où nous vivons, et qui s'annonce riche en bouleversements profonds des modèles productifs et sociaux de l'humanité, la linguistique soit en grande partie une affaire d'intelligence artificielle. Reste à savoir s'il s'agira de robots-linguistes ou de robots-statisticiens.

En tant que coordonnateur de ce numéro, je tiens à remercier chaleureusement tous les auteurs/-es. C'est également un plaisir pour moi de remercier mes collègues et amis de l'UAB, Julio Murillo et Roser Gauchola. Sans leur support, leurs encouragements et leur savoir-faire éditorial, ce numéro n'aurait jamais vu la lumière. Je remercie également les évaluateurs et les réviseurs des articles. Je ne saurais terminer sans un souvenir ému pour Teresa et ses *palavras que nunca morrerão*.