

CAHIERS DE LINGUISTIQUE

Contenu du présent fascicule :

Avant-propos	5
Xavier BLANCO ESCODA, Marie-Claude L'HOMME et Marc VAN CAMPEN-HOUDT	
Les expressions de temps productives. Contribution à leur description dans une perspective lexicologique	9
Jorge BAPTISTA et Dolors CATALÀ	
Synonymie et restrictions lexicales dans un dictionnaire historique	25
Rafael GARCÍA PEREZ	
Constitution et analyse d'un lexique de mots vietnamiens d'origine française	41
Sabine HUYHN	
Du corpus au dictionnaire. Réalisation automatique d'un outil de gestion de l'information multilingue	63
Bernard JACQUEMIN et Sabine PLOUX	
Un dictionnaire syntaxico-sémantique bilingue de la zootechnie	85
Brigitte LÉPINETTE et Françoise OLMO CAZEVIEILLE	
Les mutations de la lexicographie touristique	105
Patrick LEROYER	
Ambiguïté de la synonymie en lexicographie française	119
Jean B. NTAKIRUTIMANA	
Quelle terminologie pour la société de l'information ?	143
Christophe ROCHE	
Vers un dictionnaire de collocations multilingue	171
Amalia TODIRASCU, Ulrich HEID, Dan ȘTEFANESCU, Dan TUFİȘ, Christopher GLEDHILL, Marion WELLER, François ROUSSELOT	



9 782930 481524

ISSN : 0771-6524
ISBN : 978-2-930481-52-4
Dépôt légal : 2008/9202/17
Prix de vente : 25.00 EUR

2007 [2008] - 33/1

CAHIERS DE LINGUISTIQUE

CAHIERS DE LINGUISTIQUE

Revue de sociolinguistique
et de sociologie de la langue française

Lexique, dictionnaire
et connaissance
dans une société multilingue

Xavier Blanco Escoda,
Marie-Claude L'Homme
et Marc Van Campenhoudt, éd.

Réseau lexicologie, terminologie, traduction

2007 [2008] - 33/1

E.M.E.

Cahiers de Linguistique

Revue de sociolinguistique et de sociologie du français

Une langue ne se réalise pas de la même façon dans deux contextes différents : il y a la langue d'ici et celle de là-bas, il y a la langue du manœuvre et celle du cadre supérieur, il y a la langue du comptoir et celle de la tribune, etc. En outre, une langue évolue sans cesse sous l'effet des changements démographiques, des changements technologiques, des changements politiques, etc.

Les *Cahiers de Linguistique*, qui poursuivent le travail éditorial entrepris en 1972 par les *Cahiers de l'Institut de Linguistique de Louvain*, sont désormais plus spécifiquement consacrés à l'étude des rapports entre une langue - le français - et la société ou, plutôt, les sociétés où elle est employée. Les *Cahiers de Linguistique* paraissent à raison de deux fascicules par an d'environ 200 pages chacun. La rédaction privilégie les numéros thématiques.

Rédacteur en Chef :

Christian Delcourt (76a, route de Seraing, BE - 4122 Neupré (Christian.Delcourt@ulg.ac.be)).

Comité de Concertation :

Philippe Blanchet (Rennes), Thierry Bulot (Rennes), Jules Duchastel (Montréal), Michel Francard (Louvain), Françoise Gadet (Paris), Jean René Klein (Louvain), Jean-Marie Klinenberg (Liège), Marie-Rose Simoni-Aurembou (Paris), Pierre Swiggers (Leuven), Dominique Willems (Gand).

Correspondance et propositions d'articles ou de thématiques :

Les articles, les propositions de thématiques, les ouvrages pour recension et la correspondance sont à adresser au Rédacteur en Chef. Prière de demander la feuille de style et les conditions de publication. Les offres et les exemplaires d'échanges sont à adresser au Rédacteur en Chef.

Abonnements et commandes :

Prix de l'abonnement annuel : 50.00 €. (cf. également en troisième de couverture les conditions spéciales pour un abonnement au bouquet de revues de linguistique française), frais de port et d'emballage en sus, à adresser à :

E.M.E. (Éditions Modulaires Européennes) & **InterCommunications** s.p.r.l.
40, rue de Hanret
BE - 5380 Cortil-Wodon

Tél. : +[32]81.83 42 63 et +[32]473.93 46 57
Fax : +[32]81.83 52 63
Courriel : edition@intercommunications.be
Site : www.intercommunications.be

Un bouquet de revues de linguistique française

Quatre revues belges, spécialisées dans différentes approches de la langue française, se sont associées étroitement de manière à pouvoir fournir une livraison simultanément, deux fois par an. Elles proposent aux spécialistes du français et aux institutions un abonnement direct et groupé avec une réduction de 20 % sur le montant des abonnements à ces revues.

Rejoignez ce groupement de spécialistes qui étudient la langue française sous des angles divers. Contactez-nous, écrivez-nous, proposez-nous vos contributions ou des numéros thématiques.

« Cahiers Français et Société »

Revue de politique et d'aménagement linguistique du français.

« Le Langage et l'Homme »

Revue de didactique du français

« Cahiers de Linguistique »

Revue de sociolinguistique et de sociologie du français

« Linguistique française »

Revue d'étude de la langue et du discours

Prix de l'abonnement annuel groupé aux quatre revues : 136.00 €. , frais de port et d'emballage en sus.

Table des matières

Avant-propos	5
Xavier Blanco Escoda, Marie-Claude L'Homme et Marc Van Campenhoudt	
Les expressions de temps productives. Contribution à leur description dans une perspective lexicologique	9
Jorge Baptista et Dolors Català	
Synonymie et restrictions lexicales dans un dictionnaire historique	25
Rafael García Perez	
Constitution et analyse d'un lexique de mots vietnamiens d'origine française	41
Sabine Huyhn	
Du corpus au dictionnaire. Réalisation automatique d'un outil de gestion de l'information multilingue	63
Bernard Jacquemin et Sabine Ploux	
Un dictionnaire syntaxico-sémantique bilingue de la zootechnie	83
Brigitte Lépinette et Françoise Olmo Cazeveille	
Les mutations de la lexicographie touristique	103
Patrick Leroyer	
Ambiguïté de la synonymie en lexicographie française	117
Jean B. Ntakirutimana	
Quelle terminologie pour la société de l'information ?	139
Christophe Roche	
Vers un dictionnaire de collocations multilingue	161
Amalia Todirascu, Ulrich Heid, Dan Ștefanescu, Dan Tufiș, Christopher Gledhill, Marion Weller, François Rousselot	

Avant-propos

Xavier Blanco Escoda

Université autonome de Barcelone – Département de philologie
française et romane, Espagne, Xavier.Blanco@uab.cat

Marie-Claude L’Homme

Université de Montréal, Observatoire de linguistique Sens-Texte
(OLST), Québec, Canada, mc.lhomme@umontreal.ca

Marc Van Campenhoudt

Haute École de Bruxelles, Institut supérieur de traducteurs et
interprètes, Centre de recherche Termisti, mvc@termisti.be

Le réseau *Lexicologie, terminologie, traduction* (LTT) a été fondé à Ottawa en 1988, ce qui fait de lui le plus ancien des réseaux de langue de l’Agence universitaire de la francophonie (AUF). Plaçant les problématiques de multilinguisme au cœur de ses préoccupations, il n’offre en aucun cas une position de repli frileux aux nostalgiques d’une langue française hégémonique. Au cours de la dernière année, quelque cinq cents chercheurs issus de tous les continents ont renouvelé leur affiliation. Ils ont pour point commun d’utiliser le français comme langue de partage des résultats de leurs travaux et d’adhérer aux objectifs de l’AUF : contribuer au désenclavement des chercheurs à travers des coopérations scientifiques Nord-Sud qui répondent aux besoins de la société.

La proposition des *Cahiers de linguistique* de confier au réseau LTT la direction d’un numéro a tout de suite rencontré l’adhésion de son comité scientifique, qui nous a demandé d’en assurer la responsabilité éditoriale. En son nom, nous tenons à remercier notre collègue Christian Delcourt (Université de Liège), directeur des *Cahiers*. On reconnaîtra aisément dans cette initiative une ligne d’action qui l’a conduit à soutenir de longue date le mouvement de la francophonie universitaire.

Nos sociétés vivent une révolution qui est celle du partage des connaissances dans un monde numérique désormais sans frontières. Le produit dictionnaire classique, monolingue ou bilingue, présenté selon un modèle hérité des impératifs de l’imprimerie – toujours valorisé par l’enseignement – connaîtra encore de bons succès de librairie. Toutefois, l’image du dictionnaire se brouille progressivement auprès du public, des

produits présentés sous cette appellation étant fréquemment intégrés à d'autres applications : éditeurs de textes, correcteurs grammaticaux, téléphones mobiles, agendas électroniques... Les facilités de diffusion offertes par Internet permettent également à un grand nombre de personnes, physiques ou morales, de diffuser à très grande échelle des données lexicales de qualité variable. Dans le même temps, l'ingénierie de la connaissance développe un nombre croissant de technologies susceptibles de remettre en cause à moyen terme les techniques mêmes de conception des outils lexicographiques. On songera, certes, à la traduction automatique, à la traduction assistée ou à l'intelligence artificielle, mais aussi aux nouvelles perspectives offertes par XML et par les nouveaux développements en cours dans le cadre de l'élaboration du *web* sémantique. Les recherches entreprises ne conduiront-elles pas, *in fine*, à un hiatus de plus en plus important entre les nouvelles méthodes de la pratique lexicographique et la réalité des produits accessibles au grand public ?

Nourri de ces constats, l'appel adressé fin 2007 aux membres du réseau LTT proposait de rassembler des contributions autour du thème « Lexique, dictionnaire et connaissance dans une société multilingue ». Quatorze articles ont été soumis au réseau, neuf ont été finalement retenus au terme du processus d'évaluation scientifique. Alors que l'appel couvrait un large éventail de thématiques, dans le champ de la linguistique appliquée ou théorique, du dictionnaire classique au *web* sémantique, on se retrouve, finalement, avec une palette de sujets plus restreinte, mais non moins riche en connaissances⁽¹⁾.

Dans les articles retenus, on voit se dessiner certains axes thématiques. Parmi ceux-ci, on observe des problèmes liés aux aspects linguistiques des unités lexicales (le traitement de la synonymie, la représentation de l'information combinatoire), la représentation des données lexicographiques (le traitement lexicographique du multilinguisme, la question du support ou des supports pour les produits lexicographiques) et les méthodes informatiques visant à enrichir des ressources existantes (l'exploitation automatisée de ressources textuelles et la construction semi-automatique de nouvelles ressources lexicographiques). On note toutefois que ces thématiques sont rarement abordées de façon exclusive. En effet, l'intersection des pôles d'intérêt semble être privilégiée dans les différentes contributions à ce numéro. La mise au point d'une application informatique reposera sur une définition rigoureuse d'une propriété linguistique. Inversement, une définition formelle d'un aspect linguistique des unités lexicales débouchera sur un formalisme destiné à le représenter.

Parmi les thématiques plus spécifiques abordées dans ce numéro spécial, le lecteur trouvera des études lexicologiques détaillées menées à bien avec des méthodologies rigoureuses et novatrices. Ainsi, l'article de J. Baptista et D. Català offre une analyse minutieuse des expressions (adverbiales) de temps et aborde, notamment, les difficultés de représentation lexicographique de ces suites extrêmement productives, en adoptant une perspective contrastive français-portugais-espagnol. Adoptant

(1) Nous remercions M^{me} Nathalie Lemaire, du Centre de recherche Termisti, qui a accepté de prendre en charge la prédiction des textes.

une démarche résolument métalexicographique, S. Huynh, quant à elle, construit un lexique de mots vietnamiens empruntés au français, qu'elle analyse aussi bien du point de vue formel que sémantique.

Comme J. Baptista et D. Català, R. García Pérez s'intéresse à des questions de traitement lexicographique, dans son cas, celui des synonymes verbaux. Son étude s'inscrit dans un projet lexicographique de très grande envergure (le *Nuevo Diccionario Histórico de la Lengua Española*) et envisage la perspective diachronique en lexicographie, aussi bien par rapport à l'axe paradigmatique (synonymie) qu'à l'axe syntagmatique (restrictions combinatoires des différents synonymes). La synonymie constitue aussi l'objet d'étude principal de l'article de J. Ntakirutimana, qui vise une base de données dictionnaire pour la désambiguïsation des synonymes en tenant compte du contexte immédiat d'emploi (ou cotexte).

Les questions de traitement lexicographique de la combinatoire lexicale figurent également au centre des contributions de Br. Lépinette et Fr. Olmo, d'une part, et de celle de A. Todirascu et de ses collègues, d'autre part. Constatant certaines lacunes d'un dictionnaire terminologique de facture classique qui met l'accent sur l'information encyclopédique, Br. Lépinette et Fr. Olmo proposent de l'enrichir en y ajoutant des données sur le comportement linguistique des termes de la zootechnie, plus précisément des données de nature syntactico-sémantique. Elles montrent, par ailleurs, comment les rubriques lexicographiques présentées s'adaptent parfaitement à une perspective plurilingue.

En abordant l'axe de la spécialisation non pas à partir du lexique mais plutôt à partir du type de dictionnaire, P. Leroyer mène une réflexion sur les dictionnaires touristiques en mettant l'accent sur les transformations que subissent ces produits, attribuables sans doute à la diversification des supports mais aussi aux besoins des usagers en matière de contenus et d'accès aux contenus. L'auteur situe le multilinguisme au cœur de sa réflexion, en considérant un bon nombre de langues, dans certains cas, des langues typologiquement éloignées.

Le multilinguisme est également au cœur des travaux d'A. Todirascu et de ses coauteurs. L'équipe propose une méthode d'extraction de collocations (en allemand, en français et en roumain) à partir, notamment, de corpus alignés. En se focalisant sur les constructions verbe + nom, les chercheurs présentent une méthode d'extraction originale qui, en plus d'effectuer une première analyse de textes basée sur la statistique, tient compte, au moment du filtrage, des propriétés morphosyntaxiques des collocations telles qu'elles sont définies en grammaire systémique fonctionnelle.

La construction (semi-)automatique de ressources linguistiques multilingues à partir de corpus est présente également dans l'article de B. Jacquemin et S. Ploux, qui adoptent, toutefois, une perspective essentiellement sémantique par contraste avec des perspectives plus distributionnelles. Les auteurs proposent un modèle de navigation dans lequel les utilisateurs peuvent avoir accès aux différents sens des mots et à leur réalisation en discours, et ce, dans des langues différentes. Le modèle s'appuie sur le formalisme éprouvé des *Atlas sémantiques* qui construit des graphes représentant les aires sémantiques occupées par chacun des sens que peut recouvrir une forme lexicale.

Finalement, l'article de Chr. Roche aborde l'épineux problème de la gestion ontologique des contenus multilingues et de ses liens avec la terminologie. À la question de la traduction interlinguistique vient s'ajouter la problématique du passage de la langue naturelle au langage de représentation formel, passage que l'auteur propose d'aborder à partir d'une démarche qu'il nomme « ontoterminologie ».

La lecture des contributions que nous avons réunies dans ce numéro spécial offrira, nous l'espérons, de nombreuses pistes de réflexion, mais également des solutions concrètes à des problèmes soulevés par les pratiques actuelles en lexicographie, à savoir la gestion de données multilingues, l'acquisition de données linguistiques à partir de corpus, la représentation dictionnaire de propriétés linguistiques complexes rattachées aux unités lexicales ou aux termes spécialisés.