

Meta

Les corpus et la recherche en terminologie et en traductologie

Marc Van Campenhoudt et Rita Temmerman

Les corpus et la recherche en terminologie et en traductologie

Volume 56, numéro 2, juin 2011

URI : id.erudit.org/iderudit/1006173ar

DOI : [10.7202/1006173ar](https://doi.org/10.7202/1006173ar)

[Aller au sommaire du numéro](#)

Éditeur(s)

Les Presses de l'Université de Montréal

ISSN 0026-0452 (imprimé)
1492-1421 (numérique)

[Découvrir la revue](#)

Citer cet article

Van Campenhoudt, M. & Temmerman, R. (2011). Les corpus et la recherche en terminologie et en traductologie. *Meta*, 56(2), 223–225. doi:10.7202/1006173ar

Tous droits réservés © Les Presses de l'Université de Montréal, 2011

Ce document est protégé par la loi sur le droit d'auteur. L'utilisation des services d'Érudit (y compris la reproduction) est assujettie à sa politique d'utilisation que vous pouvez consulter en ligne. [<https://apropos.erudit.org/fr/usagers/politique-dutilisation/>]

érudit

Cet article est diffusé et préservé par Érudit.

Érudit est un consortium interuniversitaire sans but lucratif composé de l'Université de Montréal, l'Université Laval et l'Université du Québec à Montréal. Il a pour mission la promotion et la valorisation de la recherche. www.erudit.org

Les corpus et la recherche en terminologie et en traductologie

MARC VAN CAMPENHOUDT

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

RITA TEMMERMAN

Centrum voor Vaktaal en Communicatie, Toegepaste Taalkunde,
Erasmushogeschool Brussel, Bruxelles, Belgique
rita.temmerman@ehb.be.org

L'expression *linguistique de corpus* ou *corpus linguistics* semble relativement tardive au regard des liens profonds qui unissent texte et informatique. Elle semble émerger vers la fin des années 1970, sans que le lien avec la linguistique computationnelle soit nécessairement perçu comme un préalable. L'observation de l'usage réel, de la *parole*, est un fait qui lui est largement antérieur et les promoteurs d'un projet de l'ampleur du *Trésor de la langue française* ou, plus tard, ceux du *Cobuild*, n'ont pas attendu l'émergence d'un concept à la mode pour voir dans l'ordinateur un outil au service d'une linguistique descriptive. La linguistique de corpus, au sens où nous l'entendons de nos jours, est pourtant demeurée longtemps un espace confiné, fréquenté par les spécialistes de l'ingénierie linguistique et de la lexicométrie ou par des chercheurs travaillant dans le cadre de grands laboratoires, tel l'Institut national de la langue française à Nancy. Dans un univers demeuré trop longtemps – ou trop souvent – parallèle, celui de la traductologie ou des *translation studies*, l'ambition de mieux appréhender les caractéristiques linguistiques du travail du traducteur a très tôt conduit à exploiter également la puissance des ordinateurs pour aligner des textes traduits ou comparer l'écrit du traducteur et celui du rédacteur. Mais ici aussi, les difficultés d'accès à ce type d'outil ont longtemps tenu nombre de chercheurs à l'écart d'une approche systématique.

La révolution de la micro-informatique a progressivement bouleversé l'univers de la recherche, rendant accessibles des outils jusqu'alors très coûteux et bien peu ergonomiques. Dès la fin des années 1980, des logiciels comme *Micro-OCP (Oxford Concordance Program)* permettaient au chercheur individuel de travailler sur ses propres corpus sans devoir maîtriser un langage informatique. À la même époque, le monde de la traduction professionnelle commençait déjà à découvrir les algorithmes d'alignement de corpus développés par les scientifiques et à mettre au point des « mémoires de traduction », désormais utilisées pour le meilleur comme pour le pire.

Il faudra toutefois attendre le développement des autoroutes de l'information pour que l'accès facilité à d'énormes masses textuelles de qualité rende incontournables les méthodes d'une linguistique de corpus désormais indissociable de l'outil logiciel. Dans le même temps surgissait un écueil majeur : le risque de littéralement « se goinfrer » d'une nébuleuse de textes à la traçabilité incertaine, en faisant fi de tous les impératifs heuristiques d'une démarche scientifique bien pensée.

Dans un mouvement de retour en arrière, il convient de se rappeler que dans le champ des études théoriques, l'usage des corpus a été longtemps mis sur la touche au nom de préceptes bien ancrés au sein d'écoles théoriques jusqu'alors dominantes, telle celle de Noam Chomsky. De même, en terminologie, les fondements théoriques de l'« École de Vienne », tels que véhiculés par une simplification scolaire des travaux d'Eugen Wüster, de même que les pratiques normatives de l'aménagement linguistique ont conduit à largement négliger la réalité de l'usage et ses fondements sociologiques. Les recherches des vingt dernières années sur l'extraction de candidats-termes et un certain nombre d'études doctorales novatrices s'appuyant sur l'observation de faits de langue au sein de vastes corpus spécialisés ont mené à une large révision de l'approche dite « conceptuelle ».

Une page est désormais tournée et l'intérêt fondamental des concordanciers, des aligneurs et autres outils de balisage est progressivement pris en compte dans les cursus universitaires, s'agissant d'initier à une réflexion de fond sur la langue, le dire et le traduire. Dans les départements de traduction, une tendance « professionnalisante » lourde voudrait toutefois que l'on diplôme des traducteurs ou des terminologues aptes à utiliser tel ou tel logiciel commercial, appelé pourtant à évoluer rapidement, voire à disparaître au gré des modes. Pourtant, les qualités de réflexion que suppose l'exercice de ces métiers impliquent que l'on forme avant tout l'étudiant à un usage critique des logiciels commerciaux, ce que permet précisément une initiation approfondie aux méthodes de la linguistique de corpus et à son épistémologie.

Aujourd'hui, le diplômé en traduction comme le diplômé en langues et littératures se doivent de connaître toutes les possibilités de recherche systématique et rigoureuse offertes par le « bras informatisé » de la linguistique descriptive, dont les instruments paraissent désormais incontournables. En effet, ils renouvellent totalement les méthodes classiques d'observation dans tous les champs disciplinaires concernés par l'usage de la langue, qu'il s'agisse de stylistique, d'analyse de discours, de lexicologie, de phraséologie, de sémiotique ou encore de cognition.

Ces réflexions, et bien d'autres encore, nous ont conduits à organiser en avril 2009 un séminaire intitulé « La linguistique de corpus au service de la recherche en terminologie et en traductologie ». Au cours de celui-ci, les différents intervenants ont voulu faire le point sur les nombreuses avancées rendues possibles par l'exploitation de corpus écrits ou oraux dans leurs champs de recherche respectifs. La qualité de cette rencontre a débouché sur l'idée de ce numéro thématique de la revue *Meta*. Les textes ici réunis sont le résultat d'un appel à contributions restreint subséquent et d'un ample travail de sélection anonyme pour lequel nous ne saurions assez remercier les évaluateurs bénévoles.

Les articles rassemblés dans ce numéro rendent compte de recherches originales. Ils concernent neuf langues différentes (anglais, arabe, espagnol, français, galicien, néerlandais, persan, portugais et italien) et se situent presque tous dans un cadre multilingue. Le « tropisme » bruxellois de l'événement initial explique la place impor-

tante de travaux consacrés au français et au néerlandais et la nature essentiellement européenne des contributeurs. Bien des types de corpus sont envisagés. Il peut s'agir de bitextes comme de corpus comparables, de corpus textuels classiques ou plus originaux, comme les placards publicitaires (Mathieu Guidère) ou les annuaires professionnels (Dardo de Vecchi). Leur contenu peut être de nature spécialisée – monodisciplinaire ou multidisciplinaire – ou littéraire (Gabriela Saldanha) ou encore relever de la grande presse (Laurent Nicaise). Des corpus mixtes sont également exploités, comme le très original *Dutch Parallel Corpus* (Lieve Macken, Orphée de Clercq et Hans Paulussen, Gudrun Vanderbauwhede). Les domaines couverts sont nombreux et variés : actualités, économie et commerce, exobiologie, finance, informatique, médecine, pêche, publicité pour les cosmétiques.

Les aspects épistémologiques prédominent dans plusieurs communications qui mettent en avant des questions de méthode de constitution et d'exploitation des corpus (Mathieu Guidère, Lieve Macken, Orphée de Clercq et Hans Paulussen et Dardo de Vecchi, notamment). La préoccupation traductologique marque l'ensemble des thématiques abordées : évaluation de la qualité (Ghodrat Hassani), processus cognitifs (Mathieu Guidère, Sabela Fernández-Silva et Koen Kerremans), stylistique (Laurent Nicaise, Gabriela Saldanha), syntaxe et combinatoire (Anne Bertels et Serge Verlinde, Gudrun Vanderbauwhede), relations sémantiques (Jeanne Dancette) ou choix du terme et de l'équivalence (Silvia Bernardini et Adriano Ferraresi, Anne Condamines et Nathalie Dehaut, Dardo de Vecchi, François Maniez). Repartir du texte, de l'acte de communication et de traduction effectué dans un cadre culturel précis, au sein d'une communauté linguistique particulière (Laurent Nicaise), conduit à ne plus travailler dans une perspective « isolationniste » pour privilégier un fait particulier sorti de son contexte. Le terme ne peut se laisser réduire à un patron morphologique (François Maniez). Il est désormais vu comme participant à une dynamique de la langue spécialisée toujours en état de questionnement (Anne Condamines et Nathalie Dehaut). Avec la linguistique de corpus, les infinies possibilités de rapprochement, de comparaison, de mise en relation interdisent désormais de « fabuler » sur le contenu des textes, la réalité de la langue et les mécanismes de traduction.