



**HAL**  
open science

# LINGUISTIQUE DE CORPUS APPLIQUÉE : INNOVATIONS MÉTHODOLOGIQUES, THÉORIQUES ET PRATIQUES

Julien Longhi

► **To cite this version:**

Julien Longhi. LINGUISTIQUE DE CORPUS APPLIQUÉE : INNOVATIONS MÉTHODOLOGIQUES, THÉORIQUES ET PRATIQUES. Études de linguistique appliquée : revue de didactologie des langues-cultures et de lexicultureologie, 2017. halshs-01836197

**HAL Id: halshs-01836197**

**<https://shs.hal.science/halshs-01836197>**

Submitted on 12 Jul 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## LINGUISTIQUE DE CORPUS APPLIQUÉE : INNOVATIONS MÉTHODOLOGIQUES, THÉORIQUES ET PRATIQUES

Le 16 février 2018, l'AFLA (Association française de linguistique appliquée) organisait une journée intitulée « Répondre aux besoins de la société par les sciences du langage ». Plusieurs communications, dont une à laquelle nous avons participé, ont rendu compte de recherches et de collaborations avec des institutions, entreprises, ou qui se positionnent sur des sujets d'actualité ou de controverse. La dimension appliquée de la linguistique peut parfois évoquer une dimension « concrète » qui s'opposerait à la linguistique théorique ou fondamentale. L'objectif de ce volume est à la fois de contribuer à montrer la non pertinence de cette distinction théorique/appliquée, mais aussi de valoriser l'intérêt théorique et fondamental de la linguistique appliquée à des corpus issus de contextes non académiques. Par exemple, dans la lignée des recherches de Danièle Dubois, plusieurs thèses en linguistique cognitive ou psycholinguistique ont été soutenues jusqu'aux années 2010, en lien étroit avec des entreprises :

– la thèse de Gaëlle Delepaut (2007), à la SNCF, « associe des théories linguistiques et psychologiques pour l'explicitation des liens entre langage et cognition, à partir du concept de confort en train ». Le travail est réalisé à partir de corpus spécifiques sollicités (questionnaires soumis auprès de voyageurs pour recueillir un corpus sur leur ressenti individuel concernant ce concept multisensoriel) et d'« analyses linguistiques, prenant en compte dans un mouvement récursif, le lexique, la syntaxe, la morphologie et la sémantique », avec des résultats sur l'implication des locuteurs dans leur discours, la structuration du concept de confort en train, l'identification des propriétés plus ou moins typiques du confort en train. Ces résultats émergent grâce à l'articulation entre langue et discours, et le rapport au monde réel qu'établissent « les signes linguistiques, les représentations cognitives et les mesures physiques » ;

– la thèse de Caroline Cance, s'inscrit dans le cadre d'une demande industrielle de PSA Peugeot Citroën : comment se construit le jugement perceptif visuel des usagers à propos d'habitacles automobiles? Elle propose « une contribution à l'étude de ces relations de co-construction entre langue, discours, perception dans une approche pluridisciplinaire, en linguistique et psychologie cognitives, à travers l'exploration d'un questionnement appliqué ».

Ces courts descriptifs des thèses, et leur lecture plus attentive, indiquent que le gain théorique de ces analyses de corpus originaux est important, non seulement au regard des observables qui sont en jeu, mais aussi au regard de la réflexion sur les concepts d'analyse, et l'articulation des niveaux de l'analyse linguistique. Ce qui change, c'est en outre la manière dont sont appréhendés les résultats de la recherche. Comme l'indiquent Condamines et Narcy-Combes (2015), « si, dans une perspective théorique, la validation est liée à l'accueil et l'acceptation par les pairs (« oui, ces résultats s'inscrivent comme une suite de travaux pré-existants et semblent valides »); dans une perspective appliquée, la validation est double : par les pairs et par les demandeurs originaux ». Cette double validation ne doit pas être conçue comme une « dégradation » de la scientificité des résultats, qui tiendrait compte de considérations non scientifiques, mais plutôt comme un point de vue complémentaire, généralement pensé en interdisciplinarité, de la production des résultats. Par exemple, l'intégration de résultats linguistiques à des « mesures » acoustiques, visuelles, sonores, etc., peut permettre de préciser ou d'affiner les critères de l'analyse linguistique des corpus en question. Aussi, « il paraît de moins en moins possible de séparer linguistique et linguistique appliquée », puisque « les problèmes

de validation théorique ou sociale, et de la complémentarité qui les accompagnent, nous confirment que ce qui les a séparées initialement était la relative indépendance du linguiste “théorique” qui n’avait pas de compte à rendre » (Condamines et Narcy-Combes, 2015).

Des exemples d’autres recherches, qui ont beaucoup apporté aux domaines de la sémantique de textes, ou à l’analyse du discours, montrent aussi la mise à profit des savoirs théoriques très spécifiques dans des contextes applicatifs différents :

– la thèse de Bénédicte Pincemin (1999)<sup>1</sup> s’est intéressée à l’ergonomie Web adaptée à l’affichage de textes. Par l’ancrage différentiel, des améliorations de parcours des réponses du système et leur appropriation sont rendus possibles : « Dans le contexte de documents écrits à dominante scientifique et technique, quatre facettes textuelles sont définies pour guider la conception des traitements : (i) la matière linguistique du texte ; (ii) son organisation interne, structurée, close et orientée ; (iii) l’intertextualité ; (iv) le rôle constitutif des lectures et la dynamique de l’interprétation. Ce modèle motive des innovations à toutes les étapes du traitement ».

– la thèse de Marion Mauger-Parat, menée dans le cadre d’une Convention Industrielle de Formation à la Recherche (CIFRE) auprès de l’entreprise EDF R&D, a cherché à comprendre « la façon dont circulent les représentations climatiques parmi différentes sphères sociales, partant des discours scientifiques vers les discours profanes, et passant par les discours médiatiques ».

Il existe bien entendu nombre d’autres travaux qui répondent à la demande sociale, notamment dans le cadre des recherches qui articulent langage et travail, autour des projets de Josiane Boutet par exemple.

Ces innovations sont également rendues possibles par l’« outillage » de l’analyse de corpus, qui permet de traiter des volumes de textes importants, sans pour autant perdre de vue la spécificité du matériau. Dans Garric et Longhi (2009), nous indiquons que « les ressources informatiques et les nouveaux corpus qu’elles offrent renouvellent, modifient et créent les pratiques » (p. 14), à la manière du « corpus réflexif » constitué par les historiens du discours (Guilhaumou par exemple).

Ainsi, sans que les contributions présentées dans ce volume s’attachent à traiter d’un point théorique spécifique, elles abordent toutes des concepts fondamentaux pour l’analyse linguistique de corpus, ou des domaines qui font l’objet d’approfondissements : représentations linguistiques, sémantique argumentative, prototypes, perception et sensorialité, textualité, actes de langage, expertise discursive, etc. Si le prisme initial est celui de la réponse linguistique à un questionnement social, politique, ou professionnel, les conclusions sont à la fois liées à l’objet d’étude, et aux dimensions théoriques, voire épistémologiques, des niveaux d’analyse et des notions mis en jeu. Elles contribuent aussi au mouvement d’analyse réflexive sur les corpus engagée dans Garric et Longhi (2012) qui vise à « déterminer comment faire évoluer les corpus et les pratiques dans un contexte où les données et les outils offrent de nouvelles modalités avec leurs sources de complexité, notamment des paramètres d’évolution et d’hétérogénéité, susceptibles de diminuer le contrôle des observables et des résultats de l’analyse » (p. 5), s’intégrant dans une démarche qui suggère d’introduire l’hétérogénéité jusque dans l’opération de constitution des corpus établis en vue d’une même problématique. En outre, les analyses sur de tels corpus ne sont pas sans poser des questions

---

<sup>1</sup> Cette thèse a bénéficié d’une Convention Industrielle de Formation à la Recherche auprès de l’entreprise EDF.

de légitimité et surtout d'éthique, en particulier lorsque les linguistes répondent à des commandes du terrain, parfois sans maîtriser les usages de la recherche menée. La dimension applicative doit donc s'accompagner d'une réflexion sur l'expertise du linguiste dans des domaines et contextes divers, qui importent avec eux des enjeux qui peuvent complexifier à la fois l'analyse du matériau langagier, mais aussi la portée possible des analyses effectuées.

Ce volume se compose de cinq contributions<sup>2</sup>, qui abordent des contextes applicatifs différents, et qui s'ancrent dans des approches parfois différentes. Les deux premiers articles, celui de Sandrine Graf, et celui de Laurène Renaut, Laura Ascone, et Julien Longhi, abordent la spécificité de l'intervention linguistique dans des contextes professionnels (domaine de la communication, ou domaine judiciaire), avant de proposer des réflexions d'ordre méthodologique (quelles données ? quels savoirs mobiliser ?), théorique (quels niveaux d'analyse ? quelles théories mobiliser ?) et applicatif (quels premiers résultats proposer à ces domaines ?). Ils proposent donc un regard à la fois applicatif sur un domaine, mais aussi réflexif, puisqu'ils se situent à des stades d'émergence de la linguistique aux domaines concernés, et s'interrogent sur les conditions de légitimation et d'application même de l'expertise linguistique académique dans ces contextes. Les trois articles suivants, celui d'Olivier Méric et Laurent Gautier, de Valérie Rochaix, et de Matthieu Bach et Laurent Gautier, se situent tous les trois dans le cadre de projets déjà bien avancés (thèses d'O. Méric soutenue en 2016, projet ANR sur le patrimoine, projets européens et régionaux sur le vin), et permettent d'asseoir, avec le recul des résultats apportés, la pertinence du recours à l'analyse de corpus pour les objets étudiés, et le gain analytique et théorique qu'il y a à convoquer ces nouveaux corpus (recueil de paroles en situation spécifique, démarche empirique). L'ensemble de ces contributions témoigne finalement de l'exigence que requiert l'analyse de corpus appliquée, dans des contextes d'interface avec la demande sociale, industrielle, professionnelle, à la fois du point de vue des résultats, mais aussi de la rigueur scientifique pour appuyer et accompagner ces résultats, et pour en expliciter l'obtention.

*Julien LONGHI*

*Université de Cergy-Pontoise*

## BIBLIOGRAPHIE

BOMMIER-PINCEMIN, B. 1999. Diffusion ciblée automatique d'informations : conception et mise en œuvre d'une linguistique textuelle pour la caractérisation des destinataires et des documents. Thèse de doctorat en Université Paris IV (Sorbonne).

CANCE, C. 2008. Expériences de la couleur, ressources linguistiques et processus discursifs dans la construction d'un espace visuel : l'habitacle automobile. Thèse de doctorat, sous la direction de Danièle Dubois, Université de la Sorbonne nouvelle - Paris III.

CONDAMINES, A. et NARCY-COMBES, J.-P. 2015. « La linguistique appliquée comme science située », dans F. Carton, J.-P. Nancy-Combes, M.-F. Nancy-Combes, D. Toffoli, Cultures de recherche en linguistique appliquée. Riveneuve éditions.

---

<sup>2</sup> Nous remercions Olga Galatanu, Nathalie Garric, Alain Rabatel, et Albin Wagener, pour l'expertise apportée aux contributions de ce volume.

DELEPAUT, G. 2007. Contribution de la linguistique cognitive à l'identification du confort : analyse des discours des passagers sur le confort en train. Thèse de doctorat, sous la direction de Danièle Dubois, Université de la Sorbonne nouvelle - Paris III.

GARRIC, N. et LONGHI, J. 2009. « Présentation », L'analyse linguistique de corpus discursifs. Des théories aux pratiques, des pratiques aux théories. Cahiers du Laboratoire de Recherche sur le Langage. Clermont-Ferrand, n° 3, Presses universitaires Blaise-Pascal.

GARRIC, N. et LONGHI, J. 2012. « L'analyse de corpus face à l'hétérogénéité des données : d'une difficulté méthodologique à une nécessité épistémologique », Langages, n° 187.

MAUGER-PARAT, M., 2013. Les représentations climatiques dans la presse française : analyses sémiologiques et sémiotiques comparées de discours médiatiques, scientifiques et profanes. Thèse de doctorat, sous la direction de JeanDidier Urbain, Université René Descartes - Paris V.