



HAL
open science

Les conditions sociales de la constitution d'un corpus : l'impact d'une dynamique de co-construction sur l'homogénéité des données.

Athéna Dupont

► To cite this version:

Athéna Dupont. Les conditions sociales de la constitution d'un corpus : l'impact d'une dynamique de co-construction sur l'homogénéité des données.. 8es JOurnées jeunes chercheurs de la LInguistique de COrpus (JOLICO 2015), Sep 2015, Orléans, France. hal-01358039

HAL Id: hal-01358039

<https://hal.science/hal-01358039>

Submitted on 30 Aug 2016

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.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0
International License

Les conditions sociales de la constitution d'un corpus : **l'impact d'une dynamique de co-construction sur l'homogénéité des données**

Athéna Dupont¹

¹Laboratoire Ligérien de Linguistique - Université d'Orléans

Pour l'étude du français contemporain parlé, l'Enquête Sociolinguistique à Orléans (ESLO) fait référence tant par son volume (700 heures) que par la variété des situations d'enregistrement et des personnes dont la parole est recueillie (Baude & Dugua, 2011). Au nombre des « modules » construits pour appréhender la diversité des usages langagiers, le sous-corpus *ESLO 108 / Labomedia*, en cours de constitution, procède d'une démarche ethnographique fondée sur l'hypothèse qu'une association comme Labomedia serait un lieu d'émergence et un vecteur du changement linguistique. Par ses activités orientées sur les pratiques du numérique, elle réunit des intervenants jeunes et qualifiés, disposant d'un capital culturel familial consolidé par un parcours post-bac.

Lieu d'échanges qui valorise l'expression et le partage des connaissances, Labomedia n'est pas organisée comme une entreprise ou une administration mais elle entretient un lien actif avec les instances assurant la transmission des standards linguistiques – structures éducatives et médias (Bourdieu, 2001). L'enquête telle qu'elle a été conçue prescrit une temporalité qui n'est pas équivalente au temps de recueil des données, redéfinissant les modalités de la relation aux acteurs. Sur ce point, l'approche retenue pour la constitution de la ressource se situe à l'opposé du micro caché, du micro-trottoir, du micro partagé et, en partie, des techniques d'entretien qui cherchent à représenter la trajectoire d'un individu, en disposant a priori de peu, ou d'aucune information.

Les propriétés du terrain et des agents requièrent une prise en compte du lien social dans le mode de constitution des données (Bourdieu, 1993). La similarité des trajectoires du chercheur et de ses interlocuteurs, les demandes d'information sur les connaissances techniques fait du témoin un partenaire. L'intérêt porté par les membres de Labomedia pour les premiers enregistrements réalisés a fait naître des projets collaboratifs et a modifié l'entretien qui, d'élément d'archive était réapproprié en objet numérique intégré à des pratiques d'exploitation non commerciales. Cette évidence du terrain se traduit en termes de co-construction d'un corpus et non plus de captation et de restitution et modifie sa structure.

Le processus d'objectivation (Beaud & Weber, 2003), en particulier la mise à distance du caractère interpersonnel de la relation, s'est fait en intégrant au corpus des documents issus de productions personnelles, comme les notes de terrain, et de supports rédigés ou non par les membres de Labomedia – newsletters, comptes rendus, communiqués et articles provenant du wiki, du site web ou des listes de diffusion de l'association. Ce sont autant d'indices pour

appréhender la façon dont Labomedia se présente et se représente (Gumperz, 1989), et pour identifier les consommations culturelles de ses adhérents et les réseaux d'échange. Cette méthode revient à concevoir le corpus comme un mode d'organisation des données langagières, à l'intersection de la parole et des habitus, à définir le corpus oral comme un ensemble de données structurées (Habert, 2009) à partir des sources orales recueillies en co-construction.

Le travail en cours tente de questionner le point de vue positiviste qui fait du corpus une collection de données sans s'interroger sur son mode de production. En observant l'incidence sociale d'une construction qui doit prendre en compte les spécificités du terrain, on fait l'hypothèse que le corpus reflète aussi bien des attentes individuelles et collectives que les valeurs du groupe observé. Parce qu'elles participent de l'organisation et du traitement des données, le corpus recèle en lui une part du projet et de la méthodologie qui a prélué à son recueil et lui donne certaines de ses propriétés. Pour cette raison, on se propose de montrer en quoi les conditions d'enquête sont inhérentes à une réflexion critique sur les conditions de validité des corpus, en particulier ceux constitués dans un milieu qui a les moyens symboliques de contrôler son image et les éléments qu'il livre à l'enquête.

Bibliographie

BAUDE, O. & DUGUA, C. (2011). "(Re)faire le corpus d'Orléans quarante ans après : quoi de neuf, linguiste ?", *Corpus*, 10 : 99-118.

BAUDE, O. (2006). *Corpus Oraux. Guide des bonnes pratiques 2006*. Paris, CNRS Editions – Orléans, PUO.

BEAUD, S. & WEBER, F. (2003). *Guide de l'enquête de terrain : produire et analyser des données ethnographiques*. Paris, La Découverte.

BOURDIEU, P. (1993). "Comprendre", *La misère du monde*. Paris, Point-Seuil, 903-939.

BOURDIEU, P. (2001). *Langage et pouvoir symbolique*. Paris, Point-Seuil.

GUMPERZ, J. J. (1989). *Engager la conversation : introduction à la sociolinguistique interactionnelle*. Paris, Éditions de Minuit.

HABERT, B. (2005). *Instruments et ressources électroniques pour le français*. Paris, Éditions Ophrys.

HABERT, B. (2009). *Construire des bases de données pour le français : Notions* (Vol. 1). Paris, Éditions Ophrys.

LALLEMENT, M. (2015). *L'Âge du faire : Hacking, travail, anarchie*. Paris, Point-Seuil.