Lidil

Revue de linguistique et de didactique des langues

31 | 2005

Corpus oraux et diversité des approches

Présentation

Marie Savelli

Édition électronique

URL : <http://journals.openedition.org/lidil/131>

DOI : 10.4000/lidil.131

ISSN : 1960-6052

Éditeur

UGA Éditions/Université Grenoble Alpes

Édition imprimée

Date de publication : 1 juin 2005

Pagination : 5-11

ISBN : 2-914176-12-0

ISSN : 1146-6480

Référence électronique

Marie Savelli, « Présentation », *Lidil* [En ligne], 31 | 2005, mis en ligne le 22 août 2007, consulté le 24 septembre 2020. URL : <http://journals.openedition.org/lidil/131> ; DOI : <https://doi.org/10.4000/lidil.131>

Ce document a été généré automatiquement le 24 septembre 2020.

© Lidil

Présentation

Marie Savelli

- 1 La constitution de bases de données est d'actualité dans les pays de la Communauté européenne et s'intéresser au langage parlé est important et urgent pour le développement de la linguistique. Cependant, s'intéresser à l'oral reste difficile pour plusieurs raisons dont l'une des principales est la difficulté reconnue de constituer des corpus voire des sous-corpus représentatifs et de taille suffisante pour servir de support à des exploitations linguistiques. Bien que le support informatique ait considérablement facilité les choses, la constitution d'un recueil de données reste une opération très coûteuse en temps et les techniques de constitution – comme de traitement des données – doivent être sans cesse affinées pour mener à bien les diverses analyses. Nul n'est besoin de rappeler combien les grands corpus oraux, qui comptabilisent plusieurs millions de mots, manquent en France. Le retard abyssal dans ce domaine a maintes fois été dénoncé dans des articles, des revues (Rafla, 1999) ou plus récemment dans les objectifs de projets des équipes de recherches (projet dans le contrat Plan État région Rhône-Alpes, en cours). Le manque de corpus oraux disponibles et informatisés entraîne surtout une insuffisance dans les analyses et la connaissance de notre propre langue. Ce retard commence néanmoins à être comblé et l'objectif de ce numéro est de présenter le travail des équipes qui constituent actuellement des recueils de données orales ou qui se servent de banques de données existantes pour mener à bien divers types d'analyses. Ainsi faisant, ils contribuent à faire évoluer les recherches dans différents domaines de la linguistique. Deux aspects se dégagent : la valorisation des corpus, la prise en compte des données et les recherches auxquelles elles donnent lieu dans des domaines aussi variés que la psycholinguistique, l'analyse interactionnelle, la syntaxe, la didactique notamment.
- 2 Ce numéro de la revue *Lidil* est centré sur l'oral et les corpus oraux adultes et enfants. Il propose un panorama qui va de la constitution des corpus – avec des apports théoriques divers – aux analyses qu'ils sous-tendent, lesquelles nous permettent de découvrir la variété des approches.
- 3 Lukas Balthasar et Michel Bert présentent la Base CLAPI, en cours d'élaboration et qui peut être considérée comme la base centralisée de consultation de données orales pour

l'interaction en France. En cours de construction à l'ENS Lettres et Sciences humaines de Lyon, elle est l'exemple même d'une synergie réussie et hautement performante entre linguistes et informaticiens. L'intérêt de cette base de données est multiple : actualisation de la réflexion théorique à mener autour des conventions de transcription ; sur les problèmes déontologiques et juridiques liés notamment à la diffusion des données. Le lecteur trouvera dans cet article une très abondante bibliographie et des liens internet de consultation des banques de données orales actuellement disponibles sur le web.

- 4 Lorenza Mondada et Véronique Traverso assument un objectif double : travailler sur des *données orales interactionnelles* – des corpus oraux de parole en interaction recueillis dans des contextes ordinaires non provoqués par les chercheurs –, et adopter une *perspective* qui accorde une place fondamentale à l'organisation interactionnelle de cette parole. Elles se penchent sur un phénomène interactionnel bien identifié dans les corpus analysés, et « qui concerne la réalisation collective de la clôture et l'ouverture de la phase suivante notamment dans les interactions professionnelles ou institutionnelles... ». Les observations menées sont d'autant plus intéressantes qu'elles touchent un ensemble de dimensions fondamentales de l'organisation de l'interaction menée par les locuteurs. Elles portent sur le caractère reconnaissable du découpage de l'activité en phases distinctes et des moyens mis en œuvre pour délimiter les découpages phrastiques ; sur les modalités et les ressources (verbales et visuelles) qu'utilisent les interactants pour organiser leur arrivée conjointe à certains points structurants de l'activité et notamment à la fin d'une séquence. L'article pose la question de la généralisation de telles observations et de ce qu'elles peuvent nous apprendre sur le contexte des interactions.
- 5 Dominique Bassano exploite des corpus oraux d'usages enfantins avec les conventions CHILDES aménagées. On pourra juger du travail minutieux mené en amont pour le recueil des données mais qui donne la possibilité de fournir des observations longitudinales scientifiques fort originales et intéressantes. Les conclusions auxquelles parvient l'auteure sont surprenantes et à l'encontre de notre intuition. Dans la première partie de son article, elle présente d'abord l'étude de la production naturelle précoce – analyse de corpus de productions spontanées de très jeunes enfants (de 1 à 3-4 ans) –, puis, son programme de recherches et la base de données sur l'émergence du langage et la constitution du système linguistique chez les enfants français. La troisième partie illustre ce programme avec un ensemble d'analyses portant sur le développement des noms, qu'elle étudie conjointement du point de vue lexicosémantique et grammatical.
- 6 Florence Chenu et Harriet Jisa présentent aussi des recherches menées à partir des corpus oraux d'usages enfantins mais dans d'autres perspectives. L'impact des caractéristiques spécifiques aux langues dans l'input maternel a été l'objet de récentes et nombreuses recherches mais elles prolongent ce champ d'études et émettent trois hypothèses concernant l'acquisition des verbes chez 2 enfants monolingues français (24-29 mois), enregistrés tous les 15 jours en interaction avec leur mère (vidéo). Les données sont transcrites selon les conventions CHILDES, puis codées pour l'exploration de ces trois hypothèses (lemmatisation, codage syntaxique). La première hypothèse pose que la fréquence des verbes (types et tokens) dans l'input maternel influence les verbes observés dans la production de l'enfant. Tenant compte des travaux antérieurs, l'étude longitudinale proposée examine la relation entre l'usage maternel et

l'« explosion verbale » dans la production de l'enfant. Toutefois, les spécificités du français amènent les auteures à considérer deux hypothèses supplémentaires. La deuxième hypothèse pose que la position (initiale ou finale) du verbe dans la construction facilite pour l'enfant l'extraction de la forme. Les référents nouveaux en français parlé sont typiquement introduits par un item lexical en position post-verbale puis repris par des clitiques avant le verbe dans les propositions suivantes. Ainsi, en français parlé, le verbe se trouverait-il souvent dans une position finale saillante. L'article examine dans quelles mesures les verbes dans l'input maternel sont en position finale et si ces verbes font partie des premiers verbes dans la production de l'enfant. La troisième hypothèse postule que la diversité des usages syntaxiques dans l'input facilite l'élaboration des caractéristiques syntaxiques et sémantiques des verbes. Bien que l'ordre des mots pour le français soit caractérisé comme SVO, le français parlé présente une variation considérable (pronominalisations, dislocations). L'article pose la question de savoir s'il existe une relation entre la diversité d'ordre des mots dans le discours maternel et l'acquisition des verbes par l'enfant.

- 7 Mary Annick Morel développe son article sur l'examen d'une demi-minute de dialogue à bâtons rompus enregistré en audio et vidéo, et présente plusieurs niveaux d'analyse (morphosyntaxique, intonatif et mimico-gestuel) tant du côté du parleur (de la paroleuse) que du côté de celui (celle) qui écoute et réagit nécessairement, en produisant des sons, des mimiques et des gestes. Elle met en lumière les indices nouveaux (rôle des mains et du regard) qui interagissent et qu'il est indispensable de prendre en compte si l'on veut analyser la progression du dialogue. Sachant que l'intonation a pour rôle premier d'assurer la cohésion des groupes syntaxiques et des constituants du paragraphe oral, les différents paramètres de l'intonation assument chacun un rôle spécifique dans le marquage des positions interpersonnelles et dans la gestion de la formulation. Par exemple, elle montre comment l'intensité de la voix sert à manifester le droit à la parole de chacun ; comment les variations de la mélodie traduisent les anticipations du parleur à l'égard de la réception de ses propos et des réactions de celui auquel il s'adresse.
- 8 Chantal Lyche revient sur le rôle et la fiabilité des données orales et sur les choix qui sont souvent opérés indépendamment des faits sur lesquels s'appuient les études. Elle jette un regard critique sur les données subjectives, notamment en phonologie revenant sur le parler de Dell inexorablement retenu comme seul support à toutes les études sur le schwa. En contrepoint, elle expose l'un des objectifs du PFC, à savoir l'élaboration d'une base de données sur le français oral à partir d'une méthodologie commune, ce qui va favoriser des études comparatives sérieuses et une mutualisation des données. Il s'agit d'une base entièrement alignée et partiellement codée qui constituera à terme une source inestimable de données à haute fiabilité, qui devrait permettre de tester sur plusieurs points d'enquêtes la variation phonologique régionale. Dans l'attente de son achèvement, l'auteure nous livre les premiers résultats d'une étude qui porte sur les liaisons. On verra que même partiels, ces résultats sont fort intéressants et pourraient amener à revoir, en particulier, certains classements sur les liaisons obligatoires.
- 9 Robert Bouchard s'appuie sur des corpus didactiques qui présentent des caractéristiques très fortes : cadre institutionnel contraignant, définition précise des rôles, communication inégale, phénomènes polylogaux, événement communicatif long et planifié, langue sur objectif spécifique... Les corpus ont été recueillis avec un projet,

non seulement descriptif mais praxéologique : ils ont souvent pour but la modification de l'interaction « ordinaire » observée ou au contraire l'(in)validation d'« interactions innovantes ». Il est nécessaire notamment de prendre en compte l'ensemble des participants de ces leçons, l'enseignant et les élèves dans leur individualité. L'auteur développe la problématique du polylogue encore assez peu prise en compte dans les travaux de type interactionnel. Après avoir tracé un historique rapide du polylogue en général, il décrit les manifestations polylogales les plus importantes de l'échange pédagogique pour se consacrer ensuite aux manifestations polylogales et polylogiques propres aux prises de parole de l'enseignant

- 10 Paul Cappeau revient sur l'attention à porter à la constitution même des recueils de données orales. Il montre comment le travail à partir de données orales puisées dans de grands corpus a permis d'affiner, d'améliorer, de corriger les analyses qui portent sur la syntaxe du français. Il prend appui sur les corpus du GARS et un corpus réalisé à Poitiers sur les hommes politiques pour observer notamment les réalisations du sujet, les interrogatifs et les indéfinis pour mieux cerner ce que les réalisations orales peuvent apporter : elles peuvent par exemple enrichir la liste des formes à observer (ainsi *il y en a qui* doit prendre place aux côtés de *certain*, *plusieurs*, etc.), elles peuvent aussi aider à mieux comprendre comment se répartissent les emplois (ainsi on voit mieux dans les productions orales dans quels contextes on trouve l'élément *certain* anaphorique et *certain* nominal). Loin de plaider en faveur d'une syntaxe propre à l'oral, l'auteur tente de montrer que les observations recueillies conduisent à considérer un panorama bien plus riche au plan quantitatif et qualitatif (au-delà des seules productions écrites) et amène à mieux appréhender ce qu'est la syntaxe du français. Sa démonstration confirme que ce ne sont pas les travaux sur l'oral qui modifient la syntaxe mais en revanche, que les grands corpus permettent d'aborder des chapitres nouveaux en syntaxe.
- 11 Maria Caterina Manes Gallo et Marie Savelli tentent d'approcher l'interface syntaxe/sémantique dans un contexte particulier : la demande d'itinéraire piéton, en milieu urbain. Les deux corpus oraux analysés ont été recueillis dans deux situations de production différentes ce qui permet d'ébaucher des comparaisons. Il s'agit d'une part, d'un corpus recueilli *in situ* et d'autre part, d'un corpus répondant à une consigne expérimentale en laboratoire. Les données servent d'appui à deux types d'analyses complémentaires et à partir de supports théoriques différents. À la visualisation syntaxique *stricto sensu*, dégagée du modèle de Claire Blanche-Benveniste, a été superposée une seconde étude visant à affiner la fonction des constructions verbales prédicatives au niveau sémantique. Des tentatives d'enrichissement sémantique des grilles syntaxiques ont déjà été faites, mais jamais en termes de « type de procès ». D'où la proposition d'une évolution des paradigmes dégagés avec intégration d'une typologie des énoncés élémentaires fondée sur une catégorisation des types de procès. L'article suggère enfin une comparaison entre les deux types de recueil de données et leur incidence sur les discours produits.
- 12 Je souhaiterais remercier chaleureusement l'ensemble des auteur(e)s qui ont permis de présenter le vivier actuel de leurs recherches sur les corpus oraux et l'oral, même si la taille de la revue ne leur a pas permis d'être aussi explicites qu'ils le souhaitaient.

AUTEUR

MARIE SAVELLI

Université Stendhal-Grenoble 3 / Laboratoire Lidilem.