

Corpus et pathologies du langage : du recueil à l'analyse de données pour une linguistique clinique et appliquée

Corpus and language disorders: from data collection to analyses for clinical and applied linguistics

Christine Da Silva Genest et Caroline Masson



Édition électronique

URL : <http://journals.openedition.org/corpus/4374>

ISSN : 1765-3126

Éditeur

Bases ; corpus et langage - UMR 6039

Référence électronique

Christine Da Silva Genest et Caroline Masson, « Corpus et pathologies du langage : du recueil à l'analyse de données pour une linguistique clinique et appliquée », *Corpus* [En ligne], 19 | 2019, mis en ligne le 01 janvier 2019, consulté le 01 janvier 2019. URL : <http://journals.openedition.org/corpus/4374>

Ce document a été généré automatiquement le 1 janvier 2019.

© Tous droits réservés

Corpus et pathologies du langage : du recueil à l'analyse de données pour une linguistique clinique et appliquée

Corpus and language disorders: from data collection to analyses for clinical and applied linguistics

Christine Da Silva Genest et Caroline Masson

- 1 Ce numéro thématique a pour objectif de réunir un ensemble de contributions s'attachant à rendre compte de la constitution et de l'utilisation de corpus de données orales ou écrites dans le champ des pathologies du langage. Plus précisément, notre ambition est de montrer ce que les données issues de situations réelles peuvent apporter aux connaissances actuelles dans ce domaine. Nous considérons, à l'instar de (de Weck & Rodi 2005 ; de Weck & Marro 2010 ; MacLeod *et al.* 2011 ; MacWhinney *et al.* 2011 ; Delamotte 2012), que la recherche fondamentale sur l'acquisition et/ou les pathologies du langage ne peut totalement s'abstraire de la réalité des pratiques langagières effectives que ce soit dans le cadre familial, éducatif ou clinique (en orthophonie notamment). Ne pas se confronter aux données *in situ*, c'est se priver d'une source féconde pour la réflexion théorique et méthodologique. De ce fait, avoir accès et travailler sur des données recueillies en situation de production réelle présente un intérêt principal, celui de passer d'une approche déductive à une approche empirique, fondées sur des données attestées et des méthodes d'analyse spécifiques. Ces types de corpus contribuent ainsi à une meilleure connaissance du fonctionnement langagier et des développements typique et atypique, à l'amélioration des prises en charge, orthophoniques par exemple, et à la mise en œuvre d'actions professionnelles.

Décrire et évaluer le fonctionnement linguistique atypique

- 2 Dans le cadre des recherches en pathologies du langage, la manière d'appréhender les compétences langagières des locuteurs peut être placée sur un continuum (pour une revue voir de Weck & Marro 2010) allant d'épreuves contrôlées dans des situations induisant des formes et des structures verbales ciblées au recueil de données en situation la plus écologique possible (Canut & Vertalier 2008 ; Boulton *et al.* 2013 ; Morgenstern 2016). Ces approches renvoient à des visions à la fois différentes et complémentaires du fonctionnement de la langue : dans le premier cas, on cherche à évaluer la production et l'absence de production (ou la production déviante) d'un phénomène linguistique de manière décontextualisée tandis que dans le second, l'objectif est d'apprécier la façon dont le locuteur mobilise toutes ses compétences en contexte. Or, les recherches actuelles montrent que les productions linguistiques de tout locuteur (enfant ou adulte) ne peuvent pas être considérées indépendamment de leur ancrage discursif. En effet, les compétences langagières sont mobilisées différemment par les locuteurs selon les activités en cours et les genres de discours associés (Kern & Chenu 2010), au niveau, par exemple, des expressions référentielles (Salazar Orvig *et al.* 2018), des types spécifiques d'interventions verbales (da Silva *et al.* 2012), ou encore des connecteurs utilisés dans les narrations (de Weck & Rosat 2003). Les outils à développer doivent donc considérer la diversité des situations langagières et la façon dont les enfants s'adaptent à celles-ci.
- 3 La constitution de corpus offre aux chercheurs de nouvelles possibilités pour l'étude et la description des conduites langagières typiques et atypiques, quels que soient la période de la vie (l'enfance, l'adolescence, l'âge adulte ou la vieillesse) ou les modes de communication (l'oral ou l'écrit). En effet, le recueil permet d'obtenir des ressources sur les productions langagières dans des situations variées et de les utiliser à des fins d'analyses linguistiques. Celles-ci éclairent le chercheur, par exemple, sur les types de stratégies mises en œuvre par les locuteurs en situation ou sur l'évolution de compétences qui n'était pas observable dans les résultats obtenus à un test de langage (da Silva 2015). Ainsi, en ayant recours à une diversité de situations de production (provoquées, semi-spontanées, spontanées), le chercheur ou le clinicien peut appréhender le profil linguistique des locuteurs.
- 4 Le recours aux corpus peut donc être tout à fait approprié pour amener les professionnels à comprendre le fonctionnement langagier et interactionnel (da Silva Genest & Masson 2017). En effet, l'observation et l'analyse des données leur permet de mieux cerner les caractéristiques mêmes du langage oral, bien souvent méconnues et dévalorisées par la référence à la norme de l'écrit (da Silva Genest 2018), et également celles du langage écrit.

Analyser les pratiques langagières des professionnels et leurs effets

- 5 À la suite d'études sur corpus décrivant des processus atypiques (Le Normand 2007 ; Parisse & Maillart 2007 ; Royle & Thordadottir 2008), le champ des recherches s'est récemment élargi en s'intéressant aux situations cliniques de manière parallèle aux situations d'interactions familiales (Chalivet & de Gaulmyn 1998 ; Rodi 2011 ; da Silva

Genest 2017 ; da Silva Genest & Masson 2017 ; de Weck *et al.* 2017). Ceci a pour effet d'approfondir les connaissances des professionnels de santé à la fois sur leurs propres pratiques et sur la façon dont leurs patients réagissent à celles-ci.

- 6 L'utilisation de données recueillies *in situ* et la description de ces situations cliniques peuvent donc également servir la formation et la réflexion des professionnels en les conduisant à prendre conscience des modalités d'interaction proposées et à déterminer celles qui sont les plus efficaces pour aider les sujets à pallier leurs difficultés langagières. Dans cette perspective, le recours au corpus peut servir à construire avec les professionnels une démarche d'analyse des pratiques s'il est envisagé comme un moyen de mesurer l'adéquation des pratiques avec les objectifs visés. Cette démarche d'identification des procédés d'étayage et d'ajustement de l'adulte aux capacités du patient a pour but de modifier les représentations et les pratiques langagières des professionnels en vue d'instaurer une dynamique d'apprentissage dans le processus dialogique (Canut *et al.* 2018). Cela suppose de tenir compte des différents paramètres en jeu, à savoir non seulement le sujet et ses capacités cognitives et langagières, mais aussi les productions langagières du professionnel. Celles-ci vont constituer le matériau à partir duquel le patient va « travailler » pour s'approprier le langage et vont permettre de déterminer le degré d'adaptation de l'adulte aux potentiels de développement de chacun.
- 7 Il apparaît donc que la constitution de données authentiques, l'élaboration d'une méthodologie et la création d'outils d'analyses spécifiques sont nécessaires pour le développement des connaissances dans ce domaine (Morgenstern 2016).

Diffuser les corpus

- 8 Parallèlement à la constitution de données, leur diffusion s'avère tout aussi importante. Dans le cadre de la linguistique de corpus, la valorisation et la diffusion de ressources sont des questions abordées par les chercheurs depuis de longue date, en particulier la mise à disposition de données constituées par eux-mêmes, par exemple pour les études sur le français parlé (Blanche-Benveniste 1997 ; Debaisieux *et al.* 2016), la didactique des langues (Boulton 2009), les pathologies du langage (Goodwin 2000 ; Ghio *et al.* 2006 ; MacWhinney *et al.* 2010, 2011) ou l'acquisition du langage (Morgenstern & Parisse 2007). Un autre exemple est la construction collaborative de CHILDES, base de données internationale et interlangues d'interactions naturelles ou semi-guidées (MacWhinney 2000), qui a permis de mettre en relation des chercheurs de divers horizons sur des corpus unifiés. En France, le développement d'ORTOLANG (Pierrel *et al.* 2016) permet à présent aux chercheurs qui le souhaitent de déposer sur la plateforme des ressources issues de leurs travaux en vue de les diffuser et de les partager avec d'autres chercheurs. Il est ainsi devenu possible depuis quelques années d'avoir accès à des corpus constitués dans le cadre de projets de recherche. Néanmoins, cela reste limité, notamment en ce qui concerne les données de productions d'enfants francophones à développement atypique, et ne concerne qu'un nombre très limité de pathologies (troubles du spectre autistique ou troubles spécifiques du développement du langage) : seules les transcriptions de trois corpus sont accessibles (Le Normand 1997 ; Foudon *et al.* 2007 ; Bang & Nadig 2015) et les enregistrements audio ne sont pas tous disponibles. Or, rendre accessible ces corpus d'interactions pourrait contribuer à la mise en place de recherches portant sur le fonctionnement langagier atypique ou pathologique en proposant aux chercheurs, professionnels de santé et étudiants de travailler sur des données issues de différents

protocoles de recherche et permettant l'accès à des données de productions induites comme des données plus écologiques, recueillies dans un cadre scientifique et totalement annotées.

Quels apports de la linguistique de corpus pour l'étude des pathologies du langage ?

- 9 Les contributions du numéro s'attacheront donc à montrer ce que la linguistique de corpus, appliquée au champ de la linguistique clinique, apporte comme pistes de réflexion sur les plans scientifiques et professionnels ou celui de la formation (initiale ou continue). Ainsi, en considérant ces objectifs, deux grandes problématiques lient les articles du numéro.
- 10 Premièrement, celle des données en elles-mêmes : quels types de données doivent être collectés pour développer les connaissances fondamentales ou apprécier les compétences linguistiques et communicationnelles des locuteurs ? Quels sont les effets des contextes ou des tâches à effectuer sur les productions ? Quelle est la pertinence de considérer différents contextes de production ?
- 11 L'article d'Élodie Clayette et Naomi Yamaguchi relève ainsi la pertinence de proposer différentes tâches langagières (induites ou non, portant sur des mots ou sur des énoncés) pour l'analyse phonologique de la parole de sujets aphasiques. Brigitte Stanké, Marine Le Mené, Stefano Rezzonico, André Moreau, Christian Dumais, Julie Robidoux, Camille Dault et Phaedra Royle se sont intéressés quant à eux aux données développementales sur l'orthographe lexicale pour construire et développer une base de données élaborée à partir de différents supports (manuels, ouvrages de littérature, etc.) pour la communauté des chercheurs comme des professionnels. Enfin, Agnès Witko et Florence Chenu, par la présentation d'un protocole méthodologique de recueil et d'analyse des productions écrites d'adolescents présentant une dyslexie-dysorthographe, montrent l'intérêt de les analyser pour décrire les opérations sous-jacentes à l'acte d'écriture tout en dégageant les critères objectifs pertinents pour l'évaluation et la prise en charge orthophonique.
- 12 Deuxièmement, ce sont les méthodes d'analyse et la mise en évidence de critères objectifs qui sont ciblées. Les recherches portant sur des données naturelles et tout particulièrement les interactions verbales sont encore peu nombreuses. Cela est lié, non seulement, à la difficulté d'analyser des données provenant de cadres non contraignants (Parisse & Morgenstern 2010) mais aussi aux variations et à l'hétérogénéité discursive (Liégeois 2012). Sans fondement et à tort, les analyses sont souvent considérées comme basées sur de simples observations et non sur des critères précis et objectifs. Or, les contributions à ce numéro montrent bien que l'élaboration des catégories d'analyse et leur utilisation sont possibles. Elles sont le résultat d'études à la fois qualitatives et quantitatives et d'un constant va-et-vient entre théorie et corpus.
- 13 Dans cette perspective, Geneviève de Weck et Anne Salazar Orvig présentent une grille d'analyse de l'étayage linguistique et cognitif et la démarche employée dans la construction de celle-ci. Cette grille permet d'observer et de qualifier finement les conduites étayantes des adultes et les réactions des enfants à celles-ci dans des situations d'interactions familiales et de prises en charge orthophoniques. Thi Mai Tran, Maité Boyé, Sandrine Mejias et Natalia Grabar ont pour objectif de dégager de nouveaux critères cliniques fonctionnels et pertinents à partir de l'analyse de situations d'échanges

conversationnels dans le cadre des troubles démentiels de type Alzheimer. Elles montrent ainsi comment appréhender les conduites langagières des sujets atteints de cette pathologie en situations naturelles pour établir le diagnostic et la prise en charge orthophonique ultérieure. Le développement de méthodes d'analyse des compétences des sujets âgés est aussi mis en avant par Guillaume Duboisdindien, Cyril Grandin, Dominique Boutet et Anne Lacheret-Dujour qui proposent une démarche longitudinale et multimodale d'identification des marqueurs pragmatiques verbaux et non verbaux dans le discours de sujets présentant des troubles cognitifs légers. Il s'agit d'identifier les profils pragmatiques et communicationnels des sujets et de mettre en évidence l'importance de considérer toutes les modalités sémiotiques de la communication.

- 14 L'importance de travailler sur corpus pour la recherche fondamentale comme pour la pratique orthophonique est également l'objectif de la dernière contribution de Stéphanie Caët. Elle y présente une expérience pédagogique menée au département d'orthophonie de l'Université de Lille consistant à sensibiliser les futurs professionnels de santé à la méthodologie de travail sur corpus. Elle y initie les étudiants au recueil et à l'analyse de corpus dans le cadre de leur formation initiale en orthophonie : apprendre à récolter des données, à manipuler les outils, à prendre conscience des choix qu'ils seront amenés à prendre et les effets sur les productions langagières de leurs patients.
- 15 L'ensemble de ces contributions mettent toutes en évidence la nécessité de s'intéresser aux données recueillies dans diverses situations pour comprendre le fonctionnement langagier typique ou atypique. Cette approche permet de mieux appréhender les pathologies et de développer les recherches sur le sujet. En outre, alors qu'on a tendance à les opposer ou à les déconnecter, on voit bien la nécessité d'un va-et-vient constant entre recherche et clinique et la façon dont ces deux plans se complètent. Des perspectives de recherche et d'application des résultats se présentent donc favorablement pour les prochaines années.

BIBLIOGRAPHIE

- Bang J. & Nadig A. (2015). « Learning language in autism : Maternal linguistic input contributes to later vocabulary », *Autism Research* 8(2) : 214-223.
- Blanche-Benveniste C. (1997). *Approches de la langue parlée en français*. Paris : Ophrys.
- Boulton A. (2009). « Documents authentiques, oral, corpus », *Mélanges CRAPEL* 31 : 5-13.
- Boulton A., Canut E., Guerin E., Parrisé C. & Tyne H. (2013). « Corpus et appropriation de L1 et L2 », *Linx* 68-69 : 9-32.
- Canut E. & Vertalier M. (2008). « Des données représentatives... De quoi en acquisition du langage ? Constitution de données à observer et objectifs d'analyse », *Verbum* 30(4) : 299-312.
- Canut E., Masson C. & Leroy-Collombel M. (2018). *Accompagner l'enfant dans son apprentissage du langage. De la recherche en acquisition à l'intervention des professionnels*. Paris : Hachette.

- Chalivet A. & de Gaulmyn M.M. (1998). « Speech-therapy for elderly people : Construction of coherency », *Pragmatics* 8(2) : 203-219.
- Da Silva C., Rezzonico S., Corlateanu C., Gendre S. & Ingold J. (2012). « Influence de la situation dans des interactions mère-enfant dysphasique et mère-enfant tout-venant », *CORELA - RJC Cotexte, contexte, situation* [En ligne] publié le 7 mai 2012 : <http://corela.edel.uni-poitiers.fr/index.php?id=2349>.
- Da Silva C. (2015). « Compétences pragmatico-discursives des enfants dysphasiques en situation de jeu symbolique : évaluation et évolution », *Colloque « Narration et interaction »*, Université Paris Descartes, 10-12 juin 2015, Paris.
- Da Silva Genest C. (2017). « Influence des activités sur les pratiques langagières en orthophonie/logopédie », *Travaux Neuchâtelois de Linguistique* 66 : 157-173.
- Da Silva Genest C. (2018). *Le Grevisse de l'Orthophoniste*. Louvain-la-Neuve : De Boeck Supérieur.
- Da Silva Genest C. & Masson C. (2017). « Apport de la linguistique de corpus à l'étude des situations cliniques : utilisation de ressources écologiques pour évaluer les pratiques professionnelles », *Studii de lingvistică* 7 : 89-112.
- Debaisieux J.M., Benzitoun C. & Delofeu H.J. (2016). « Le projet ORFEO : Un corpus d'études pour le français contemporain », *Corpus* 15 : 91-114.
- Delamotte R. (2012). « Narration enfantine : la diversité des récits et de leur perception », in R. Delamotte & M.A. Akinci (éd.) *Récits d'enfants : Développement, genre, contexte*. Rouen : Presses Universitaires de Rouen et du Havre, 119-144.
- De Weck G. Sublon A. & Fox G. (2017). « The use of personal storytelling in speech and language therapist-adolescent Interactions », in S. Pekarek Doehler, A. Bangerter, G. de Weck, L. Filliettaz, E. Gonzalez-Martinez & C. Petitjean (éd.) *Interactional competences in institutional settings. From school to the workplace*. Palcrave : Macmillan, 59-88.
- De Weck G. & Marro P. (2010). *Les troubles du langage chez l'enfant. Description et évaluation*. Issy-les-Moulineaux : Elsevier Masson.
- De Weck G. & Rodi M. (2005). « Évaluation des capacités pragmatiques et discursives », in B. Piérart (dir.) *Le langage de l'enfant. Comment l'évaluer ?*. Louvain-la-Neuve : De Boeck Supérieur.
- De Weck G. & Rosat M.C. (2003). *Troubles dysphasiques : comment raconter, relater, faire agir à l'âge préscolaire*. Paris : Elsevier Masson.
- Foudon N., Reboul A. & Manificat S. (2007). « Language acquisition in autistic children : a longitudinal study ». *CamLing2007 : University of Cambridge Postgraduate Conference in Language Research*.
- Ghio A., Teston B., Viallet F., Jankowski L., Purson A. et al. (2006). « Corpus de parole pathologique, état d'avancement et enjeux méthodologiques ». *Travaux Interdisciplinaires du Laboratoire Parole et Langage d'Aix-en-Provence (TIPA)* 25 : 109-126.
- Goodwin C. (2000). « Gesture, aphasia, and interaction », in D. McNeill (éd.), *Language and Gesture*. Cambridge : Cambridge University Press, 84-98.
- Kern S. & Chenu F. (2010). « Contextes de production et LAE en français : caractéristiques pragmatiques et syntaxiques ». *Actes du CMLF 2010*, 1527-1537.
- Le Normand M.T. (1997). « Early morphological development in French children », in A.S. Olofsson & S. Strömqvist (éd.) *Cross-linguistic studies of dyslexia and early language development*. Luxembourg : Office for Official Publication of the European Communities, 59-79.

- Le Normand M.T. (2007). « Évaluation de la production spontanée du langage oral et de l'activité sémantique du récit chez l'enfant d'âge préscolaire », *Rééducation Orthophonique* 231 : 53-72.
- Liégeois L. (2012). « De l'analyse au partage des données, quel (s) format (s) choisir ? L'exemple d'un corpus d'interactions parents-enfant », *COLDOC 2012 : Traitement de corpus linguistiques*, 128-142.
- MacLeod A.A.N., Sutton A. Trudeau N. & Thordardottir E. (2011). « The acquisition of consonants in Québécois French : A cross-sectional study of pre-school aged children », *International Journal of Speech-Language Pathology* 13(2) : 93-109.
- MacWhinney B. (2000). *The CHILDES Project : Tools for Analyzing Talk. 3rd Edition*. Mahwah, NJ : Lawrence Erlbaum Associates.
- MacWhinney B., Fromm D., Forbes M. & Holland A. (2011). « AphasiaBank : Methods for studying discourse », *Aphasiology* 25/11 : 1286-1307.
- Morgenstern A. (2016). « Pratiques langagières et comportements du patient en milieu familial : apport des méthodes ethnographiques multimodales pour la recherche en médecine », *Ethics, Medicine and Public Health* 2 : 641-649.
- Morgenstern A. & Parisse C. (2007). « Codage et interprétation du langage spontané d'enfants de 1 à 3 ans », *Corpus* 6: 55-78.
- Parisse C. & Maillart C. (2007). « Phonology and syntax in French children with SLI : A longitudinal study », *Clinical Linguistics & Phonetics* 21(11-12) : 945-951.
- Parisse C. & Morgenstern A. (2010). « Transcrire et analyser les corpus d'interactions adulte-enfant », in E. Veneziano, A. Salazar Orvig & J. Bernicot, *Acquisition du langage et interaction*. Paris : L'Harmattan, 201-222.
- Pierrel J.M., Parisse C., Blanchard J., Petitjean E. & Pierre F. (2016). « ORTOLANG : a French Infrastructure for Open Resources and TOols for LANGuage » *Linköping Electronic Conference Proceedings*, Aix-en-Provence, 26-28 October 2016, CLARIN Common Language Resources and Technology Infrastructure 136: 102-112.
- Rodi M. (2011). « Interactions logopédiste-enfant : Comment se construisent des échanges potentiellement acquisitionnels dans une situation de jeu symbolique ? », *Glossa* 110 : 49-68.
- Royle P. & Thordardottir E. T. (2008). « Elicitation of the passé composé in French preschoolers with and without specific language impairment », *Applied Psycholinguistics* 29(3): 341-365.
- Salazar Orvig A., Marcos H., Heurdier J. & da Silva C. (2018). « Referential features, speech genres and activity types », in M. Hickmann, E. Veneziano & H. Jisa (éd.) *Sources of variation in first language acquisition : Languages, contexts, and learners*. Trends in Language Acquisition Research (TiLAR). Amsterdam/Philadelphia : John Benjamins Publishing Company, 219-242.

AUTEURS

CHRISTINE DA SILVA GENEST

Université de Lorraine – EA 3450 DevAH & UMR 7118 ATILF

CAROLINE MASSON

Université Sorbonne Nouvelle – Paris 3 - EA Clesthia 7345 & UMR 7118 ATILF