Former les étudiants en orthophonie à la constitution de corpus : pourquoi ? Comment ?

Training speech and language therapy students to corpus building: why? How?

Stéphanie Caët

Édition électronique

URL : <http://journals.openedition.org/corpus/4258>

DOI : 10.4000/corpus.4258

ISSN : 1765-3126

Éditeur

Bases ; corpus et langage - UMR 6039

Référence électronique

Stéphanie Caët, « Former les étudiants en orthophonie à la constitution de corpus : pourquoi ? Comment ? », *Corpus* [En ligne], 19 | 2019, mis en ligne le 01 janvier 2019, consulté le 08 septembre 2020. URL : <http://journals.openedition.org/corpus/4258> ; DOI : <https://doi.org/10.4000/corpus.4258>

Ce document a été généré automatiquement le 8 septembre 2020.

© Tous droits réservés

Former les étudiants en orthophonie à la constitution de corpus : pourquoi ? Comment ?

Training speech and language therapy students to corpus building: why? How?

Stéphanie Caët

Introduction

- 1 Dans cet article, nous proposons de nous pencher sur les enjeux d'une formation des étudiants en orthophonie à la constitution de corpus (enregistrement et transcription). Nous suggérons que cette formation peut notamment permettre de :
 - Leur faire prendre conscience de la multitude des choix qui sont posés par celui, chercheur et/ou praticien, qui étudie les productions langagières, et de l'impact de ces choix sur les données et les analyses.
 - Les former à une observation attentive et les préparer à une analyse réflexive de productions langagières attestées.
 - Développer une réflexion sur leurs propres pratiques langagières et professionnelles futures.
 - Leur faire manipuler des outils qui pourront leur être utiles dans le cadre de leur activité professionnelle.
 - Les sensibiliser aux règles de bonnes pratiques concernant la collecte, l'exploitation et l'éventuelle diffusion de données.
 - Les initier à une méthodologie de recherche particulière afin qu'ils puissent mener une activité qui, au même titre que l'évaluation et la prise en soin, relève pleinement des activités de l'orthophoniste : l'étude scientifique du langage et de ses troubles (Comité Permanent de Liaison des Orthophonistes 1994).
- 2 Nous nous centrerons sur la constitution de corpus oral ou multimodal, pouvant amener des questions spécifiques liées aux techniques de recueil, au caractère personnel voire sensible des données collectées en contexte de pathologie, à la

transcription nécessairement subjective de ces données, et sur le lien avec la pratique clinique.

- 3 Nous illustrerons les points soulevés par des exemples d'application pédagogique mise en œuvre dans le Département d'Orthophonie de Lille en 2017-2018. La méthode pédagogique adoptée évoluera sans aucun doute encore sensiblement dans les prochaines années, grâce au dialogue continu avec les praticiens, d'autres enseignants qui pratiquent avec leurs étudiants la constitution et l'analyse de corpus en orthophonie ou en sciences du langage ainsi que les étudiants de 5^e année en orthophonie constituant un corpus pour leur mémoire de fin d'études.
- 4 L'approche pédagogique que nous adoptons est une approche inductive d'une part (Michel 2002), et « par compétences » d'autre part (Chauvigné & Coulet 2010). C'est en effet à partir des situations proposées et des actions qui sont à réaliser (enregistrer, transcrire et analyser des données) que les étudiants sont amenés à se questionner et à chercher des réponses à ces questions et des solutions aux problèmes rencontrés. Ce faisant, ils appliquent et renforcent les connaissances théoriques acquises précédemment et acquièrent des connaissances nouvelles et des compétences à la fois spécifiques à la discipline et généralisables à d'autres enseignements théoriques ou cliniques, à des situations rencontrées en stage, à leur mémoire de fin d'études, et, nous l'espérons, à leur pratique future. Du fait de cette approche inductive, les étapes du travail qui est proposé aux étudiants ne reflètent pas le processus de la recherche en linguistique de corpus (cf. Atkin, Cleans & Ostler 1992). Ainsi, quand ils collectent leurs données et quand ils commencent à les transcrire, les étudiants ne savent souvent pas encore ce qu'ils vont analyser alors que dans le processus de la recherche, même si l'approche est « corpus driven », la collecte et la transcription des données sont réalisées suivant une approche théorique et/ou clinique spécifique.
- 5 Nous pensons que la constitution de corpus peut favoriser un investissement personnel actif et interactif par l'étudiant de l'enseignement et des apprentissages. Nous tentons de rendre cela possible en leur proposant un recueil de données personnel, une transcription en binôme (les étudiants créent donc de toutes pièces le support de leurs apprentissages), une liberté de choix quant au domaine ou au phénomène langagier étudié en fonction de leurs centres d'intérêts et des données disponibles dans leurs corpus, des interactions entre étudiants et avec l'enseignant prévues sur les temps d'enseignement et grâce aux outils de l'environnement numérique de travail.
- 6 À la fin de cet article, nous revenons sur des inquiétudes exprimées par les étudiants eux-mêmes concernant cette pratique de constitution de corpus oral, comme autant d'appels adressés par ces futurs professionnels aux chercheurs et/ou aux institutions pour que soient développés des outils et des cadres d'activité professionnelle facilitant la participation des orthophonistes à la recherche sur les pathologies du langage, leur évaluation, leur prise en charge et l'évaluation de cette prise en charge.

1. Corpus et formation des futurs orthophonistes

- 7 De par leur objet commun, les sciences du langage et l'orthophonie entretiennent des relations ténues. Parce que les praticiens sont amenés à traiter les données langagières recueillies auprès de leurs patients, Witko (2010) présente la relation entre orthophonie et linguistique de corpus comme une relation « naturelle » (§ 16). Elle suggère que l'apport de la linguistique de corpus à l'orthophonie est double, puisqu'elle offre des

outils théoriques et méthodologiques pour le traitement des données recueillies par les professionnels dans le cadre de leur pratique d'une part et pour la recherche sur les pathologies du langage et leur prise en soin d'autre part (cf. aussi Canut, Leroy-Collombel & Masson 2017 ; Da Silva et Masson 2017).

- 8 C'est donc sans doute tout aussi naturellement que la notion de corpus a été convoquée dans le nouveau référentiel de formation publié en 2013, dans le descriptif d'unités d'enseignements (UE) à la fois théoriques (à visée applicative toutefois : UE Développement du langage et psycholinguistique et UE Connaissances fondamentales en sciences du langage appliquées à l'orthophonie – semestre 3) et cliniques (UE Bilan et évaluation en orthophonie – semestre 5). L'usage du terme est associé aux étapes de recueil des données (spontanées, en complément des données recueillies grâce aux outils standardisés), de transcription (en API) et d'analyse (au niveau pragmatique). On remarque que le mot « corpus » n'apparaît pas dans le descriptif des UE cliniques sur l'intervention orthophonique ni dans l'UE sur l'évaluation des pratiques professionnelles, deux lieux où l'analyse de corpus pourrait être pratiquée (Canut *et al.* 2017 ; Da Silva & Masson 2017).

Exemple – Corpus et formation des étudiants à Lille

À Lille, les enseignements de l'UE Développement du langage et psycholinguistique sont consacrés à l'acquisition de connaissances théoriques dans ce domaine. Dix heures portent sur le développement des productions langagières des enfants et l'interaction. Dans ce cadre, les étudiants découvrent des études menées sur corpus et en mesurent les apports et les limites (Demuth 2008). De nombreuses illustrations sont tirées de corpus librement accessibles, comme le corpus CoLaJE (Morgenstern et Parisse 2012). C'est dans le cadre de l'UE Connaissances fondamentales en sciences du langage appliquées à l'orthophonie que les étudiants transcrivent et analysent des productions orales et multimodales. Dans cette UE, la plus grande part des enseignements reviennent sur les notions acquises lors des 2 UE de sciences du langage de 1^{re} année de manière appliquée sur corpus de productions élicitées, spontanées ou semi-spontanées, typiques ou pathologiques, d'adultes ou d'enfants, en monologue ou en interaction (dont des interactions patient-orthophoniste). Parallèlement, les étudiants sont invités à travailler sur des enregistrements audio ou vidéo qu'ils ont eux-mêmes réalisés lors de leurs stages de 1^{re} année en école maternelle et primaire, à les transcrire et à les analyser. Ce sont les réflexions sur ce travail parallèle que nous présentons ici, en nous focalisant sur les étapes de recueil et de transcription.

2. Recueil de productions orales ou multimodales

- 9 Le recueil de données orales ou multimodales nécessite de se poser des questions relevant à la fois de son objectif scientifique ou clinique et du cadre législatif dans lequel il est mené.

2.1 Questions de législation

- 10 Avant de pouvoir recueillir des données orales ou multimodales, il est nécessaire 1) d'informer le sujet quant au cadre dans lequel s'inscrit le recueil, aux modalités d'enregistrement, aux objectifs scientifiques ou cliniques et aux usages qui seront faits des données (anonymisation, exploitation, conservation, protection, diffusion) et 2) d'obtenir son consentement ainsi éclairé (cf. Baude *et al.* 2006). Initier les étudiants en orthophonie à la pratique de recueil de données orales ou multimodales dès le début

de leur cursus est l'occasion d'une première sensibilisation à ces « bonnes pratiques », car ce sont des questions qu'ils seront amenés à rencontrer à d'autres reprises au cours de leur formation (s'ils enregistrent un patient pour une étude de cas clinique dans le cadre de travaux dirigés, ou lors de la réalisation de leur mémoire de fin d'études) et s'ils enregistrent leurs patients au cours de leur pratique clinique et ce, même si la collecte de données dans la formation et la pratique orthophonique est encadrée par le secret médical.

- 11 L'enregistrement, préconisé par Tran et Bécavin (2001) et présenté par Witko (2010 : § 24) comme une activité « naturelle », « fréquente » voire « incontournable » du professionnel, semble en effet être largement pratiqué par les orthophonistes : dans son mémoire de fin d'études, Reynaud-Michaelides (2016) observe que 70 % des 424 orthophonistes ayant répondu à un questionnaire quant à leurs pratiques en matière de recueil, de transcription et d'analyse de corpus, enregistrent les productions de leurs patients. L'enregistrement leur permet notamment de réécouter les productions du patient au moment de la rédaction du compte-rendu de bilan orthophonique ou de revenir avec le patient et éventuellement son entourage sur ses/leurs productions. Il permet également de garder une trace des productions spontanées en complément des productions élicitées. Pour Witko (2010), une observation longitudinale des productions du patient (forme, structure, usage) et de la dynamique de l'interaction est ainsi possible, contribuant à l'orientation des axes de remédiation. L'enregistrement pourra également permettre au praticien de mener un travail de réflexion sur ses propres pratiques langagières (cf. de Weck 2003 ; Rodi 2011 ; da Silva 2010) et de contribuer à l'étude scientifique des troubles et des prises en charge à partir de corpus en pathologie.
- 12 Ces questions d'ordre législatif ne sont pas sans conséquence sur les observations cliniques ou scientifiques qui peuvent être menées. Comme le rappellent Baude *et al.* (2006) par exemple, informer le sujet cible quant aux objectifs de l'enregistrement peut ôter à la situation le caractère naturel que l'on souhaitait observer, ce qui n'est pas sans questionner les étudiants au moment de l'analyse de leur corpus, un questionnement qu'ils seront de nouveau amenés à avoir lors de leur pratique professionnelle. En effet, l'information du patient peut avoir un impact d'autant plus important sur les données que l'insécurité linguistique du patient peut être grande du fait que l'observation porte sur l'objet même de ses difficultés (cf. Tran et Bécavin 2001). Aussi, comme le souligne Witko (2010), l'instauration d'un climat de confiance entre l'orthophoniste et le patient est essentielle. Enfin, l'outil peut être introduit par le praticien auprès du patient comme un outil professionnel « ordinaire » voire un « participant à l'interaction » (Morgenstern 2016), ce qui se fera d'autant plus simplement que l'enregistrement est pratiqué de manière régulière (da Silva Genest & Masson 2017).

2.2. Questions de pratiques scientifiques et cliniques

- 13 Si l'enregistrement constitue la première étape technique de la constitution de corpus, il résulte d'un ensemble de choix posés en amont et de considérations théoriques qui les sous-tendent : pourquoi enregistrer ? Qui enregistrer ? Dans quelle situation ? Combien de temps ? Vais-je réaliser un enregistrement audio ou vidéo ? Quel outil utiliser ? Est-ce que je serai présent au moment de l'enregistrement ?

Exemple – Modalités de recueil des données

À Lille, une même situation d'enregistrement a été déterminée pour l'ensemble des

étudiants pour pouvoir ensuite mutualiser les réflexions. C'est donc au moment de l'analyse des données que les étudiants vont pouvoir mesurer l'implication des choix précédemment posés pour eux, comprendre pourquoi ces éléments sont généralement déterminés en amont et en fonction des objectifs scientifiques (cf. Sinclair 1996 cité par Habert *et al.* 1997) et/ou cliniques (da Silva Genest & Masson 2017) et que toutes les étapes constitutives de l'étude de corpus sont liées les unes aux autres (cf. Aldolphs & Knight 2010).

Ils sont invités à réaliser un enregistrement audio ou vidéo (suivant l'autorisation parentale reçue) d'un enfant de grande section de maternelle parlant d'un dessin qu'il vient de réaliser. L'enregistrement d'un enfant est intéressant à plusieurs égards. Au moment de l'enregistrement, certains vont expérimenter la difficulté que peut parfois rencontrer un enfant à prendre la parole devant un micro ou une caméra (cf. Blanche Benveniste & Pallaud 2001). Au moment de la transcription, ils vont s'interroger sur la transcription de productions non standard (cf. section 3.2 ci-après). Au moment de l'analyse, ils vont questionner l'impact de leur présence et de celle de l'outil d'enregistrement sur les productions de l'enfant et réfléchir ensemble aux moyens de palier ce « paradoxe de l'observateur » (cf. Labov 1972) dans le cadre de la pratique clinique. Aucun autre critère n'est donné quant à la sélection des sujets. Certains choisissent toutefois d'enregistrer des enfants qui selon eux parlent « très bien » ou au contraire des enfants qui leur semblent rencontrer des difficultés. Dans un cas comme dans l'autre, les étudiants pourront être surpris au moment de la transcription et de l'analyse de découvrir ce que l'enfant qui ne parle « pas bien » peut faire et ce qui chez l'enfant qui parle « bien » est encore en cours d'acquisition (cf. section 3.1). Nous invitons les étudiants à interagir avec l'enfant au cours de l'enregistrement. Il s'agit donc d'interactions auxquelles eux-mêmes prennent part (situation proche de celles qu'ils rencontreront avec les patients), ce qui leur permettra de mener une réflexion sur leurs propres productions, au niveau structurel, fonctionnel et interactionnel. La situation dyadique permet en outre de ne pas les mettre en difficulté pour les étapes suivantes de transcription et d'analyse. Le choix d'enregistrer l'enfant parlant d'un dessin est déterminé par la volonté de mutualiser la collecte des données et de croiser les observations que les étudiants pourront également faire dans le cadre de l'UE sur le développement du graphisme. L'intérêt que cela représente a cependant des limites pour l'analyse linguistique lorsque les étudiants n'ont par exemple pas pu réaliser un enregistrement vidéo : l'usage des déictiques renvoyant à ce qui est représenté rend parfois difficile l'analyse des productions. C'est un point sur lequel il est intéressant de revenir avec les étudiants en leur rappelant les réflexions préalablement menées sur les avantages et les limites de différents types de recueil (cf. Behrens [2008] sur les méthodes utilisées dans les études sur l'acquisition du langage). La présence d'un support à l'interaction que l'enfant a lui-même créé facilite toutefois sa prise de parole dans ce contexte, et l'expérience menée reste une introduction au recueil et au traitement de productions spontanées que Rondal (2003) appelle de ses vœux pour l'évaluation orthophonique. Au moment de l'analyse, nous insisterons sur le fait qu'il s'agit d'une situation particulière et que seul un fragment des compétences de l'enfant est observable. Nous les sensibilisons dans le même temps à l'importance d'une formulation non généralisante lors de la mise par écrit des analyses. Nous ne transmettons pas aux étudiants de consigne exacte à donner à l'enfant. Ceci pourra les amener à s'interroger sur l'influence de la consigne donnée à l'enfant sur le type de discours produit (Blanche Benveniste & Pallaud 2001). Nous demandons aux étudiants de réaliser un enregistrement de 3 minutes minimum. Bien que le temps ne soit pas une mesure très pertinente en linguistique de corpus (cf. Baude *et al.* 2006 : 28), cette durée est un cadre qui les rassure, au moment de l'enregistrement comme au moment de la transcription. Cependant, si le débit de l'enfant est lent, elle pourra à l'inverse les déstabiliser au moment de l'analyse. Ce sera l'occasion de leur montrer qu'en fonction du domaine linguistique ou du phénomène analysé, la

densité du corpus peut être importante (cf. Tomasello & Stahl 2004) mais qu'il peut y avoir beaucoup à observer même si l'enfant parle peu, et que des analyses qualitatives riches des productions individuelles et de l'interaction peuvent être menées (Debras 2018 ; Canut & Vertalier 2008). Ces cas rejoignent des situations de bilan qu'ils seront amenés à expérimenter notamment avec des patients qui parlent peu ou pas. Les étudiants sont invités à enregistrer les productions de l'enfant avec leur téléphone portable car ils ont l'habitude de l'utiliser pour réaliser des films et les enfants sont de plus en plus habitués à être filmés de cette façon par leurs familles. L'usage de cet outil permet donc d'éviter de part et d'autre une éventuelle inquiétude induite par la technique (cf. Habert 2004) et laisse entrevoir aux étudiants la possibilité de l'utiliser lorsqu'ils seront praticiens. Après que l'enregistrement a été réalisé, on pourra questionner les étudiants sur d'autres aspects qu'ils ont eux-mêmes, de façon plus ou moins consciente, déterminés, comme le point de départ de l'enregistrement (ont-ils enregistré la consigne qu'ils ont donnée à l'enfant ?) ou ce qu'ils ont fait entrer dans le champ s'ils ont réalisé un enregistrement vidéo (l'enfant uniquement ? l'enfant et son dessin ? Ont-ils cherché une prise de vue leur permettant d'être dans le champ également ?). On pourra alors par exemple attirer leur attention sur l'importance d'avoir accès au contexte et aux productions non verbales de l'interlocuteur pour pouvoir interpréter certaines productions du sujet cible. En vue de la rédaction de leurs analyses, nous leur demanderons de critiquer la méthodologie utilisée et de réfléchir à une méthodologie alternative.

3. La transcription

- 14 Une fois l'enregistrement réalisé, la transcription est une étape centrale de la constitution de corpus pour la formation des étudiants en orthophonie.

Exemple – Transcrire séparément puis mettre en commun

Parce que le travail de transcription peut être long et parce qu'il peut mener à des questionnements intéressants à partager, nous invitons les étudiants à travailler en binôme. Après avoir écouté/visionné chacun de leurs enregistrements, nous leur demandons de sélectionner celui qui leur paraît le plus intéressant et de le transcrire séparément. Les étudiants mettent ensuite en regard leurs deux transcriptions. À ce moment-là, ils découvrent par eux-mêmes 1) qu'ils n'ont pas toujours entendu ou vu les mêmes choses (cf. Hriba, Baude & Dugua 2017 ; Cappeau 2008), ce qui va les conduire à prêter une attention renouvelée aux données collectées et à s'interroger sur la subjectivité de la perception auditive, leurs représentations du langage, de la langue et sur leurs attentes vis-à-vis du sujet et 2) qu'ils n'ont pas toujours représenté de la même façon les mêmes observables, ce qui va les amener à s'interroger sur les choix qui sont posés dans l'acte de transcription (Ochs 1979). Nous développons ces points dans les sections 3.1 et 3.2 ci-dessous.

3.1. Questionner ses représentations

- 15 L'écoute ou le visionnage multiple que permet et nécessite la transcription va faire apparaître des éléments auxquels les étudiants n'avaient peut-être d'abord pas prêté attention : un geste qui complète un énoncé verbal, une intervention verbale ou non verbale de l'interlocuteur. Ceci les conduit à s'interroger sur leurs représentations du langage et notamment sur la place du non-verbal (cf. McNeil 2000).
- 16 En transcrivant, les étudiants s'interrogent également sur les attentes qu'ils avaient vis-à-vis du sujet (cf. Gadet 1996 ; Blanche Benveniste 2010). Ainsi, ceux qui avaient choisi d'enregistrer un enfant qui selon eux ne parlait pas très bien vont parfois être

étonnés de voir que certes, au niveau phonologique, l'enfant présentait des spécificités qui pouvaient troubler l'intelligibilité, mais que ses énoncés présentent tout de même un certain degré de complexité syntaxique. Prendre conscience de ces attentes ne manquera pas de nourrir leur réflexion sur l'écoute dans la formation et la pratique orthophonique, une écoute souvent « orientée vers la recherche d'indices en faveur d'une éventuelle pathologie langagière » (cf. Tran & Bécavin 2001 : 31). Grâce à la transcription et à l'analyse de corpus, ces attentes et cette écoute « orientée » vont pouvoir être confrontées à un ensemble d'observables qui pourront être mesurés et situés les uns par rapport aux autres au sein du système du sujet cible, donnant ainsi plus de « crédibilité » à l'analyse qualitative (Witko 2010 : § 18).

- 17 La transcription va en outre les conduire à questionner leurs représentations de la langue orale. Si les corpus sont utiles pour la formation d'adultes en langues étrangères (cf. Chambers 2009), ils présentent également un intérêt non négligeable pour la formation des étudiants francophones au français parlé. Bien que natifs de leur langue, ils ont, comme beaucoup de locuteurs, une représentation de leur langue principalement basée sur l'écrit et sur des formes d'écrits relativement normées (écrits scolaires/universitaires, écrits journalistiques) (Blanche Benveniste 2010).

Exemple – Transcrire ses propres productions

Nos étudiants tendent parfois à attribuer certaines caractéristiques de la langue orale ordinaire (élisions, dislocations, absence du « ne » de négation, hésitations, reprises, recherches lexicales) au langage de l'enfant. C'est un point de départ intéressant pour les inviter à se questionner sur le point de référence qu'ils adoptent quand ils étudient les productions de l'enfant (langage adressé à l'enfant vs. certaines formes de français écrit). Transcrire leurs propres productions permet aux étudiants de prendre conscience « de l'intérieur » des spécificités de la langue orale « standard », « normale ». Ils découvrent également que les « accidents de la parole » (cf. Tran & Bécavin 2001) font partie d'un oral non pathologique, et qu'il sera nécessaire, lors de l'étude de productions de patients, enfants et adultes, d'essayer de distinguer ce qui relève du trouble et ce qui relève de la modalité dans laquelle est réalisée une tâche (cf. Loufrani 2000). Les analyses qu'ils feront ensuite les amèneront à réfléchir aux facteurs possiblement explicatifs de ces spécificités. S'interrogeant sur les facteurs externes, ils pourront être amenés à réfléchir aux métadonnées qu'il peut être important de collecter au moment du recueil pour pouvoir développer certaines hypothèses.

- 18 Le caractère systématique de la transcription va permettre aux étudiants de mesurer la fréquence de phénomènes qu'ils croyaient peu fréquents, comme les dislocations. Questionner les représentations qu'ils ont de la langue orale est une étape essentielle de leur formation car lors de leur pratique future, ils vont accompagner des patients vers un usage de la langue qui doit être un usage « réaliste », proche des repères qui lui sont pertinents pour évoluer dans son environnement social (Tran & Bécavin 2001).
- 19 Parallèlement, la transcription peut amener les étudiants à s'interroger sur leurs représentations de la syntaxe et de la syntaxe en acquisition. Par exemple, selon les conventions adoptées, la transcription peut prévoir une annotation des constructions potentiellement figées produites par l'enfant (Behrens 2008). Il est ainsi possible d'évoquer les hypothèses constructivistes appliquées à la syntaxe adulte (Goldberg 1995) et à la pathologie du langage (Leroy 2013).

3.2. Transcrire : poser des choix

- 20 De même que de nombreux choix sont posés en amont de l'enregistrement, de nombreux choix sont posés en amont de la transcription (outils, conventions), orientés par l'approche théorique adoptée et les analyses (domaine langagier, analyses qualitatives et/ou quantitatives) qui vont être menées.

Exemple – Transcrire avec CLAN

À Lille, nous formons les étudiants à transcrire avec CLAN (MacWhinney 2000) car c'est un programme libre, utilisé par une communauté internationale de chercheurs, relativement facile à prendre en main, dont l'interface ressemble aux outils de traitement de texte connus des étudiants, qui présente une certaine souplesse concernant l'alignement entre le texte de transcription et l'enregistrement (on peut aligner ou non, de manière plus ou moins précise) et qui permet de travailler à la fois avec des enregistrements audio et des enregistrements vidéo. Les analyses qui peuvent ensuite être menées comme le calcul de la longueur moyenne d'énoncés (LME) ou les mesures de diversité lexicale sont des analyses parfois menées par les orthophonistes au cours de leur pratique (Schelstraete & Maillart 2004). Les transcriptions peuvent en outre être exploitées avec d'autres outils comme PRAAT, PHON et ELAN. Enfin, la base de données CHILDES, dont les données sont lisibles avec CLAN, contient des données de parole normale et pathologique qu'il peut être intéressant pour les futurs professionnels de connaître, de savoir explorer et, pourquoi pas, de développer par la suite.

- 21 L'utilisation d'un outil informatique spécifique est souvent nouvelle pour ces étudiants de premier cycle. Si la majorité possède un ordinateur personnel, certains sont mal à l'aise voire effrayés quand il s'agit d'installer et de prendre en main un nouveau programme. Pourtant, avec le développement des batteries informatisées d'évaluation orthophonique, ces futurs praticiens seront sans doute fréquemment confrontés à l'usage de programmes informatiques spécifiques. Cette première expérience de création de corpus informatisé peut donc contribuer à « faire reculer la peur [...] face à la “pensée technique” (savoir lire une documentation, un manuel de référence, télécharger et installer un logiciel, etc.) » (cf. Habert 2004 : 9).
- 22 L'alignement du texte de transcription avec l'enregistrement est une technique qui a révolutionné les recherches sur l'oral, notamment parce qu'il permet de réécouter très facilement et aussi souvent que nécessaire l'enregistrement. Ceci permet de revenir aux productions effectives des locuteurs et au contexte. Réécouter facilement un segment permettra ainsi de compléter la transcription et/ou l'analyse des productions, ce qui est essentiel pour l'orthophoniste souvent confronté à des problèmes d'écoutes (cf. Loufrani 2000 ; Tran & Bécavin 2001).

Exemple – Aligner ou ne pas aligner ?

Au début de notre expérience, nous demandions aux étudiants d'aligner leur transcription avec leur enregistrement, pour les raisons que nous venons d'évoquer. Cependant, un certain nombre d'étudiants passaient beaucoup de temps sur cette étape et se laissaient décourager, suggérant que le rapport coût/intérêt de l'alignement n'était pas suffisamment important à ce moment-là (cf. Habert 2004), avec le risque de détourner ces futurs professionnels de la pratique de constitution de corpus. Par ailleurs, il nous a semblé que cette étape était moins proche de la réalité de la pratique orthophonique, dont le temps est particulièrement compté (da Silva & Masson 2017). Aussi, nous n'avons pas demandé aux étudiants des deux dernières années de réaliser cette étape. Nous en soulignons l'importance auprès d'eux par des exemples présentés en cours et les invitons à réécouter, revisionner leur enregistrement au moment de l'analyse.

- 23 La transcription d'un enregistrement audio ou vidéo place les étudiants face à un ensemble de décisions à prendre, allant de la représentation visuelle de la langue orale (cf. par exemple conventions du GARS avec une représentation non linéaire) à la segmentation des productions, en passant par l'adaptation de l'orthographe standard à la langue orale (avec, par exemple, le marquage de certaines formes d'assimilations comme « j'sais pas » prononcé [ʃɛpa] mais la question des « trucages orthographiques » soulevés par Blanche-Benveniste et Jeanjean [1987] se pose alors), le marquage éventuel de constructions figées ou qui pourraient être considérées comme telles (cf. Morgenstern & Parisse [2007] pour les productions d'enfants), la transcription de la prosodie, des pauses, des hésitations, des reprises, des rires, des raclements de gorge, des gestes, des chevauchements de parole... Même la transcription en API, dont on pourrait penser qu'elle représente de manière relativement fidèle les productions du sujet ne permet pas toujours de rendre compte exactement de ce qui est perçu. Les étudiants prennent alors conscience que la transcription est non seulement le reflet d'une perspective théorique sur le langage (cf. Ochs 1979) mais aussi de la subjectivité du transcripteur (cf. Blanche-Benveniste & Jeanjean 1987).

Exemple – Adopter les conventions CHAT

Dans un premier temps, nous invitons les étudiants à reprendre leurs transcriptions initiales et à réfléchir aux choix qu'ils ont posés : comment ont-ils représenté les productions et l'interaction ? Ont-ils utilisé l'orthographe, une orthographe adaptée à l'oral, l'API ? Ont-ils marqué les répétitions, les rires, les pointages... ? Dans un second temps, afin que les étudiants découvrent par eux-mêmes les conventions de transcription associées à CLAN (format CHAT), nous leur présentons des transcriptions déjà réalisées issues du corpus CoLaJE (Morgenstern & Parisse 2012). Ce n'est que dans un troisième temps que nous leur transmettons un document détaillant ces conventions précisant lesquelles doivent être respectées pour que le programme fonctionne et lesquelles sont spécifiques au traitement de données orales ou de productions d'enfants. Les étudiants sont alors invités à reprendre leurs transcriptions, à les insérer dans un fichier CLAN et à les adapter suivant ces conventions. Un ensemble de documents créés pour qu'ils puissent travailler en autonomie est mis à leur disposition ainsi qu'un forum pour qu'ils puissent partager leurs questions et leurs réponses. Face à l'étendue des possibles que dévoilent les conventions de transcription CHAT, les étudiants sont parfois tentés de vouloir « trop » transcrire. Leur objet d'analyse n'étant pas encore déterminé, nous les invitons donc à réaliser une transcription relativement minimale d'abord, proches de celles du corpus CoLaJE (Morgenstern & Parisse 2012) puisqu'il s'agit d'interactions adulte-enfant, avec transcription obligatoire des énoncés de l'enfant en API pour qu'ils pratiquent la transcription phonétique et pour qu'ils mesurent l'écart entre la transcription orthographique et ce qui est effectivement produit. Une fois leur sujet d'analyse déterminé, les étudiants sont parfois amenés à compléter leur transcription, soit pour y faire apparaître des éléments nouveaux (des gestes par exemple – ils découvrent alors que les conventions CHAT et le programme CLAN présentent des limites pour l'annotation du non segmental – cf. Adolphs & Knight [2010]), soit pour uniformiser et systématiser l'annotation de certains phénomènes (cf. Thompson 2004) comme par exemple les élisions : ils découvrent alors que les transcriptions sont des objets vivants amenés à évoluer en fonction des perspectives adoptées sur les données (cf. Traverso 2008).

- 24 La transcription place également les étudiants face à la question de l'interprétation des productions du sujet cible, en particulier lorsque le sujet cible et le transcripteur ne partagent pas la même variante de la langue ou lorsque le sujet cible est un enfant en cours d'acquisition de sa langue maternelle (cf. Morgenstern & Parisse 2007) et elles

sont fréquentes dans le cas de pathologie du langage (cf. Tran & Bécavin 2001). Au moment de la transcription, s'ils utilisent l'orthographe pour transcrire, les étudiants vont devoir interpréter (ou choisir de ne pas interpréter si le doute est trop important) les productions du sujet, ce qui impactera les analyses qui vont ensuite être menées.

- 25 Créer des espaces (cours, forum en ligne) où les étudiants peuvent mettre en mot les problèmes de transcription qu'ils rencontrent va leur permettre de mobiliser les concepts linguistiques et la terminologie acquise dans les UE précédentes, de travailler à une mise en mot qui soit descriptive et non prescriptive témoignant d'une prise en compte des spécificités de l'oral et du langage de l'enfant et de l'exercice d'un regard critique sur les différents choix opérés.

Conclusion

- 26 La constitution et l'analyse de corpus est l'occasion pour les étudiants en orthophonie de découvrir un certain nombre de questions qui font partie intégrante de la pratique orthophonique. En particulier, la sensibilisation à la problématique des choix qui sont à poser au moment de l'enregistrement et de la transcription, choix qui vont avoir un impact sur les analyses conduites (cf. Mondada 2002), n'est qu'une introduction à la démarche réflexive à laquelle ils vont ensuite être formés dans les enseignements cliniques et qu'ils vont être amenés à adopter lors de leur pratique future (Schelstraete & Maillart 2012).
- 27 Ce travail de constitution et d'analyse de corpus est un travail difficile pour les étudiants parce qu'il n'y a pas de réponse juste ou fausse, d'où la nécessité de souligner 1) l'importance d'explicitier les choix qui sont posés tout au long du processus et d'être conscient de l'impact de ces choix, 2) l'importance de se référer à la littérature pour appuyer ses choix et ses analyses.
- 28 Bien que la pratique de l'enregistrement soit préconisée par certains orthophonistes chercheurs (cf. Tran & Bécavin 2001), bien que certaines études suggèrent que c'est une pratique effectivement ancrée dans la pratique orthophonique (Witko 2010 ; Reynaud-Michaelides 2016), bien que des études mettent en évidence l'intérêt de l'analyse de corpus pour la pratique orthophonique (Witko 2010), les futurs professionnels montrent parfois des inquiétudes voire des réticences à pratiquer l'enregistrement et la transcription de données. Les raisons invoquées sont fondées mais des réponses peuvent être apportées. Ainsi, certains craignent une rupture de la relation entre le patient et le professionnel. À cette question, Witko (2010) répond que l'établissement d'une relation de confiance avec le patient permet un enregistrement sans méfiance. La question du temps que nécessite la transcription est également souvent évoquée. Il est certain que c'est un travail supplémentaire non rémunéré que le professionnel s'impose. À cette question, différentes réponses peuvent être apportées : au cours de la formation (initiale ou continue), on peut présenter l'étendue des possibles quant à l'outil utilisé pour transcrire ou aux informations effectivement transcrites pour que le professionnel puisse les adapter à ses besoins ; au niveau de la recherche en informatique, le développement de technologies facilitant la transcription pourra également apporter des éléments de soutien à cette pratique ; au niveau de la recherche en orthophonie, la présentation de travaux montrant l'apport de l'analyse de corpus sur l'évaluation et la prise en soin ou sur l'évaluation des pratiques professionnelles peut accroître l'intérêt pour cette pratique et en faire ainsi diminuer

le rapport intérêt/coût. Mais sans doute, pour être entendues, faut-il que ces réponses soient apportées par les praticiens eux-mêmes, car ils disposent de la légitimité nécessaire aux yeux des futurs professionnels. C'est la raison pour laquelle il est également important, tout au long de la formation initiale, d'inviter les étudiants à lire les travaux scientifiques publiés par les orthophonistes qui pratiquent l'analyse de corpus et de les inviter à questionner leurs maîtres de stage sur leurs pratiques de l'enregistrement, de transcription et d'analyse des enregistrements.

- 29 Les retombées de ce travail en 2^e année et les bénéfices de l'approche pédagogique adoptée sont difficiles à mesurer, en particulier parce qu'il n'y a plus d'enseignements en sciences du langage les années suivantes. On relève cependant que de plus en plus d'étudiants se lancent sans crainte dans la réalisation de mémoires de fin d'études reposant sur la constitution de corpus oral ou multimodal et qu'ils parviennent à convaincre les personnes enregistrées de l'intérêt de partager les enregistrements avec la communauté scientifique¹.
- 30 Nous n'avons pas abordé dans cet article les questions que soulèvent le partage et la diffusion des données collectées dans le cadre de la pratique orthophonique. En sciences du langage, c'est une pratique à laquelle les chercheurs sont fortement invités, afin de rendre accessibles les outils de la recherche (cf. Baude *et al.* 2006), mais qui reste une pratique freinée par des aspects notamment juridiques. En orthophonie, la diffusion de données met en outre en jeu la relation thérapeutique d'une part et le secret médical d'autre part, deux aspects supplémentaires susceptibles de freiner le partage de données collectées dans le cadre de la pratique orthophonique, pourtant nécessaire au développement de la recherche en pathologie du langage et en orthophonie. Dans la continuité de la réforme des études en orthophonie, il importe à présent de développer des cadres d'activité favorisant les collaborations entre professionnels et chercheurs.

BIBLIOGRAPHIE

Aldolphs S. & Knight D. (2010). « Building a spoken corpus : what are the basics ? », in A. O'Keeffe & M. McCarthy (éd.) *The Routledge handbook of corpus linguistics*. Londres, New York : Routledge, 38-51.

Atkins S., Clear J. & Ostler N. (1992). « Corpus design criteria », *Literary and linguistic computing* 7(1) : 1-16.

Baude O., Blanche-Benveniste C., Calas M.-F., Cappeau P., Cordereix P., Goury L., ... Mondada L. (2006). *Corpus oraux, guide des bonnes pratiques 2006*. Paris, Orléans : CNRS Éditions, Presses Universitaires Orléans.

Behrens H. (2008). *Corpora in Language Acquisition Research : History, methods, perspectives*. Amsterdam, Philadelphia : John Benjamins Publishing.

Blanche-Benveniste C. (2010). *Approches de la langue parlée en français*. Paris : Éditions Ophrys.

- Blanche-Benveniste C. & Jeanjean C. (1987). *Le français parlé : transcription et édition*. Paris, France : Didier.
- Blanche-Benveniste C. & Pallaud B. (2001). « Le recueil d'énoncés d'enfants. Enregistrements et transcription », *Recherches sur le français parlé* 1(16) : 11-37.
- Cappeau P. (2008). « Établir des corpus oraux : ce que nous apprennent les transcriptions », *Verbum* 30(4) : 343-353.
- Canut E., Masson C. & Leroy-Collombel M. (2017). « Efficience de l'étayage dans les pratiques professionnelles : effets sur le développement du langage d'enfants atteints de déficience intellectuelle », *Langage et Pratiques* 59 : 69-86.
- 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.
- Chambers A. (2009). « Les corpus oraux en français langue étrangère : authenticité et pédagogie », *Mélanges Crapel* 31 : 15-33.
- Chauvigné C. & Coulet J.-C. (2010). « L'approche par compétences : un nouveau paradigme pour la pédagogie universitaire ? », *Revue française de pédagogie. Recherches en éducation* 172 : 15-28.
- Da Silva Genest C. & Masson C. (2017). « L'apport de la linguistique de corpus à l'étude des situations cliniques : l'utilisation de ressources écologiques », *Studii de Lingvistica* 7 : 89-112.
- Da Silva C. (2011). « Orientations dialogiques dans une interaction orthophoniste-enfant dysphasique », in *Colloque international « Dialogisme : langue, discours », septembre 2010, Montpellier*.
- Debras C. (2018). « Petits et grands corpus en analyse linguistique des gestes », *Corpus* 18. Consulté à l'adresse <http://journals.openedition.org/corpus/3287>.
- Demuth K. (2008). « Exploring corpora for language acquisition research », in H. Behrens (éd.) *Corpora in Language Acquisition Research : History, methods, perspectives*. Amsterdam, Philadelphia : John Benjamins Publishing, 199-206.
- De Weck G. (2003). « Pratiques langagières, contextes d'interaction et genres de discours en logopédie/orthophonie », *Revue Tranel (Travaux neuchâtelois de linguistique)* 38 : 25-48.
- Gadet F. (1996). *Le français ordinaire*. Paris, France : A. Colin, 1996.
- Goldberg A. E. (1995). *Constructions : A Construction Grammar Approach to Argument Structure*. Chicago : University of Chicago Press.
- Habert B. (2004). « Outiller la linguistique : de l'emprunt de techniques aux rencontres de savoirs », *Revue française de linguistique appliquée* 9(1) : 5-24.
- Habert B., Nazarenko A. & Salem A. (1997). *Les linguistiques de corpus*. Paris, France : A. Colin, impr. 1997.
- Hriba L., Baude O. & Dugua C. (2017). « Le transcripateur transcrit : retour d'expérience à partir du corpus des ESLO », *Actes des 9^{es} Journées Internationales de la Linguistique de corpus*, 31.
- Labov W. (1972). *Sociolinguistic patterns*. Philadelphia : University of Pennsylvania Press.
- Leroy S. (2013). *Troubles de la généralisation dans les grammaires de construction chez des enfants présentant des troubles spécifiques du langage* (Thèse de doctorat). École doctorale Connaissance, langage et modélisation, Nanterre, France.

- Loufrani C. (2000). « Exploitation des données sur les corpus d'origine pathologique », in M. Bilger (éd.) *Corpus : Méthodologie et applications linguistiques* (vol. 3). Paris : Honoré Champion, 168-179.
- MacWhinney B. (2000). *The CHILDES project : tools for analyzing talk* (vol. 1-2). Mahwah, N.J., États-Unis d'Amérique : Lawrence Erlbaum.
- McNeill D. (2000). *Language and gesture* (vol. 2). Cambridge : Cambridge University Press.
- Michel M. (2002). « La démarche inductive en pédagogie », *Le Portique. Revue de philosophie et de sciences humaines* 9. Consulté à l'adresse <http://journals.openedition.org/leportique/182>.
- Mondada L. (2002). « Pratiques de transcription et effets de catégorisation », *Cahiers de praxématique* 39 : 45-75.
- 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(4) : 641-649.
- Morgenstern A. & Parisse C. (2007). « Codage et interprétation du langage spontané d'enfants de 1 à 3 ans », *Corpus* 6 : 55-78.
- Morgenstern A. & Parisse C. (2012). « The Paris Corpus », *Journal of French Language Studies* 22 : 7-12.
- Ochs E. (1979). « Transcription as theory », in E. Ochs & B. Schieffelin (éd.) *Developmental pragmatics*. Academic Press, 43-72.
- Reynaud-Michaelides N. (2016). *Recueil et analyse de productions orales spontanées et semi-spontanées dans la pratique orthophonique : intérêts théoriques et réalité clinique* (Mémoire d'orthophonie). Université de Lille.
- Rodi M. (2011). « Interactions logopédiste-enfant : Comment se construisent des échanges potentiellement acquisitionnels dans une situation de jeu symbolique », *Glossa* 110 : 49-68.
- Rondal J.-A. (2003). *L'évaluation du langage*. Bruxelles, Belgique : P. Mardaga.
- Schelstraete M.-A. & Maillart C. (2004). « Utilisation de Childe dans la pratique logopédique », *Les Cahiers de la SBLU* 16 : 45-49.
- Schelstraete M.-A. & Maillart C. (2012). « Les défis d'une formation universitaire clinique en logopédie », in *L'orthophonie et la logopédie au fil des ans-Regards sur des pratiques*. Issy-les-Moulineaux : Masson. Disponible sur ORBI. Consulté à l'adresse http://www.logoclinique.ulg.ac.be/upload/Txt_regardsFINAL-1.pdf.
- Thompson P. A. (2005). « Spoken language corpora », in M. Wynne (éd.) *Developing linguistic corpora : a guide to good practice*. Oxford : Oxbow, 59-70.
- Tomasello M. & Stahl D. (2004). « Sampling children's spontaneous speech : How much is enough ? », *Journal of child language* 31(1) : 101-121.
- Tran T.M. & Bécavin A.-S. (2001). « Evaluer l'expression orale : réflexion sur la pratique orthophonique », *Glossa* 78 : 22-33.
- Traverso V. (2008). « Analyser un corpus de langue parlée en interaction : questions méthodologiques », *Verbum-Revue de Linguistique* 30(4) : 313-328.
- Witko A. (2010). « Corpus et pathologie du langage. Quelle approche en orthophonie-logopédie ? », *Cahiers de praxématique* 54-55 : 119-144.

NOTES

1. En parallèle de ce travail auprès des étudiants, nous avons mené un travail important avec le référent CNIL de l'Université pour que les données personnelles des personnes filmées dans le cadre de mémoires soient protégées et nous avons par ailleurs mis en place un groupe de travail sur les interactions professionnel-enfant qui réunit des linguistes, des orthophonistes et des référents de l'école primaire, à la suite de quoi l'une de nos collègues orthophonistes a décidé de constituer, avec l'aide de ses stagiaires, un corpus d'interaction orthophoniste(/stagiaire)-enfant. C'est un travail qui sera prochainement proposé à l'ensemble des étudiants de 5^e année dans le cadre de leurs UE optionnelles.

RÉSUMÉS

Cet article rend compte d'une expérience pédagogique consistant à proposer à des étudiants de 2^e année en orthophonie de constituer leur propre corpus de productions orales ou multimodales (enregistrement audio ou vidéo et transcription de cet enregistrement). Il met en évidence un certain nombre de questionnements auxquels cet exercice conduit les étudiants et souligne les liens avec leur pratique future.

This paper reports on a pedagogical experimentation consisting in proposing speech therapy students to create their own corpus of spoken or multimodal productions (audio or video recording and transcription of this recording). It highlights a number of questions to which this exercise may lead students and underlines the links with their future practice.

INDEX

Mots-clés : orthophonie, formation initiale, constitution de corpus

Keywords : speech therapy, initial training, corpus building

AUTEUR

STÉPHANIE CAËT

Université de Lille & UMR 8163 - STL - Savoirs Textes Langage
stephanie.caet@univ-lille.fr