

Enseigner la création et l'utilisation de corpus langagiers à des étudiant-e-s en orthophonie : partage d'expérience pédagogique

Christophe dos Santos (*UMR 1253, iBrain, Université de Tours, Inserm, Tours*).

Le nouveau référentiel de formation du certificat de capacité d'orthophonie publié en septembre 2013 au bulletin officiel, spécifie entre autres, le contenu de chaque Unité d'Enseignement (UE) par module et sous-module. En ce qui concerne le sous-module 1.1 « sciences du langage », l'UE 1.1.4 intitulé « Connaissances fondamentales en sciences du langage appliquées à l'orthophonie », précise notamment, comme élément de contenu pour cette UE : « Présentation des outils d'analyse du discours et de corpus ». C'est dans le cadre de cette UE et pour répondre à cette demande qu'a été mis en place à l'école d'orthophonie de Tours un cours dénommé « Analyse linguistique clinique ». Ce cours vise principalement trois objectifs : une évaluation de la maîtrise de la transcription de la parole, nécessaire pour la pratique orthophonique, acquise lors de la première année ; une évaluation de la capacité à appliquer les connaissances acquises dans le domaine des sciences du langage, autres que la simple transcription, à l'analyse d'échantillons de langage normal et/ou pathologique ; et, l'initiation à des logiciels d'analyse du langage.

Pour tenter d'atteindre ces objectifs, la question de l'interface ou logiciel à utiliser s'est posée. Afin d'évaluer à la fois la phonétique/phonologie et la morphosyntaxe, le choix de TalkBank (MacWhinney, 2007) s'est imposé. Ce projet/système qui a pour objectif de faciliter le partage et l'accessibilité à des enregistrements et des transcriptions de productions normales ou pathologiques, héberge également des corpus de productions francophones. Dans cet objectif, TalkBank permet de mettre à disposition de nombreuses bases de données qui balayent un large champ de possibilités, de l'aphasie au développement du langage chez l'enfant. De plus, TalkBank propose, entre autres, deux outils, CLAN (Computerized Language ANalysis ; MacWhinney, 2000) et Phon (Rose et al., 2006 ; Rose & MacWhinney, 2014) qui facilitent, pour le premier, l'analyse morphosyntaxique des énoncés, et, pour le second, l'analyse phonologique des données. Enfin, les bases de données et outils disponibles ont été largement utilisés, et ce dans de nombreux domaines (bilinguisme, surdité, trouble du langage, développement phonologique...) ce qui a conduit à des milliers de publications scientifiques. Pour ne citer que deux exemples de publications en français, grâce aux outils de Talkbank, Delage (2008) a notamment réalisé une analyse syntaxique de productions spontanées d'enfants sourds moyens et légers, et MacLeod et collègues (2015)

ont pu établir le profil phonologique de sept enfants francophones ayant un trouble du développement des sons de la parole.

L'une des principales difficultés rencontrées pour qui doit former à ces outils est l'appropriation par les étudiants de ces outils. Cette appropriation par l'étudiant se réalise au fur et à mesure de la pratique qu'elle soit sur le lieu de formation ou à l'extérieur, mais au préalable, l'étudiant doit être convaincu du sens de ces travaux pratiques. L'importance d'introduire les outils par l'intermédiaire de cliniciens qui les utilisent ou les ont utilisés est essentielle, surtout s'ils proviennent de la sphère francophone. La langue peut en effet être également une barrière à cette appropriation, le site et les outils étant exclusivement en anglais. Ainsi, les exemples de travaux comme ceux de Delage (2008) ou MacLeod et collègues (2015), ne sont pas anodins car tous deux issues de la pratique clinique et du monde francophone.

Le développement des études sur corpus a permis et permet de s'interroger à la fois sur la pratique clinique et sur la méthodologie à adopter. En effet, les productions peuvent varier considérablement d'un individu à l'autre, en fonction de la situation, ou tout au long de la vie (Ferguson, Craig & Spencer, 2009). L'utilisation de corpus permet aux étudiants de réaliser pleinement les différences d'usages du français ainsi que ses différentes variétés (les étapes du développement langagier leur étant déjà connus). Ces variations qui peuvent s'éloigner de la norme sont ainsi comparées à ce qui relève de la parole pathologique, de l'agrammaticalité. Seule une méthodologie rigoureuse permet une comparaison valide. L'utilisation d'outils comme Phon rend compte de la nécessaire structuration des données et de leur traitement (segmentation, transcription, alignement) avant analyse (Rondal & Seron, 1999). Les extraits à choisir pour application en cours doivent nécessairement refléter au moins en partie ces dimensions afin d'inciter à l'interrogation des pratiques cliniques.

Pour finir, ces outils complets proposent, en général, un ensemble de programmes intégrés qui simplifie grandement l'analyse des données. Le formateur doit faire un choix parmi ces fonctionnalités en fonction à la fois du public visé mais également des extraits choisis. L'appropriation de l'outil se construit également par l'expérience que des analyses beaucoup plus fines que celles permises par les approches traditionnelles, ou que le traitement de corpus de dimensions beaucoup plus imposantes est possible. Peut-être qu'au-delà de la simple appropriation, certains de ses étudiants contribueront un jour à ces bases de données qui permettront sans doute à terme l'établissement de normes dans divers domaines.

Références bibliographiques

- Delage, H. (2008). Étude du langage spontané d'enfants sourds moyens et légers : Vers un évitement de la complexité syntaxique ? In Lidilem (ed.) : *Autour des langues et du langage : Perspective pluridisciplinaire*, 177-184. Grenoble : Presses Universitaires de Grenoble.
- Ferguson, A., Craig, H., Spencer, E. (2009). Exploring the Potential for Corpus-Based Research in Speech-Language Pathology. *Proceedings of the HCSNet Workshop on Designing the Australian National Corpus*. Somerville, MA: Cascadilla Proceedings. 30-36.
- France. Ministère de l'Enseignement Supérieur et de la Recherche. Enseignement supérieur et recherche. Certificat de capacité d'orthophoniste. Régime des études : décret n° 2013-798 du 30-8-2013 - J.O. du 1-9-2013. BO, bulletin officiel de l'Enseignement Supérieur et de la Recherche, 05 septembre 2013, n°32
- MacLeod, A. A., Hémond, M., Meziane, S., & Rose, Y. (2015). Le profil phonologique d'enfants ayant un trouble du développement des sons de la parole. *Rééducation Orthophonique*, 263, 87-104.
- MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk. Third Edition*. Mahwah, NJ: Lawrence Erlbaum Associates.
- MacWhinney, B. (2007). The TalkBank Project. In J. C. Beal, K. P. Corrigan, & H. L. Moisl (Eds.), *Creating and Digitizing Language Corpora: Synchronic Databases, Vol. I.* (pp. 163-180). Houndmills, Basingstoke, Hampshire: Palgrave-Macmillan.
- Rondal, J. A., & Seron, X. (1999). *Troubles du langage: bases diagnostiques et rééducation*. Bruxelles : Mardaga.
- Rose, Yvan, Brian MacWhinney, Rodrigue Byrne, Gregory Hedlund, Keith Maddocks, Philip O'Brien & Todd Wareham. 2006. Introducing Phon: A Software Solution for the Study of Phonological Acquisition. In David Bamman, Tatiana Magnitskaia & Colleen Zaller (eds.), *Proceedings of the 30th Annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press. 489-500.
- Rose, Yvan & Brian MacWhinney. 2014. The PhonBank Project: Data and Software-Assisted Methods for the Study of Phonology and Phonological Development. In Jacques Durand, Ulrike Gut & Gjert Kristoffersen (eds.), *The Oxford Handbook of Corpus Phonology*. Oxford: Oxford University Press. 308-401.