Constitution, analyse et exploitation de corpus: au carrefour de la recherche et de l'enseignement dans les centres de langue français

Saandia Vanessa Ali, Leonardo Contreras Roa* et Cristian Valdez**

CLLE-ERSS-Cognition, Langue, Langage, Ergonomie
LIDILE-Linguistique, Ingéniérie, Didactique des langues.

Mots clés : linguistique de corpus, constitution/exploitation des corpus, corpus écrits et oraux, corpus d'apprenants, secteur Lansad, interaction entre recherche et pratique de l'enseignement.

L'utilisation des corpus dans l'enseignement des langues suscite l'enthousiasme au sein de la communauté scientifique depuis déjà plusieurs années (cf. Braun 2005, Aston 2002, Capeau & Gadet 2007, Guilquin et al 2007, entre autres). Toutefois les applications et nouvelles pratiques potentiellement proposées par ces recherches restent relativement peu représentées dans le tissu universitaire français et en particulier dans les centres de langue qui se trouvent pourtant au cœur des réflexions innovantes autour de l'enseignement des langues. En effet, la mise en place d'expérimentations et l'utilisation des corpus par les enseignants de langue, qui pour la plupart ne sont pas des chercheurs en linguistique de corpus, restent chronophages, complexes et nécessitent des formations en amont pour les enseignants mais aussi pour les apprenants (Tyne 2012).

En tant que chercheurs en linguistique de corpus et enseignants dans un centre de langues, la prise en considération de l'interaction entre la recherche et la pratique d'enseignement constitue un axe essentiel de notre activité. Ainsi, dans cette communication, nous chercherons à montrer comment le corpus peut servir de vecteur dans la création d'un « espace commun de recherche et d'apprentissage » dans le sens où le corpus et l'enseignement s'enrichissent mutuellement. Pour ce faire, nous nous placerons en amont du processus de création d'activités didactiques basées sur les corpus, en interrogeant le dialogue entre la recherche fondamentale et le développement de méthodes et d'outils dédiés à l'enseignement des langues.

L'hypothèse qui sera mise à l'épreuve avance que l'apport réciproque entre corpus et eneignement se retrouve à tous les stades de la recherche - celui de la collecte et de la constitution du corpus, celui de la transcription et de l'annotation, et enfin au moment de l'analyse et de l'exploitation du corpus ainsi constitué.

Dans le but d'illustrer nos propos et de tenter de valider notre hypothèse, nous analyserons chaque étape mentionnée ci-dessus à la lumière d'exemples concrets tirés de quatre recherches en cours.

La première recherche, appliquée à l'enseignement de l'espagnol, présente l'élaboration de grands corpus écrits visant à enrichir et à informer l'enseignement d'un point grammatical précis : celui de l'utilisation de *ser* et *estar* en espagnol langue étrangère. La seconde cherche à

constituer et à analyser des corpus oraux d'apprenants d'anglais afin d'interroger l'enseignement de l'intonation en anglais pour des groupes d'apprenants francophones et hispanophones. La troisième montre comment on peut former les futurs enseignants en LANSAD par la collecte, l'élaboration, l'étiquetage et l'analyse de corpus oraux d'apprenants. La dernière expérience propose d'utiliser les modélisations extraites d'un corpus achevé et annoté afin de développer un dispositif de formation hybride pour l'enseignement des compétences prosodiques en anglais basé sur la didactisation des systèmes de reconnaissance automatique de la parole.

Références bibliographiques

Aston, G. 2002. "The learner as corpus designer". In B. Kettemann & G. Marko (eds), *Teaching and Learning by Doing Corpus Analysis*. Amsterdam: Rodopi, pp. 9-25.

Braun, S. 2005. From pedagogically relevant corpora to authentic language learning contents, *ReCALL* 17(1): 47-64.

Cappeau, P. & F. Gadet. 2007. L'exploitation sociolinguistique des grands corpus: Maître-mot
et pierre philosophale, *Revue Française de Linguistique Appliquée* 121: 99-110.

Guilquin, G, Granger, S & Paquot, M. 2007. Learner corpora: The missing link in EAP pedagogy, *Journal of English for Academic Purposes 6*: 319–335

Tyne, H. 2012. "Corpus work with ordinary teachers: Data-driven learning activities". In J Thomas & A. Boulton (eds.), *Input, Process and Product: Developments in Teaching and Language Corpora*, Brno: Masaryk University Press, pp.114-129.

Tyne, H.2009. Corpus oraux par et pour l'apprenant. *Mélanges CRAPEL* 31: 91-111.