LIDILE Linguistique Ingénierie et Didactique des Langues

Thèse Najib Arbach

Constitution d'un corpus oral de FLE : enjeux théoriques et méthodologiques

Date et lieu de soutenance : 2 mai 2015, Université Rennes 2 **Directrice :** Marie-Claude Le Bot (Université Rennes 2) **Président du jury :** Paul Cappeau (Université de Poitiers) **Membres du jury :**

- Dominique Legallois (Université de Caen)
- Élisabeth Richard (Université Rennes 2)

Résumé

Les méthodologies de constitution de corpus linguistiques ont été amplement étudiées, mais sont moins abondantes quand il s'agit de corpus oraux ; ces méthodologies sont encore plus rares en ce qui concerne l'interlangue orale. Le projet CIL (Corpus Inter Langue), en cours de finalisation à l'Université Rennes 2 et sous la supervision de l'équipe d'accueil LIDILE (EA 3874), vise à la constitution d'un corpus de productions écrites et orales d'apprenants en FLE et ALE. Cette thèse concerne le corpus oral de FLE du projet global (CIL-FLE). Partant du constat que l'intérêt des linguistes pour la langue orale a systématiquement été en retard par rapport à celui porté à la langue écrite, nous nous intéressons dans un premier temps à l'étude de l'oralité dans différents domaines de la linguistique d'un point de vue historique et épistémologique. Le second chapitre est consacré à la linguistique de corpus de manière générale et au corpus en tant qu'objet linguistique en particulier. En ce qui concerne la linguistique de corpus, nous tentons de présenter les différentes méthodologies auxquelles les linguistes ont recours lorsqu'il s'agit de consulter des données : introspection, élicitation ou consultation de données authentiques. Le concept de corpus est ensuite analysé selon un ensemble de critères définitoires que nous étudions en détail, afin de proposer une définition du corpus linguistique. Le troisième et dernier chapitre est la mise en application des constats théoriques dans la constitution du corpus CIL-FLE : nous détaillons les constituants du corpus, les protocoles de collecte et d'archivage. C'est au protocole de transcription que nous nous intéressons en particulier, en insistant sur les difficultés de la transcription de l'interlangue. Le corpus CIL-FLE, qui représente environ 105000 mots, représente le fruit de ce travail et sera ainsi détaillé.

Mots-clés : Corpus d'apprenants, parole spontanée, transcription de l'interlangue, français (langue) -- étude et enseignement -- allophones, corpus linguistique, communication orale, français (langue) -- langue parlée http://www.theses.fr/2015REN20014

Accéder à la thèse

Publication

Najib Arbach, Saandia Ali. Aspects théoriques et méthodologiques de la représentation des corpus. CORELA - COgnition, REprésentation, LAngage, CERLICO-Cercle Linguistique du Centre et de l'Ouest (France), 2014, 10.4000/corela.3029. hal-01616804