LIDILE Linguistique Ingénierie et Didactique des Langues

Najib Arbach's thesis

Constitution of an oral corpus of FLE: theotic and methodological issues

Najib ARBACH

Date and place of defense: May 2, 2015, Rennes 2 University **Director**: Marie-Claude Le Bot (Rennes 2 University) **President of the jury**: Paul Cappeau (University of Poitiers) **Jury members**:

- Dominique Legallois (University Caen)
- Elisabeth Richard (University Rennes 2)

Abstract

The need to design linguistic corpora to support research in linguistics has triggered the development of numerous studies exploring various approaches and methodologies regarding good practices for written corpus building. Fewer studies are available when it comes to spoken data and those that concern the interlanguage of learners are even rarer. The CIL project (Corpus Inter Langue), under completion at the University of Rennes2 and supervised by a research team specialising in the fields of linguistics and pedagogy (LIDILE), aims at building a large corpus of written and spoken productions in EFL and in FFL. This phd dissertation mainly focuses on the FFL (French as a Foreign Language) corpus (CIL-FLE).The first chapter of the thesis is dedicated to the study of oral speech as a linguistic object from both a historical and an epistemological perspective. The second chapter tackles the question of corpus linguistics generally speaking as well as the concept/ notion of corpus as a linguistic object. Regarding corpus linguistics, we will review and explore the diverse approaches and methods that are used so as to carry out research enquiries: introspection, elicitation or consultation of authentic data. The concept of corpus is then analysed according to/following a series of criteria which we will closely examine in order to propose a definition of the linguistic corpus. The third and last chapter will implement the former theoretical findings through the description of the CIL corpus design. Thus, corpus constituents, transcription and archiving protocols will be described in detail. We are particularly interested in the transcription protocol and we will insist on the difficulties encountered when attempting to transcribe learners 'data. Finally, the CIL-FLE corpus, which contains approximately 105 000 words and was developed all along this phd, will be described.

Keywords: Corpus of learners, spontaneous speech, transcription of interlanguage, French (language) -- study and teaching -- allophones, linguistic corpus, oral communication, French (language) -- spoken language http://www.theses.fr/2015REN20014

Accéder à la thèse

Publications

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