

Investigating L1-induced learner variability: Using the Web as a source of comparable data

Magali Paquot & Cédric Fairon

Learner interlanguage is characterised by a high degree of variability (Ellis 1994) that has commonly been accounted for by a number of external and internal factors, among others the influence of the first language (L1). Assessing L1 influence on learner variability is not an easy task and necessitates methodological rigor. Jarvis (2000) argues that transfer studies should minimally consider three potential effects of L1 influence: (1) similarities shared by members of an L1 population, (2) differences between several L1 populations and (3) performance similarities between the first language and the interlanguage of a given L1 population. The aim of an ongoing project on Phraseology and Foreign Language Learning and Teaching at Louvain is to apply this framework to all subcorpora of the *International Corpus of Learner English* (ICLE) (Granger et al. 2002). The difficulty however is to find L1 comparable corpora for each of the 11 mother tongue backgrounds represented in the corpus. In our presentation we will introduce an innovative method which makes it possible to extract corpora automatically from the Web using *Really Simple Syndication* (RSS) codes. The RSS technology is typically used by websites that change or add content regularly, e.g. newspaper websites. The open source software *Corporator* (Fairon forthcoming) exploits the RSS organisation in columns (business, sports, opinions, etc.) to extract and classify textual data sources. We will show how this method can be used to build dynamic corpora of editorials in several languages which can serve as comparable corpora to the ICLE argumentative essays. The benefits of the method will be demonstrated on the basis of a study of the potential influence of the French imperative on the heavy overuse of multiword units with “let’s” by French learners (Paquot 2005).

References

- Ellis R. (1994) *The Study of Second Language Acquisition*. Oxford: Oxford University Press.
- Fairon C. (forthcoming) *Corporator: du fil d’information RSS au corpus spécialisé*.
- Granger S., Dagneaux, E. & Meunier, F. (eds) 2002. *The International Corpus of Learner English*. CD-ROM and Handbook. Presses universitaires de Louvain: Louvain-la-Neuve. Available from <http://www.i6doc.com>
- Jarvis S. (2000) Methodological rigor in the study of transfer: identifying L1 influence in the interlanguage lexicon. *Language Learning* 50(2): 245-309.
- Paquot M. (2005) EAP vocabulary in learner corpora: a cross-linguistic perspective. Paper presented at *Phraseology 2005: The Many Faces of Phraseology*, Louvain-la-Neuve, Belgium, 13-15 October 2005.