

A variationist, corpus linguistic analysis of lexical richness in Flanders

A number of tests have been developed for measuring the lexical knowledge and use of language learners (see for example Read, 2000). Yet, while most of these tests focus on the extent of vocabulary acquisition of (typically L2) language users from an applied linguistic perspective, relatively little research has been conducted to investigate the distribution of lexical knowledge from a sociolinguistic, variationist point of view. This talk reports on research that is part of an ongoing PhD project, which attempts to apply and, where necessary, adapt existing measures, such as the type-token ratio (TTR), to the lexical knowledge of adult mother tongue speakers. A corpus-driven, quantitative methodology is proposed, analysing the *CGN* corpus (Corpus of spoken Dutch, Schuurman et al., 2003) for a number of sociovariational parameters such as register, region, and sex. In a multidimensional statistical analysis, this research project will attempt to explore possible correlations between these extralinguistic parameters and lexical richness. It will be demonstrated that a number of methodological complications have to be taken into account, partly with regard to the somewhat uneven distribution of linguistic material in the corpus, but mostly with regard to the measuring techniques (e.g. the text length-dependency of the TTR).

References:

- Biber, D. (1995). *Dimensions of register variation: a cross-linguistic comparison*. Cambridge: Cambridge University Press.
- Laufer, B. & Nation, P. (1995). Vocabulary size and use: lexical richness in L2 written production. *Applied Linguistics*, 16:307-322.
- Malvern, D.D. & Richards, B.J. (1997). A new measure of lexical diversity. In Ryan, A. and Wray, A. (Eds.), *Evolving models of language*. Clevedon: Multilingual Matters, p. 58-71.
- McCarthy, M. & Carter; R. (1997). Written and spoken vocabulary. In Schmitt, N. and McCarthy, M. (Eds.), *Vocabulary: Description, acquisition and pedagogy*. Cambridge: Cambridge University Press, p. 20-39.
- Nation, P. (1993). Vocabulary size, growth and use. In Schreuder, R. and Weltens, B. (Eds.), *The bilingual lexicon*. Amsterdam/Philadelphia: John Benjamins, p. 115-134.
- O'Loughlin, K. (1995). Lexical density in candidate output on direct and semi-direct versions of an oral proficiency test. *Language Testing*, 12 (2): 217-237.
- Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.
- Ure, J. (1971). Lexical density and register differentiation. In Perren, G.E. and Trim, J.L.M. (Eds.), *Applications of linguistics: Selected papers of the second international congress of applied linguistics, Cambridge, 1969*. Cambridge: Cambridge University Press, p. 443-452.

Author:

Sofie Van Gijsel
RU Quantitative Lexicology and Variational Linguistics, University of Leuven
Dept. Linguïstiek
Faculteit Letteren
Blijde-Inkomststraat 21
3000 Leuven
Belgium

e-mail: Sofie.VanGijsel@arts.kuleuven.ac.be
phone: 0032 (0)16 324824