A Corpus-based Study of Colloquial Flemish

Koen Plevoets, Dirk Speelman, & Dirk Geeraerts

RU Quantitative lexicology and Variational Linguistics
Department of Linguistics, University of Leuven
Blijde Inkomststraat 21
3000 Leuven, Belgium

e-mail: koen.plevoets@arts.kuleuven.ac.be
dirk.speelman@arts.kuleuven.ac.be
dirk.geeraerts@arts.kuleuven.ac.be

Historically speaking, Flanders did not develop a standard language of its own but adopted the Dutch standard that already existed in The Netherlands. In the more colloquial linguistic registers, however, Flanders has recently witnessed the emergence of the so-called ‘tussentaal’ (lit. ‘inbetween-language’), which is a supraregional language variety that still makes heavy use of phonological and morphological elements from the original Flemish dialects.

Focussing on the morphology, our presentation (which continues the line of variational research started in Geeraerts et al. 1999) attempts to uncover the underlying register dimensions that structure the everyday speech characteristics in Flanders, thus establishing some ‘measure of colloquialism’. In addition, we will study the distribution of this colloquialism measure over several sociolinguistic parameters, such as region, age, sex, educational level, and occupational level. The investigation reveals that the ‘tussentaal’ register is far from homogeneous, but is itself structured along different dimensions.

The method of inquiry is corpus-based. The latent dimensions are accounted for by means of the statistical techniques of principal components, correspondence, and factor analysis; the variational investigation involves logistic regression. The corpus of data is the recently released ‘Corpus Gesproken Nederlands’, which consists entirely of transcriptions of 10 million spoken Dutch words.