Explaining register and sociolinguistic variation in the lexicon: Corpus studies on Dutch

This study combines the fields of sociolinguistics and corpus linguistics in investigating global lexical variation in two large corpora. It expands the knowledge on the role of register and sociolinguistic factors (country, gender, age, and education level) in shaping the way lexical characteristics vary in both written and spoken Dutch. The study specifically targets lexical productivity and derivational morphology.

In corpus linguistics the emphasis is on the effects of register on global text characteristics. The emphasis in variationist studies in sociolinguistics is on the impact of social factors on specific linguistic variables. The combination of these fields proves to be successful. Both language use and the language user emerge as important sources of lexical variation.

Concerning register, the highest derivational and lexical productivity are found in the most formal registers of spoken and written Dutch. Concerning social factors, the most important finding on differences between the Netherlands and Flanders is that variation patterns are primarily word-bound, and can probably be traced back to divergent lexical choices in expressing specific concepts. A high derivational and lexical productivity, a high Type-Token Ratio, and a high proportion of nouns, all characteristics of a more ‘informational’ speech style, characterize men’s speech. A high proportion of verbs and most common words, typical of a more ‘involved’ speech style, characterize women’s speech. Older highly educated speakers are most productive, mainly in situations that evoke the use of more ‘informational’ language, indicating that a speaker’s lexical knowledge increases during the lifetime.