What can large monolingual corpora tell us about dialect/standard code-switching?

NIKOLAY KHAKIMOV
University of Freiburg

Code-switching between dialects and standard varieties has been researched in various contact situations in Europe at least since Blom and Gumperz (1972). Dialect/standard code-switching exhibits patterns that differ from patterns involving distantly related languages (Spanish and English). Evidence has accumulated that dialect/standard code-switching data often challenge morphosyntactic constraints regulating other code-switching, such as the equivalence constraint (Poplack 1980) and the MLF model (Myers-Scotton 1993).

Using a Russian-German corpus of bilingual speech, Khakimov (submitted) demonstrates that the co-occurrence frequency of lexical pairs in one of the involved languages affects choices in code-switched clauses, suggesting that frequency distributions in the standard and in the dialect can influence dialect/standard code-switching, as in the following:

(1) Swiss German dialect and Swiss Standard (Krohn 2008: 111, quoted in Petkova 2011)
   Säb Weibervolk war ein *kräftiges Mentsch* […].
   Stand. G.: Seine Frau   war ein kräftiger Mensch.
   'His wife was an athletic person.'

(2) Sicilian dialect and Italian (Alfonzetti 1992: 307)
   È così, ê figghi râ   signora L. macari, chiddu piccolo, ca ora appoi
   Stand.I.:   ai figli   della signora L. anche, quello piccolo, che
   diventò grande.
   'So, and also the sons of Mrs. L., that little one, who has grown now.'

The switched fragments in (1-2) include words co-occurring frequently in the respective varieties, i.e., *kräftiges Mentsch* 'athletic person' in (1), and *diventò grande* 'became big' in (2). Whether co-occurrence frequency determines the use of such structures in the situation of dialect/standard code-switching needs further studies.

In this talk, I will present a method that allows to investigate the impact of co-occurrence frequency on the structure of code-switching. Specifically, I will show how large monolingual corpora can be used to investigate phenomena emerging in situations of language contact. The talk will focus on modelling the competition between word strings that vary in frequency. This was investigated in a bilingual Russian-German corpus comprising natural conversations and informal group interviews of 20 young Russian-speaking adults living in Germany. We focused on the prepositional phrase, where the switch is either before or after the preposition:

(3) smotreli v *englischunterricht*   (4) byla *in tschechien*
   watched in English class   was in Czechia

German nouns occurring in the examined phrases were analyzed in the large German deWaC corpus with regard to their use with prepositions. The collocational distribution of each target noun was modelled by odds, the ratio of the frequency of the target string to the summed frequencies of the other strings with the same noun.

The study of prepositional phrases included priming as a further predictor of switch-placement. Exposure to the target preposition in the previous discourse influenced the choice of the language for preposition realization. To study the interplay of these factors, the generalised linear mixed model was employed. The statistical model was of a high predictive power. The results are interpreted as evidence for (1) the cognitive reality of multi-word strings and (2) the role of lexical priming in production.