Evaluating the similarity of dialect grammars: a corpus-based variationist approach
Benedikt Szmrecsanyi (KU Leuven)

There is a comparatively extensive literature on determining the grammatical similarity of dialects based on dialect atlas data (see, e.g., Spruit, Heeringa & Nerbonne 2009), but measuring similarity in naturalistic corpus data is a trickier task without well-trodden methodologies. In this talk, I sketch a corpus-based variationist method for calculating the similarity of dialect grammars: what counts is not if and/or how often people use particular constructions, but how – that is, subject to which probabilistic constraints – they choose between “alternate ways of saying ‘the same’ thing” (Labov 1972:188). As a case study, I discuss similarity patterns between regional varieties of English, fueled by a variationist analysis of three major alternations in the grammar of English: the dative alternation, as in (1), the genitive alternation, as in (2), and the particle placement alternation, as in (3).

(1) The dative alternation (see Röthlisberger, Grafmiller & Szmrecsanyi to appear)
   a. I’d given Heidi my T-Shirt (the ditransitive dative variant)
   b. And I’d given the key to Helen (the prepositional dative variant)

(2) The genitive alternation (see Heller, Szmrecsanyi & Grafmiller 2017)
   a. the country’s economic crisis (the s-genitive)
   b. the economic growth of the country (the of-genitive)

(3) The particle placement alternation (see Grafmiller & Szmrecsanyi in preparation)
   a. you can just cut the tops off (verb-object-particle order)
   b. cut off the flowers (verb-particle-object order)

These alternations are studied in corpus material covering some nine international varieties of English (British English, Canadian English, Irish English, New Zealand English, Hong Kong English, Indian English, Jamaican English, Philippine English, and Singapore English). Analysis shows that the probabilistic grammars regulating variation between the (a) and (b) variants are overall fairly similar across varieties of English – in other words, we seem to be dealing with a rather solid “common core” (in the parlance of Quirk et al. 1985:33) of the grammar of English. That said, we do find more or less subtle differences both between regional varieties as well as between syntactic alternations in terms of how hospitable they are to probabilistic indigenization.

References