

Benedikt Szmrecsanyi

Variation squared

Inspired by work in comparative sociolinguistics (e.g. Tagliamonte 2001) and quantitative dialectometry (e.g. Nerbonne, Heeringa & Kleiweg 1999), this talk sketches a corpus-based, variationist method (Variation-Based Distance & Similarity Modeling – VADIS for short) to rigorously quantify the similarity between varieties and dialects as a function of the correspondence of the ways in which language users choose between "alternate ways of saying 'the same' thing" (Labov 1972: 188). In other words, the basic idea is to measure inter-speaker variability by assessing intra-speaker variability: hence "variation squared". To showcase the potential of the method, I present a case study that investigates three syntactic alternations (the dative alternation, the genitive alternation, and the particle placement alternation) in some nine international varieties of English. Key findings include the following: (a) probabilistic grammars are remarkably similar and stable across the varieties under study; (b) there is a fairly robust split between "native" (a.k.a. Inner Circle) varieties, such as British English, and "non-native" (a.k.a. Outer Circle) varieties, such as Indian English; and (c) there is not much coherence across alternations, which suggests that probabilistic grammars in the aggregate are a collection of independent and independently constrained alternations.

References

- Labov, William. 1972. *Sociolinguistic patterns*. Philadelphia: University of Philadelphia Press.
- Nerbonne, John, Wilbert Heeringa & Peter Kleiweg. 1999. Edit Distance and Dialect Proximity. In David Sankoff & Joseph Kruskal (eds.), *Time Warps, String Edits and Macromolecules: The Theory and Practice of Sequence Comparison*, v–xv. Stanford: CSLI Press.
- Tagliamonte, Sali. 2001. Comparative sociolinguistics. In Jack Chambers, Peter Trudgill & Natalie Schilling-Estes (eds.), *Handbook of Language Variation and Change*, 729–763. Malden and Oxford: Blackwell.