Computational Linguistics and Variationist Linguistics
John Nerbonne
University of Groningen & University of Freiburg

Computational linguistics techniques have often been used in dialectology, especially edit-distance measures of pronunciation differences transcriptions, stemming for comparing lexical choices, part-of-speech tagging for comparing syntaxes, and a range of machine-learning techniques for detecting structure in underlying distributions. These have become part of the dialectometry 'tool kit' (Wieling & Nerbonne 2015). Given the Chambers-Trudgill thesis that language variation ought to constitute a single discipline – whether the variation be geographically or socially conditioned – we should expect similar sociolinguistic studies to be fruitful. In fact we have now a number of studies which suggest that computational techniques may be fruitfully applied in sociolinguistics, but these are much less common (Nguyen et al. to appear).

One factor that may have slowed the uptake of computational techniques, namely the focus of one line of sociolinguistic research on studying individual sound changes in progress. Most dialectometric work proceeds by examining aggregate differences among varieties (Goebl 1984; Nerbonne 2009), making the study of individual changes less primary. But we note that computational techniques for isolating individual differences have also been developed (Prokic et al. 2012), and we note several areas of socio-linguistics where the aggregate perspective seems promising.

Finally, given the problematic distinction between languages and dialects, it is not surprising that "dialectometric" techniques have found interest among researchers in language documentation (Mennecier et al. to appear, Snoek 2014).

Hans Goebl (1984) Dialektometrische Studien. 3 Vol. Tübingen: Max Niemeyer.

Philippe Mennecier, John Nerbonne, Evelyne Heyer and Franz Manni (to appear) A Central Asian language survey. Collecting data, measuring relatedness and detecting loans. *Language Dynamics and Change* 2016

Nerbonne, John (2009) Data-Driven Dialectology *Language and Linguistics Compass* 3(1), 175-198. DOI: 10.1111/j.1749-818x2008.00114.x

Dong,Nguyen, A. Seza Doğruöz, Carolyn P. Rosé & Franciska de Jong (to appear). Computational Sociolinguistics: A Survey. *arXiv preprint arXiv:1508.07544*. Accepted to appear in *Computational Linguistics*.

Jelena Prokić, Çağrı Çöltekin & John Nerbonne. Detecting Shibboleths (2012) In: Miriam Butt & Jelena Prokić (eds.) *Visualization of Language Patterns and Uncovering Language History from Multilingual Resources*. Workshop at 13th EACL. Avignon, France. 72-80.

Conor Snoek (2014) Review of Gabmap: doing dialect analysis on the web. *Language Documentation and Conservation* 8:192–208

Wieling, Martijn and John Nerbonne (2015) Advances in Dialectometry. *Annual Review in Linguistics* 1(1):243-264. doi:10.1146/annurev-linguist-030514-124930

Martijn Wieling, Simonetta Montemagni, John Nerbonne and R. Harald Baayen (2014) Lexical Differences between Tuscan Dialects and Standard Italian: A Sociolinguistic Analysis using Generalized Additive Mixed Modeling. *Language* 90(3): 669-692