The near horizon in dialectology John Nerbonne Groningen, Freiburg and Tübingen

This programmatic lecture will present innovations that the subdiscipline dialectometry (Séguy 1971; Goebl 1982; Wieling & Nerbonne 2015) has contributed in order to sketch the most challenging emerging research questions these contributions have enabled. We focus on pronunciation.

Dialectometry has been a concentrated effort to innovate in the *methods* of dialectological research. *Pace* some researchers, it has not been an effort to found a new field or even to revolutionize its parent subdiscipline, dialectology. Having said that, we note that some non-methodological innovations have also been enabled and suggest below that the time has come to suggest further possibilities, or at least, a shift in theoretical attention. The essential step forward in dialectometry has been to MEASURE dialect differences – the different lexicalizations of the same concept (Séguy 1971) or the degree of difference in pronunciations of the same words (Heeringa 2004). This has at times meant resorting to the simplest of measures, same vs. different, but it has also stimulated the development of sensitive measures of pronunciation differences optimized for use with phonetic transcriptions (Heeringa 2004; Wieling et al. 2012) or for acoustic recordings (Barthelds et al. 2020).

A second step in dialectometry's development has been to refocus attention from details of dialect differences to the level of aggregate differences between varieties, i.e. between data collection sites or even entire dialect areas (Nerbonne 2009). Since the early papers measured only at a categorical level, the step toward examining aggregates was natural. The earliest work (Séguy, Goebl) collected differences from all linguistic levels, but it is now customary to focus on single linguistic levels such as lexis (word choice), phonetics (pronunciation), morphology or syntax, where most work has been devoted to the first two levels. The shift in focus to aggregate differences stimulated in turn the question as to the degree to which individual elements contributed to them. We'll report on a recent collaboration aimed at identifying characteristic sounds in varieties (Rubehn et al. 2024).

Dialectometry borrowed from computational linguistics analytical procedures such as the application of edit distance to phonetic transcriptions (Heeringa 2004), but also an insistence on evaluating work for reliability. It has further followed psychometrics in seeking independent validation for analyses.

The combination of these innovations has enabled forays into the general relation between geography and language variation. In an effort to detect the dialect regions that early dialectology postulated as the primary organizing element in the distribution of linguistic variation, Goebl (1982) turned to clustering, which is still of service (Prokić & Nerbonne 2008), but following Embleton (1993), who introduced multidimensional scaling (MDS) to dialectology, Leinonen (2011) used MDS to conclude in an analysis of Swedish that regions played little role in the overall continuum she adduced from a large sample. Dialectometry has also enabled a gratifying burgeoning interest in the influence of geography on dialectal distribution. Grieve (2018) champions the role of geostatistics for dialectometry, and Burridge (2017) researches the impact of physical geography on dialect variation.

Morphology and syntax are still woefully understudied, and there is no clear candidate for an analysis technique that could generate comparable syntactic analyses of informal material. The analysis of the geographic distribution of variation is often interpreted as showing diffusion, but too little concentrated work has been carried out on the dynamics of this diffusion. We close with an appeal for additional research energy devoted to dialectal dynamics, a focus of current collaboration, which would amount to a shift in theoretical attention (Lameli et al. in preparation).

References

- Bartelds, Martijn, Caitlin Richter, Mark Liberman & Martijn Wieling 2020. A new acoustic-based pronunciation distance measure. *Frontiers in Artificial Intelligence*, *3*, 39.
- Burridge, James 2017. "Spatial evolution of human dialects." *Physical Review X* (7.3): 031008.
- Embleton, Sheila 1993. "Multidimensional scaling as a dialectometrical technique: Outline of a research project." *Proc. 1st Int. Conf. Quantitative Linguistics, QUALICO*. Berlin:Springer. 267-276.
- Goebl, Hans 1982. Dialektometrie; Prinzipien und Methoden des Einsatzes der numerischen Taxonomie im Bereich der Dialektgeographie. (Philosophisch-Historische Klasse Denkschriften 157) Vienna: Verlag der Österreichischen Akademie der Wissenschaften.
- Grieve, Jack 2018. "Spatial statistics for dialectology." In Boberg, C. et al. (eds) *The handbook of dialectology*: 415-433.
- Heeringa, Wilbert J. 2004 *Measuring dialect pronunciation differences using Levenshtein distance*. PhD Dissertation, Rijksuniversiteit Groningen.
- Lameli, Alfred, Simonetta Montemagni and John Nerbonne (in preparation) Dialectal Dynamics, spec. issue of Languages. (in preparation to appear in 2025).
- Leinonen, Therese 2011. "Aggregate analysis of vowel pronunciation in Swedish dialects." *Oslo Studies in Language* 3(2).
- Nerbonne, John 2009. Data-driven dialectology. Language and Linguistics Compass, 3(1):175-198.
- Prokić, Jelena, and John Nerbonne 2008. "Recognising groups among dialects." *International journal of humanities and arts computing* 2(1-2): 153-172.
- Rubehn, Arne, Simonetta Montemagni, and John Nerbonne. 2024. "Extracting Tuscan phonetic correspondences from dialect pronunciations automatically." *Language Dynamics and Change* 14.1: 1-33.
- Séguy, Jean 1971. "La relation entre la distance spatiale et la distance lexicale." *Revue Lingusitique Romane* 35: 335-357.
- Spruit, Marco René 2006. "Measuring syntactic variation in Dutch dialects." *Literary and linguistic computing* 21(4): 493-506.
- Wieling, Martijn, Eliza Margaretha & John Nerbonne 2012. "Inducing a measure of phonetic similarity from pronunciation variation." *Journal of Phonetics* 40(2): 307-314.
- Wieling, Martijn 2012. A quantitative approach to social and geographical dialect variation. PhD Dissertation, Rijksuniversiteit Groningen.