

# Chapter 1

## Embracing the future of dialects

Marie-Hélène Côté

Université Laval

Remco Knooihuizen

University of Groningen

John Nerbonne

University of Groningen, University of Freiburg

The conference *Methods in Dialectology XV* was held in Groningen on 11–15 August 2014. In October 2014 we issued a call for a contribution to a volume of proceedings, which led to a gratifying number of excellent reactions. This brief introduction tells a bit more about the conference and provides some orientation to the papers in the volume.

### 1 The conference

The conference was the fifteenth in the series *Methods in Dialectology*, which started in 1972. It has “generally alternated between Europe and Canada”.<sup>1</sup> Following its predecessors, *Methods XV* issued a broad call for contributions to the conference, emphasizing that areal, social and historical perspectives have all been regarded as tributaries to the discipline of variationist linguistics at least since Chambers and Trudgill’s programmatic work (Chambers & Trudgill 1980, Chambers & Trudgill 1998: Chapter 12).<sup>2</sup> Shortly after the conference (in September 2014) we issued a call for papers for this volume among those who presented at *Methods XV*. We asked for the usual length conference papers, but we also allowed the submission of brief papers (in this volume, papers of six to ten pages).

---

<sup>1</sup> *Methods and Methods 14*, [http://westernlinguistics.ca/methods14/methods\\_14.html](http://westernlinguistics.ca/methods14/methods_14.html), consulted 20 April 2015.

<sup>2</sup> See Kleiner (2014) for an independent report on *Methods XV*.



The conference featured five plenary lectures. Jacob Eisenstein of the Georgia Institute of Technology talked on “Dialectal variation in online social media”, and Frans Gregersen of the University of Copenhagen, delivered a lecture entitled “A matter of scale only?” on the different temporal scales analyzed in variationist linguistics on the one hand and historical linguistics on the other. Mark Liberman of the University of Pennsylvania sketched new technical possibilities for collecting and analyzing linguistic data automatically in a lecture entitled “The dialectology of the future”, Naomi Nagy of the University of Toronto presented her research on “Heritage languages as new dialects”, and Brigitte Pakendorf of the Université Lyon 2 “Lumière” spoke on “Dialectal variation and population genetics in Siberia”. A tutorial on Gabmap (Nerbonne et al. 2011) was given by Wilbert Heeringa and Therese Leinonen, while a workshop on integrating perceptual dialectology and sociolinguistics with geographic information systems was organized by Lisa Jeon, Patricia Cukor-Avila, Chris Montgomery and Patricia Rektor. Special sessions on various topics were organized, including one on open access publishing by Martin Haspelmath of Language Science Press. There were 140 single-paper presentations during four-and-a-half days.

The organizers of the conference were especially happy to include – we think for the first time – a poster session consisting of fourteen posters, two of which were awarded prizes named after Lisa Lena Opas-Hänninen, a frequent participant at *Methods*, and co-organizer of *Methods XI* in Joensuu, Finland in 2002. These poster prizes for young scholars were generously funded by the Alliance for Digital Humanities Organizations, and papers by both recipients may be found in this volume, “Imitations of closely related varieties” by Lea Schäfer, Stephanie Leser and Michael Cysouw, and “Infrequent Forms: Noise or not?” by Martijn Wieling and Simonetta Montemagni.

Cambridge University Press generously offered to underwrite two prizes for best papers by young scholars. The “Chambers prizes” are named after Jack Chambers, one of the most prominent figures in variationist linguistics of the last half century, and a source of energy, wisdom and inspiration for the *Methods* series. These papers are also included in the volume, namely Anne-Sophie Ghyselen’s “Structure of diaglossic language repertoires: Stabilization of Flemish *tussentaal*?” and Simon Pickl’s “Fuzzy dialect areas and prototype theory. Discovering latent structures in geolinguistic variation”.

## 2 The papers

### 2.1 Dialects' Future

Traditional, geographically defined dialects are losing ground, in particular to standard languages in Europe. Naomi Nagy's paper "Heritage languages as new dialects" examines the speech of Cantonese, Faetar, Italian, Korean, Russian and Ukrainian immigrants to Canada from a perspective complementary to the usual European one (Auer, Hinskens & Kerswill 2004), where we see basilectal varieties being eroded under the influence of standard speech, but one which Trudgill (2004) has pursued in depth. In contrast to the "intimate contact" between dialect and standard in Europe, Nagy emphasizes the need to incorporate methods and perspectives from the study of language contact (Hickey 2010). She points to evidence of diversion from varieties in native countries and, true to the focus of *Methods* conferences, devotes the lion's share of her paper to the presentation of methods in use in a large Toronto project.

In fact the erosion of very specific varieties has given rise to the study of "regiolects" (Auer & Hinskens 1996) – forms of speech intermediate between the basilectal varieties of a village or small town and standard languages, typically used in national communications such as radio and television. As Auer (2005: 22) noted, the forms of speech may not be homogeneous and stable enough to deserve the name "variety". Anne-Sophie Ghyselen's prize-winning paper on Belgian Dutch *tussentaal* examines how stable this intermediate form of speech has become, concluding that *tussentaal* is too heterogeneous and unstable to be regarded as a variety, just as Lenz (2003) concluded for regional forms used in the Eifel.

### 2.2 Methodological contributions

#### 2.2.1 Dialectometry

Dialect "areas" constituted the standard means of presentation of dialectological wisdom about the influence of geography of variation for many decades even if it was recognized that continua were also to be found in the data, and that areas, when found, were often delimited by vague borders. Simon Pickl's prize-winning paper "Fuzzy dialect areas and prototype theory. Discovering latent structures in geolinguistic variation" suggests that it is time to eschew dialectometric techniques such as clustering, which always yields sharp partitions among data collection sites, in favor of techniques such as factor analysis, which give rises to

areas with “fuzzy” borders. He links these ideas to prototype theory in cognitive science and Berruto’s notion of “condensation areas”.

Andrea Mathussek examines “the problem of field worker isoglosses” as she encounters these in the *Sprachatlas von Mittelfranken* (‘Dialect Atlas of Middle Franconia’, SMF, Munske & Mathussek 2013). Mathussek emphasizes that the field workers were aware of the potential problems and actively took measures to try to avoid idiosyncrasies in transcription, e.g. transcribing the same data as an exercise and then comparing the results, but differences remained. Mathussek used the web application Gabmap (Nerbonne et al. 2011), which is based on dialectometric techniques, to show that the field worker effects persisted even into aggregate levels of comparison. It was crucial for tracking the effects that Gabmap supports the identification of characteristic elements of clusters (Prokić, Çöltekin & Nerbonne 2012).

Simonetta Montemagni and Martijn Wieling focus on lexical dialectology and apply an alternative calculation for identifying characteristic features in “Tracking linguistic features underlying lexical variation patterns: A case study on Tuscan dialects”, namely one based on graph theory (Wieling & Nerbonne 2011). They note that dialectometry identifies groups similar to those in traditional Tuscan dialectology, but go on to identify which words are most characteristic, introducing *en passant* the innovation in combining the measures of how representative and how distinctive features are. They combine not additively, as earlier work had, but multiplicatively, effectively ensuring that only features that score highly on both components are regarded as characteristic. Montemagni and Wieling also attend to age differences in their analyses.

Jelke Bloem, Martijn Wieling and John Nerbonne apply a technique developed in dialectometry, namely a quantitative measure of how characteristic a speech trait is, to a non-dialectological problem, namely automatically identifying characteristic features of non-native English accents, in their paper of the same title. It has long been recognized that there are parallels between traditional dialects and socially delimited varieties on the one hand and contact varieties on the other (Trudgill 1986), but the authors likewise claim that the introduction of dialectometric techniques into the study of foreign accents may improve the latter by providing aggregate perspectives in an area that has largely relied on the study of a small number of phenomena.

Tyler Kendall and Valerie Fridland’s “Mapping the perception of linguistic form: Dialectometry with perceptual data” proposes a collaboration between two of the most innovative strands within modern variationist linguistics, namely perceptual phonetics and dialectometry. They focus on the varying boundaries

of vowel perception within the US and examine inter alia the relation between perception and production boundaries. Given perceptual dialectology's standard attention to social factors (Niedzielski 1999), their collaboration also entails understanding how dialectometry and sociolinguistics might join forces, an area which has received too little attention thus far (Nerbonne et al. 2013). On the dialectometric side they make extensive use of the geo-statistical techniques Grieve, Speelman & Geeraerts (2011) have championed.

While Philipp Stoeckle does not identify his contribution "Two dimensional variation in Swiss German morpho-syntax" as dialectometrical, he makes use of the Delaunay-Voronoi techniques made popular by Goebel (2006, and references there) and he aggregates over 57 different syntactic items to obtain an index of variation, effectively the degree to which forms at a given site agree with the most frequent one. This provides insight into a second dimension in his study of variation in addition to the geographic "one". The paper is also notable for its quantitative attention to syntax, an area where Spruit (2008) still stands as one of the few more substantial works. Given the syntactic focus, there are not lots of alternatives to Stoeckle's measure of local variability, but Kretzschmar, Kretzschmar & Brockman's work (2013) on the Gini coefficient would be an interesting alternative.

As Martijn Wieling and Simonetta Montemagni note in their note "Infrequent forms: Noise or not?" opinions differ as to the value of including infrequent forms. Goebel (1984) introduced an inverse frequency measure to count infrequent items as stronger indications of dialectal similarity, and Nerbonne & Kleiweg (2007) provide empirical confirmation of the wisdom of this step. But corpus-based approaches often insist on the opposite, effectively ignoring infrequent items due to their inherent unreliability. It may turn out that some differences are due to the different data collection techniques. After all, since there's no guarantee of having exactly commensurable items in corpus-based work, some "trimming" is inevitable, while the use of questionnaires and check lists ensures that information on even infrequent items normally will be elicited.

Christoph Wolk and Benedikt Szmrecsanyi provide a very useful overview and comparison of approaches in "Top-down and bottom-up advances in corpus-based dialectometry". The earliest work was done by Szmrecsanyi, who collected frequencies of 57 morphosyntactic features, specifying the features ahead of time in a "top-down" manner, converted these to relative frequencies and applied a logarithmic transformation to prevent frequent elements from dominating the measure. In a probabilistic variant, generalized additive models are used to predict the values, and the predicted values are used, effectively smoothing the log

relative frequencies. The third, bottom-up technique uses bigrams of part-of-speech tags (POS tags) in the entire corpus, obviating the need to select features ahead of time. The reliability of the POS bigrams is assayed using a resampling procedure, yielding features for analysis. Wolk (2014) promises all the details!

### 2.2.2 Other methods

Lea Schäfer, Stephanie Leser and Michael Cysouw report on two interesting data sets collected to investigate the mechanisms of imitating closely related language varieties in “Mechanisms of dialect imitation”. The poster presentation won one of the “best poster” awards at the conference. Purschke (2011) was one of the earliest works on dialect imitation, but Schäfer and colleagues build on Myers-Scotton’s (1993) model of code-switching between different languages, and their goal is to learn not only about the language being imitated (the “target”), but also about the imitator’s usual speech (the “matrix”), acknowledging that other varieties may also be influential in how the imitation is realized. 600 subjects participated in an internet survey in which they imitated dialect speech, and the researchers quantified imitation features in an effort to understand what is imitated.

In “Spontaneous dubbing as a tool for eliciting linguistic data: The case of second person plural inflections in Andalusian Spanish”, Victor Lara Bermejo introduces a new methodology for eliciting linguistic data, whereby informants dub short scenes shown on videos and accompanied by a description and a lead sentence designed to trigger specific syntactic structures. This technique appears particularly useful for eliciting linguistic features that prove too rare in traditional sociolinguistic interviews, while maintaining a level of spontaneity that is not compatible with pre-established questionnaires. The methodology is successfully applied to the case of the second person plural pronoun in Andalusian Spanish, which neutralizes the standard distinction between the formal *ustedes* and the informal *vosotros*. The Andalusian usage is shown to be doubly variable, in the choice of pronoun and in the agreement patterns of *ustedes* between 2nd and 3rd person. The standard variants appear to be spreading hierarchically, typically conditioned by age and educational background.

Ivana Škevin’s paper “Dialect levelling and changes in semiotic space” introduces Lotman’s (1985 [2015]) concept of semiotic space as an additional explanatory factor in dialect levelling. Drawing on fieldwork in Betina, Croatia, she shows that much of the traditional Romance-based vocabulary in Dalmatian dialects is being lost. In many cases, this is not due to accommodation to or influence from Standard Croatian, but simply because the concepts these lexical items

signify have lost importance in the speakers' daily lives. The effect of this change in semiotic space is similar to that of dialect levelling: the traditional dialect loses many of its salient characteristics.

A multi-method approach to the study of variation is presented in Ares Llop Naya's "The future of Catalan dialects' syntax: A case study for a methodological contribution". Llop combines a revision of existing linguistic studies on Catalan with data from speaker recordings, popular dialect literature, grammaticality judgments and even folk linguistics to arrive at a refined analysis of the constraints on the use of the dialectal negative marker *cap*. Although this work is in its early stages, it clearly shows that methodological innovation can also lie in the combination of existing methods.

Keiko Hirano investigated the use of Japanese vocabulary in the native English speech of English teachers in Japan for her paper "Code-switching in the Anglophone community in Japan". Her corpus of conversations between 39 native English-speaking teachers in the Fukuoka area contained over 1200 of such code-switches. Analysis of the data shows that the use of Japanese lexicon increases the longer a speaker has lived in Japan, and that it correlates positively with the strength of a speaker's social network with other English teachers, both native speakers and Japanese. Hirano suggests many code-switches involve group phraseology and proposes a community-of-practice explanation for this trend.

Two papers take advantage of the ultrasound tongue imaging technique and illustrate its relevance in dialectological studies. In "Tongue trajectories in North American English short-a tensing", Christopher Carignan, Jeff Mielke and Robin Dodsworth take a new look at the classic /æ/ variable. While the different regional realizations of /æ/ and their segmental conditioning are relatively well known, the phonetic motivations for the patterns observed remain unclear. The authors compare the articulatory trajectories of /æ/ before different coda consonants, with speakers from regions known to exhibit different patterns of /æ/ tensing. The results suggest in particular that different North American dialects have phonologized patterns of vowel-consonant coarticulation to different degrees. More generally, the authors emphasize the attractiveness of the ultrasound imaging technique in dialectology, due to its low cost and transportability.

Lorenzo Spreafico applies the same technique to another variable in "s/-retraction in Italian-Tyrolean bilingual speakers: A preliminary investigation using the ultrasound tongue imaging technique". The author investigates the of /s/ by Tyrolean speakers in the Italian region of South Tyrol, as opposed to the apical articulation characteristic of Italian. He compares tongue shapes during the production of /s/ in /sV/ vs. /sCV/ contexts in Italian and Tyrolean words by Italian-

dominant, Tyrolean-dominant and balanced bilingual speakers. Differences are observed across contexts, languages and speakers, suggesting that the articulation of /s/ is influenced by the degree of contact with Italian, even with minimal or no perceptual effects. Further studies are required to clarify the sociophonetic relevance of such results.

### 2.3 Japanese dialectology

Four papers report on current developments in Japanese dialectology. The first of these, “Developing the Linguistic Atlas of Japan Database and advancing analysis of geographical distributions of dialects” by Yasuo Kumagai details ongoing work on the digitization of the materials collected for the Linguistic Atlas of Japan between 1966 and 1974. Over half a million data cards are being digitized, including multiple responses, comments, and additional material that did not make it into the initial publication of the LAJ. Kumagai showcases some of the work that this updated material allows, such as investigating the geographical distributions of standard forms or the degree of linguistic similarity between locations; the emerging patterns are related to extralinguistic factors like transport networks.

Two papers make use of longitudinal data derived from a comparison of LAJ material with more recent linguistic surveys. In her paper “Tracing real and apparent time language changes by comparing linguistic maps”, Chitsuko Fukushima overlays linguistic maps from four surveys in the Niigata area to investigate diachronic change. The superimposition of the maps shows isoglosses moving in real time, with Western Japanese dialect forms first spreading to Niigata from Kyoto, and then retreating again. The maps also show transitional stages of changes in progress.

Takuichiro Onishi’s “Timespan comparison of dialectal distributions” investigates the wave theory of linguistic change by comparing LAJ data with two more recent surveys. He finds that, firstly, the spread of a change occurs in a rapid burst, rather than gradually and continually, and that secondly, dialect change need not spread from a central area to the periphery, but may also show an inverse pattern. These three papers on Japanese dialectology together show a wealth of data in the process of being unlocked for advanced analysis.

The final paper in this section is Ichiro Ota, Hitoshi Nikaido and Akira Utsugi’s “Tonal variation in Kagoshima Japanese and factors of language change”. The authors discuss the effect of various phonological and social factors in an ongoing change in the tonal system of Kagoshima Japanese (KJ). The traditional KJ system differs in important respects from that of Standard Japanese, both varieties



sharing a basic contrast between accented and unaccented words. The accented and unaccented patterns appear to be associated with different social meanings, as an asymmetry is observed between change toward the accented pattern of SJ and change toward the unaccented pattern, interpreted respectively as ‘de-dialectization’ and ‘de-standardization’. The paper also points to the role of mass media in language change (see Sayers 2014 and the ensuing debate).

## Acknowledgments

The conference was organized by Charlotte Gooskens, Nanna Haug Hilton, Bob de Jonge, John Nerbonne and Martijn Wieling, who were very capably assisted especially throughout the final months by Alexandra Nteliflippidi and Mara van der Ploeg. We were able to call on Joan Beal, the chair of the standing committee for the Methods conference series, whenever an element of tradition seemed obscure.

We received generous financial support from The Royal Netherlands Academy of Arts and Sciences; the University of Groningen in collaboration with the City of Groningen; the Netherlands Organization for Scientific Research; the Center for Language and Cognition, Groningen; CLARIN-NL, the Dutch branch of the European Common Language and Technology Infrastructure program, John Benjamins Publishing company, and Brill Publishers. The Alliance of Digital Humanities Organizations sponsored prizes for the two best posters by young scholars, and Cambridge University Press underwrote two Chambers prizes, which were awarded for the two best papers by young scholars.

We thank Oscar Strik, Groningen, and Sebastian Nordhoff, Language Science Press, for technical assistance with this first volume in the series *Language Variation*<sup>3</sup> of Language Science Press. We are especially grateful to the many colleagues who volunteered their time to review submissions, judge their quality with respect to publication, and provide authors with notes on how to improve their work. These were Birgit Alber, Will Barras, Charles Boberg, Miriam Bouzouita, Silvia Brandao, David Britain, Marc Brunelle, Paul de Dekker, Vittorio Dell’Aquila, Veronique De Tier, Jacob Eisenstein, Hans Goebel, Charlotte Gooskens, Jack Grieve, Lauren Hall-Lew, David Heap, Wilbert Heeringa, Steve Hewitt, Kris Heylen, Daniel Ezra Johnson, Roland Kehrein, Tyler Kendall, Brett Kessler, Nicolai Khakimov, Stefan Kleiner, Bill Kretzschmar, Haruo Kubozono, Laurence Labrone, Therese Leinonen, Andreas Lötscher, Andrea Mathussek, Jeff Mielke,

---

<sup>3</sup> <http://langsci-press.org/catalog/series/lv>

Simonetta Montemagni, Chris Montgomery, Naomi Nagy, Enrique Pato, Simon Pickl, Jelena Prokić, Simon Pröll, Christoph Purschke, Stefan Rabanus, Daniel Recasens, Lori Repetti, Andrés Salanova, Lea Schäfer, Oliver Schallert, Yves Scherrer, Koen Sebregts, James Stanford, Femke Swarte, Erik Tjong Kim Sang, Kristel Uiboed, Hans Van der Velde, Øystein A. Vangsnes, Helmut Weiß, Martijn Wieling and Heike Wiese. They have improved the volume immensely!

## References

- Auer, Peter. 2005. Europe's sociolinguistic unity, or: A typology of European dialect/standard constellations. In Nicole Delbecque, Johan van der Auwera & Dirk Geeraerts (eds.), *Perspectives on variation: Sociolinguistic, historical, comparative*, 7–42. Berlin / New York: Mouton de Gruyter.
- Auer, Peter & Frans Hinskens. 1996. The convergence and divergence of dialects in Europe. New and not so new developments in an old area. *Sociolinguistica* (1996) 10. 1–30.
- Auer, Peter, Frans Hinskens & Paul Kerswill (eds.). 2004. *Dialect change: Convergence and divergence in European languages*. Cambridge, New York: Cambridge University Press.
- Chambers, Jack & Peter Trudgill. 1980. *Dialectology*. Cambridge, New York: Cambridge University Press.
- Chambers, Jack & Peter Trudgill. 1998. *Dialectology*. 1st edn. Cambridge, New York: Cambridge University Press.
- Goebel, Hans. 1984. *Dialektometrische Studien: Anhand italo-romanischer, rätoromanischer und galloromanischer Sprachmaterialien aus AIS und ALF*. Tübingen: Niemeyer.
- Goebel, Hans. 2006. Recent advances in Salzburg dialectometry. *Literary and Linguistic Computing* 21(4). 411–435.
- Grieve, Jack, Dirk Speelman & Dirk Geeraerts. 2011. A statistical method for the identification and aggregation of regional linguistic variation. *Language Variation and Change* 23(2). 193–221.
- Hickey, Raymond (ed.). 2010. *The handbook of language contact*. Hoboken, NJ: Wiley.
- Kleiner, Stefan. 2014. Bericht. Methods in Dialectology XV 11.-15. August 2014, Groningen. *Zeitschrift für Dialektologie und Linguistik* 81(1). 61–66.
- Kretschmar, William A., Brendan A. Kretschmar & Irene M. Brockman. 2013. Scaled measurement of geographic and social speech data. *LLC: Journal of Digital Scholarship in the Humanities* 28(1). 173–187.

- Lenz, Alexandra N. 2003. *Struktur und Dynamik des Substandards: Eine Studie zum Westmitteldeutschen (Wittlich/Eifel)*. Stuttgart, Wiesbaden: F. Steiner.
- Lotman, Juri. 1985 [2015]. On the semiosphere. *Sign Systems Studies* 33(1). 215–239.
- Munske, Horst Haider & Andrea Mathussek. 2013. *Handbuch zum Sprachatlas von Mittelfranken: Dokumentation und Auswertung*. Heidelberg: Universitätsverlag Winter.
- Myers-Scotton, Carol. 1993. *Duelling languages: Grammatical structure in code-switching*. Oxford: Oxford University Press.
- Nerbonne, John & Peter Kleiweg. 2007. Toward a dialectological yardstick. *Journal of Quantitative Linguistics* 14(2). 148–166.
- Nerbonne, John, Rinke Colen, Charlotte Gooskens, Peter Kleiweg & Therese Leinonen. 2011. Gabmap – a web application for dialectology. *Dialectologia* Special Issue II. 65–89. <http://www.gabmap.nl/>.
- Nerbonne, John, Sandrien van Ommen, Charlotte Gooskens & Martijn Wieling. 2013. Measuring socially motivated pronunciation differences. In Lars Borin & Anju Sacena (eds.), *Approaches to measuring linguistic differences*, 107–140. Berlin: De Gruyter.
- Niedzielski, Nancy. 1999. The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology* 18(1). 62–85.
- Prokić, Jelena, Çağrı Çöltekin & John Nerbonne. 2012. Detecting shibboleths. In *Proceedings of the EACL 2012 Joint Workshop of LINGVIS & UNCLH*, 72–80. Association for Computational Linguistics.
- Purschke, Christoph. 2011. *Regionalsprache und Hörerurteil. Grundzüge einer perceptiven Variationslinguistik* (ZDL-Beihefte 149). Stuttgart: Franz Steiner.
- Sayers, Dave. 2014. The mediated innovation model: A framework for researching media influence in language change. *Journal of Sociolinguistics* 18(2). 185–212.
- Spruit, Marco René. 2008. *Quantitative perspectives on syntactic variation in Dutch dialects*. University of Amsterdam PhD thesis.
- Trudgill, Peter. 1986. *Dialects in contact*. Oxford: Blackwell.
- Trudgill, Peter. 2004. *New-dialect formation. the inevitability of colonial Englishes*. Edinburgh: Edinburgh University Press.
- Wieling, Martijn & John Nerbonne. 2011. Bipartite spectral graph partitioning for clustering dialect varieties and detecting their linguistic features. *Computer Speech & Language* 25(3). 700–715.

Wolk, Christoph. 2014. *Integrating aggregational and probabilistic approaches to language variation*. Freiburg: University of Freiburg PhD thesis. <https://www.freidok.uni-freiburg.de/data/9656/>.