Modern Journal of Language Teaching Methods (MJLTM)
ISSN: 2251-6204

Geography And Spatial Analysis In Diachronic Linguistics

.Krasikova E.N
.Lomteva T.N
.Kamensky M.V
.Kalashova A.S
.Chepurina I.V
.Nagamova N.V

North Caucasus Federal University,Russia, Stavropol, Pushkin Str.,1

ARTICLE INFORMATION
Original Research Paper
Received May. 2018
Accepted July. 2018

Keywords:
Historical
Linguistics
geography analysis,
spatial analysis
mapping
regional
studies
dialectology
dialectometry
geolinguistics
continental and subcontinental scales

ABSTRACT

The article deals with spatial patterns associated with language contact, language spread and isolation. They have been included in traditional theories of historical linguistics. The development of mapping and spatial analysis tools, combined with innovations in quantitative approaches to diachronic linguistics, has introduced a new era in linguistic geography. Geographic research in historical linguistics, however, is largely carried out in the separate traditions of various linguistic subfields. Many commonalities exist between the questions asked and the methods applied in each of these subfields, and increased interaction across dialectometry, phylogenetics and areal linguistics has the potential to reinvigorate linguistic geography and accelerate progress on questions about geography and the spatial outcomes of language change and language expansion. The article summarizes existent methods of analysis of spatial and geographical language changes and gives a possibility to continue the research. At greater time depths, the same processes responsible for dialect variation lead eventually to sufficient diversification to create separate languages and subgroups. Though the literature in historical linguistics contains fewer explicit discussions of spatial patterns and geographic relationships than appear in dialectology, geographic contact and isolation are still important contributors to the processes that create deeper linguistic relationships such as families.

1. Introduction

Diachronic linguistics has traditionally focused on language change as a temporal phenomenon, but spatial analysis has begun to play an increasingly important role in understanding the processes and outcomes of language change. The interaction of time and space in the diversification of languages and diffusion of linguistic material has been a fundamental element of historical linguistic theory, yet spatial patterns and processes have played only a peripheral role in most historical linguistic research until the mid-to-late 20th century. Geographic perspectives on diachronic linguistics have made more frequent and meaningful appearances in research on historical linguistic topics, however, as computational tools for map visualization and spatial data analysis have

become more accessible.

The field of linguistic geography, in spite of the interdisciplinary overtones of its name, has been practiced primarily by linguists, with limited interaction by geographers. Whereas linguists are interested in the internal workings of language systems, geographers tend to treat language as a unitary variable and have not engaged significantly in research on its internal complexity. This state of affairs was lamented by Wagner [Wagner 1958: 87] and remained unchanged throughout the remainder of the 20th century in spite of his agitation for more geographer involvement in the sorts of questions that interest historical linguists. Even among linguists, however, linguistic geography is a fractured field. Spatial patterns and geographic relationships are typically secondary to linguists' other research aims such as diagnosing linguistic relationships, modeling the dynamics of change, or describing structural diversity among the world's language. As a result, geography is relegated to the periphery of various linguistic subfields more often than it is treated as a unified topic in diachronic linguistics.

Methodology

The geographic questions asked in various subfields of linguistic might appear on the surface to be quite divergent. Yet the geographic component of most of these areas of research focuses on the detection and analysis of "areal signals". In most diachronic linguistic research, these signals present as a tendency for spatially near values or language varieties to be more similar than spatially distant ones. Geographers refer to this basic pattern, described in Tobler's First Law of Geography [Tobler 1970: 236], as spatial autocorrelation. Spatial autocorrelation can arise through the geographic spread or diffusion of linguistic material across lects, through the splitting of proto-languages into geographically proximal daughter languages, or through the reinforcement of shared retentions as a result of regular contact and communication. Though the relative importance of diffusion, divergence, and accommodation or convergence may vary in different linguistic studies, the nature of spatial organization in linguistic data and these basic mechanisms through which it arises are relevant to linguistic relationships at multiple historical scales. The identification and interpretation of such spatial patterns serves as a point of commonality among the various divisions of linguistic geography.

The notion of scale is an important element in the detection and analysis of geographic patterns. In linguistics, the spatial scales of patterns typically correlate with the time depths of the associated phenomena. Micro-variation among dialects occurs on a regional scale and reflects a recent stratum of history, while global patterns in language distribution reflect much deeper histories of language spread and population dynamics. The subfields of diachronic linguistics in which geography plays a role fall naturally into hierarchy of spatio-temporal scales. The following study reviews the integration of geographic perspectives and methods into historical linguistic research at several of these scales. First, we discuss research into the patterns of dialect diversity that occur on a regional scale. Then, we look at deeper linguistic changes associated with language families and linguistic areas and surveys the integration of geographic analysis into studies of phylogenetic relations and language spread. Next, we discuss research on global patterns in language geography. While much of the work examined in this section is focused on the analysis of synchronic patterns, the generalizations about language geography that can be developed through the examination of global patterns are useful for modeling language change. Finally, we discuss some methodological and theoretical aspects of linguistic geography that unites research at all these scales and concludes with remarks on the unification of linguistic geography and the advancement of spatial analysis in historical linguistics.

In no other area of linguistics has geography figured as prominently and enduringly as in the study of dialect diversity. The patterns of variation that emerge within languages at the dialect scale are typically limited

in both their spatial extents and in the time depth of the associated changes. By surveying older speakers, dialectologists often further minimize the temporal distance between patterns of variation in synchronic data and the processes of innovation and diffusion responsible for them. The typically fine spatial resolution of dialect data and the focus of this subfield on language changes nearest the surface of linguistic stratigraphy lend themselves to detail in the study of language change. This in turn facilitates the examination of theories about how elements of the social and physical landscape may condition the progression of changes within and among communities. Though dialectology is often treated as an independent subfield of linguistics, methodological and theoretical developments in this field are quite relevant to historical linguistics more generally.

Discussion

Traditionally, the geographic component of dialect geography involved plotting data from questionnaires on the map pages of linguistic atlases and then drawing isoglosses between surveyed locations to represent the approximate boundaries between competing linguistic forms. By overlaying isoglosses representing many dialect features on the map, the dialect geographer gauges not only where these isoglosses happen to coincide but also how dense or diffuse a particular bundle of isoglosses is. This qualitative procedure for evaluating isogloss patterns was the standard method for analyzing dialect variation and diagnosing dialect area boundaries through the late 19th and early-to-mid 20th centuries, but later scholars [Schneider 1989: 132] were troubled by the subjectivity of this method. Dialectology's emphasis on lexical relics also earned the field a reputation for antiquarianism and created a schism between atlas-oriented dialectologists and the more general field of linguistics [Campbell 1995: 187].

During the mid-20th century, the prevailing perspectives and methods in dialect geography shifted dramatically in response to the rise of variationist sociolinguistics and the influence of human geography, which in that era sought to understand spatial motivations for patterns in human behavior and culture. Though it included among its several social explanatory factors only one very basic geographic division, Labov's [Labov 2007: 138] study of sound change on Martha's Vineyard altered scholarly perspectives on language variation by demonstrating a quantitative approach for studying sources of variation in speech patterns and highlighting the social workings of language change. With this new tradition developing in sociolinguistics and models of spatial diffusion being created by human geographers, dialectology took a turn toward explaining spatial variation in language [Trudgill 1983: 98], rather than simply visualizing and describing dialectal boundaries.

Geolinguistics, as this branch of dialectology is known, developed an explicit focus on measuring spatial variation in linguistic data and probing the nature of spatial dialect patterns. Whereas isoglosses occasionally defy the principles of traditional dialect geography by intersecting in unexpected configurations instead of falling into tidy bundles, the geolinguistic approach applies less starkly categorical modes of analysis to investigate general geographic patterns in dialect data [Bailey 1993: 364]. In many cases, geolinguistic studies investigate complexities in spatial variation that are not well suited to isogloss mapping.

The geolinguistic arm of variationist linguistics has questioned many of the assumptions that underlie traditional dialectology. Fundamentally, the emphasis has shifted from treating dialects as basic systems to understanding the histories of individual features and collections of features. The concept of boundaries between dialect areas has been thoroughly questioned, both in terms of the data interpretation processes through which they are diagnosed [Ormeling 1992: 59] and their suitability for representing the actual variation that separates lects. Examination of the process of geographic diffusion has also evolved beyond a comparison of basic wave and gravity models, incorporating the additional ideas of environmental and social barriers and amplifiers to

diffusion [Bailey 1993: 376].

Current progress in dialect geography owes much to the branch of geolinguistics called dialectometry. This area of research seeks to infer patterns of variation from large datasets by analyzing features quantitatively and in aggregate, rather than focusing on limited datasets that describe individual linguistic characteristics [Nerbonne 2013: 229]. The establishment of dialectometry work in this area has progressed steadily since that time [Nerbonne 2013: 234]. However, advances in the projection of dialect variation onto geographic space and quantitative analysis of spatial patterns in dialect data have accelerated during the 21st century, presumably in response to technological innovations.

Because dialectometry typically quantifies linguistic relationships in terms of the aggregate difference between pairs of varieties, the comparison of linguistic and geographic distance matrices is widely used to assess spatial patterns in dialect variation. A standard statistic for comparing pairwise linguistic distances to pairwise geographic distances is the Mantel test. This procedure correlates distances (using familiar statistics such as Pearson's product moment correlation) and implements a permutation test to correct for the non-independence of distance measures. Mantel tests that compare language to geographic distance essentially measure the strength and significance of spatial autocorrelation in linguistic data. Other methods for comparing distance matrices are occasionally used, such as the PERMANOVA model used by Szmrecsanyi [Szmrecsanyi 2011: 58] to conduct analysis of variance with both spatial distance and dialect area membership as model parameters.

Linguistic distance, in these models, can be characterized in a number of ways. A common metric for expressing linguistic dissimilarities between pairs of locations is Levenshtein distance [Nerbonne 2013: 59]. Essentially a string-edit algorithm, Levenshtein distance quantifies the number of changes required to transform one linguistic form to another. Basic variations of this metric count insertions, deletions, and replacements of sounds, while more complex implementations can involve phonetic differences and weighting schemes [Heeringa 2004: 357].

The most basic representation of the spatial distance between lects is Euclidean distance or the Pythagorean calculation of the length of a straight line connecting two pairs of latitude/longitude coordinates. Meta-analysis by Nerbonne estimates that this simple characterization of geography accounts for 16% to 37% of the linguistic variation in dialect studies. This constitutes substantial support for the idea that geography plays a crucial role in language change and the genesis of linguistic diversity, but leaves much to be explained. Case studies presented by Nerbonne and Szmrecsanyi improve upon the prediction of linguistic variation by adding dialect area membership as an additional model parameter, following Shackleton's suggestion that a dialect area analysis may be more appropriate for the English study areas than a continuum model. Other measures of geographic distance adjust for barriers and conduits of contact or cultural diffusion model parameters [Nerbonne 2013: 234].

While models that compare basic distance measures like Euclidean distance to linguistic differentiation can be considered implementations of a basic wave model of diffusion, spatial distance can also be weighted according to the size of populations to implement a gravity model of linguistic diffusion. Trudgill's formulation of the gravity model is tested by Groningen dialectometrists [Nerbonne 2013: 232], but it is not found to outperform basic wave-like diffusion for the prediction of variation in Dutch. Better support for the gravity model is found in Szmrecsanyi's study of morphosyntactic variation in British English.

Other models of geographic distance reflect environmental and infrastructure constraints on contact. In addition to testing the gravity model, Szmrecsanyi compares linguistic variation to Google Maps travel time between locations [Szmrecsanyi 2011: 68]. The use of travel time in modeling dialect differentiation directly

captures the role of distance in mediating social and linguistic contact, yet this measure of contemporary contact turns out to be a poor predictor of British dialect variation. More complex models of the geographic and social context of language variation are made possible by the generalized additive models and mixed effects models employed by Weiling for Dutch dialect variation. While spatial distance is, as expected, the primary predictor of Dutch variation, demographic factors such as community size and average age also impact dialect differences. The use of fixed effects and random effects in this type of model allows for nuanced conclusions about the dynamics of language change – e.g. effects of frequency and part of speech – and the interaction of these effects with geographic and social factors. The sociolinguistic orientation of this model is useful where demographic data exists and social factors can be evaluated in light of political and historical facts. Additionally, the flexibility of the generalized additive model implemented in this paper for examining spatial variation in pronunciation allows for the characterization of non-linear relationships between linguistic and geographic distances. For languages in particularly remote or rugged territories and those for which demographic and social history is not well documented, environmentally specified models such as the cost-distance models may be more applicable.

The practice of identifying dialect areas and boundaries has also been updated with new, quantitative methodologies. Multidimensional scaling (MDS) is used to reduce the dimensionality of datasets and represent linguistic distances in a format that can be compared to geographical maps. Hierarchical clustering is also used to identify basic divisions between groups of varieties. Several methods have been developed to overcome issues with instability in clustering results, including bootstrapping methods borrowed from biology and new noisy clustering techniques [Nerbonne 2013: 231].

Dialectometry has produced many quantitative tools for investigating geographic patterns and processes in language variation and change, and several mapping techniques have also emerged as standard practices in this area of linguistics. At the most basic level, network maps represent the strength of pairwise linguistic distances with lines of varying levels of darkness, which connect the relevant locations. Continuum-like language variation can be represented by choropleth maps that represent MDS linguistic distances with red-green-blue color values [Nerbonne 2013: 229]. Dialect area maps represent the dialect boundaries identified by clustering by drawing lines between neighboring locations and using the weight or darkness of these boundary lines to represent the separation between pairs of neighboring locations.

Geographic proximity has been both explicitly and implicitly considered by historical linguists in reconstructing historical relationships, notably by Campbell's [Campbell 1995: 197] principle that "neighboring languages often turn out to be related".

Conflict

The quantitative methods employed in linguistic phylogenetics to infer historical relationships and investigate processes of evolution, like dialectometric methods, rely on similarities between languages. Although traditional reconstruction of language phylogenies relied on systematic correspondences in linguistic material, the extent to which the similarities used in modern computational phylogenetic methods involve systematic correspondence varies according to the selection of characters and coding of linguistic data in such studies. Generally, though, similarities between languages are generated by two general mechanisms (leaving aside, for the moment, the impacts of universal cognitive constraints and chance). Common genealogical inheritance, or vertical transmission, leads to similarities such as the correspondences identified by the traditional comparative method. In opposition to this is horizontal transmission, or the borrowing of material from one language to another, often through contact facilitated by geographic proximity. Detailed study by Labov [Labov 2007: 368]

associates these two fundamental mechanisms of language evolution with the difference between child language acquisition and adult language learning, suggesting that they should lead to different outcomes. However, recent debate among historical linguists has revolved around the relative contributions of these two mechanisms to the relationships found among the world's languages.

A strong position in this debate is represented by Dixon's [Dixon 2002: 245] punctuated equilibrium model. This theory places the majority of language evolution in long periods of relative sociohistorical stability, during which areal diffusion of linguistic material is the primary mechanism for generating linguistic diversity. In this model, the cladistics splits that generate family tree structures are exceptional, occurring only in association with infrequent historical events that significantly disrupt the social or geographic state of affairs. The contrast between Dixon's treatment of Australia as a linguistic area and the responses to his punctuated equilibrium and Australian areality proposals illustrates the debate regarding the extent to which horizontal transmission influences linguistic history. Consensus in the field supports the continued use of the family tree model, which may indeed interact with spatial diffusion of linguistic material, though to a more limited extent than suggested by Dixon [Dixon 2002: 256]. Spatial analysis has been applied on an even more limited basis in evaluating the relative roles of vertical and horizontal transmission in evolutionary studies; spatial methods are typically only employed in post-hoc tests for reality, if at all.

One prominent debate about reality and genealogy demonstrates this use of spatial methods to aid in the interpretation of phylogenetic characterizations of the relationships among Island Melanesia languages. This debate centers on the application of Bayesian phylogenetic methods by Dunn to structural feature data in order to characterize the relationships between the languages of this archipelago. The dialogue launched by Donohue and Musgrave's response to this work questions the nature of identified relationships between non-Austronesian languages in the study, suggesting that the signal identified by phylogenetic inference may in fact reflect areal diffusion rather than genealogical descent. Subsequent testing for spatial autocorrelation in the strengths of linguistic relationships provided further evidence to bear on this matter. These spatial autocorrelation tests took the form of linear correlations between linguistic distance (characterized as normalized Hamming distance) and spatial distance for pairs of languages. Dunn shows spatial distance to be a significant and reasonably strong predictor of linguistic relationships between pairs of Austronesian languages and pairs of non-Austronesian languages, but a far weaker predictor for mixed pairs. Given the commingled geographic distribution of these two groups of languages across the archipelago, these results suggest that the spatial signals identified within these groups are associated with processes of genealogical diversification. Weak overall spatial autocorrelation, particularly for mixed pairs, further suggests that convergence through areal diffusion of structural features is of limited explanatory value here. Donohue [Donohue 2013] reject this interpretation of the spatial pattern, construing the spatial signal instead as direct evidence of areal diffusion. The compatibility of spatial autocorrelation with both phylogenetic and areal explanations makes debates such as this particularly unsatisfying, as the disagreement lies not in the detection of a spatial signal but the interpretation of that signal.

Both phylogenetic and areal approaches to language evolution would benefit from further development of methodologies for analyzing spatial patterns in linguistic data. In particular, procedures for detecting spatial autocorrelation in residual variation not accounted for by phylogeny would aid in distinguishing geographic signatures of diversification from the effects of areal diffusion. For families that are well described and not too temporally deep for phylogenetic signals to surface in lexical data, the comparison of patterns in structural and lexical data could be useful for estimating the proportions of spatial signal that derive from genealogy and areal diffusion. In other cases, the first step toward a nuanced understanding of these language dynamics may be better

undertaken through spatial analysis of the distributions of individual linguistic features. For example, where deep genealogies are insecure, areal and genealogical influences on feature distribution could be studied with techniques like the neighbor graph procedure in Towner. Whereas linguistic and spatial distances can be compared with basic statistical procedures like Mantel correlations, other spatial statistics, such as join count statistics for binary data or Moran's I for continuous variables, are appropriate for identifying spatial autocorrelation in features associated with individual locations rather than pairs of locations.

Studies of language family expansion have led to a more natural integration of the temporal and spatial components of language evolution, and a more rapid development of geographically sophisticated methodologies. Interdisciplinary interest in prehistoric population expansions has been growing, with recent proposals such as the language-farming dispersal hypothesis receiving considerable attention. Efforts to answer questions about the timing and geography of population dispersals have resulted in a growing body of phylogenetic research that bears on questions of linguistic geography. Gray, for example, test hypotheses regarding the population of the Pacific by using Bayesian methods to construct phylogenetic trees and evaluating theories against the topology, dates, and geography associated with their phylogeny. Comparing patterns of cladogenesis (splitting of branches) and anagenesis (lengthening of branches) evident in the phylogeny with geography, they are able to link patterns of genealogical diversification evident in their reconstructions with a pulse-pause theory of Pacific settlement. Similar methods have been used to provide evidence for an Anatolian origin for Indo-European [Gray 2003: 482]; however, the extension of this linguistic technique for assessing population geographies to other parts of the world is limited largely by data availability.

Some of the most recent advances that fall under the umbrella of linguistic geography also come from studies of linguistic phylogenetics. Bayesian models for phylogeographic inference employed by Bouckaert for Indo-European and Currie for Bantu simultaneously model the spatial, temporal, and genealogical outcomes of language family expansion. Bouckaert [Bouckaert 2012: 959] adapted a viral epidemiology model to infer from lexical cognate data the ancestral nodes of the Indo-European family tree and the locations associated with these nodes. Latitude and longitude are modeled as evolving along phylogenetic tree branches according to a relaxed random walk model of isotropic diffusion. The geographic component of the model differentiates between land and water, but otherwise treats the Eurasian continent as a uniform surface. This methodology is an important advance in the use of geographic information in modeling linguistic prehistory, yet its essentially isotropic simulation of geographic diffusion and its failure to consider variation in settlement density and migration rates leave room for the development of more geographically sophisticated methods. Such refinements of the geographic component of this model require a complex understanding of human-environment relations and may not be feasible without further empirical study of global patterns in language diversity and the environmental variables that favor it.

Conclusion

Geographic Information System (GIS) technology has also made mapping easier for linguists. Basic GIS functions like the coordination of spatial datasets and the visualization of linguistic information in map form are broadly useful for illustrating historical facts and phenomena. However, these tools will become even more useful with the development of larger, open databases of language location information. However, the integration of mapping and geographic analysis in linguistic research, as represented by this work, still lags behind its potential. The availability of better and more complete language map information and the advancement of cartographic standards in linguistics will facilitate innovation in geographically oriented approaches to language change.

References

1. Bailey, Tom Wikle, and Lori Sand. 1993Some patterns of linguistic diffusion. Language variation and change 5. Pp. 359–390.
2. Bednarek Monika 2013. Integrating visual analysis into corpus linguistic research. Corpus linguistics 2013. Pp. 37–39.
3. Bouckaert, Remco, Philippe Lemey, Michael Dunn, Simon J. Greenhill, Alexander V. Alekseyenko, Alexei J. Drummond, Russell D. Gray, Marc A. 2012 Suchard and Quentin D. Atkinson. Mapping the origins and expansion of the Indo-European language family. Science 337. Pp. 957–960.
4. Campbell, Lyle 1995. The Quechumaran hypothesis and lessons for distant genetic comparison. Diachronica 12. Pp. 157–200.
5. Dixon, R.M.W. 2002. Australian languages: their nature and development. Cambridge: Cambridge University Press. Pp. 245–256.
6. Donohue, Mark & Musgrave, Simon & Whiting, Bronwen & Wichmann, Soren 2011. Typological feature analysis models linguistic geography. Language. 87. 369-383.
7. Gooskens, Charlotte 2004. Norwegian dialect differences geographically explained. Papers from the Second International Conference on Language Variation in Europe, ed. by B. L. Gunnarson, L. Bergstrom, G. Eklund, S. Fridella, L. H. Hansen, A. Karstadt, B. Nordberg, E. Sundgren and M. Thelander. Uppsala, Sweden: Uppsala University. Pp. 195–206.
8. Gray, Russell D., A.J. Drummund and Simon J. Greenhill 2009. Language phylogenies reveal expansion pulses and pauses in Pacific settlement. Science 323. Pp. 479–483.
9. Haynie, Hannah 2014. Geography and Spatial Analysis in Historical Linguistics. Language and Linguistics Compass.
10. Heeringa, Wilbert 2004. Measuring dialect pronunciation differences using Levenshtein distance. Ph.D. thesis. Rijksuniversiteit Groningen.
11. Heeringa, Wilbert 2001. Dialect areas and dialect continua. Language Variation and Change. 13. 375 - 400.
12. Labov, William 2007. Transmission and diffusion. Language 83. Pp. 344–387.
13. Maha N. Alharthi 2013 A corpus-based study for assessing the collocational competence in learner production across proficiency levels. Corpus linguistics 2013. Pp. 9 –13.
14. Nerbonne, John 2013. How much does geography influence language variation? Space in language and linguistics: geographical, interactional, and cognitive perspectives, ed. by Peter Auer, Martin Hilpert, Anja Stukenbrock and Benedikt Szmrecsanyi. Berlin: De Gruyter. Pp. 220–236.
15. Ormeling, Ferjan 1992. Methods and possibilities for mapping by onomasticians. Discussion Papers in Geolinguistics 19–21. Pp. 50–67.
16. Schneider, Edgar W., and William A. Kretzschmar. Jr. 1989. LAMSAS goes SASsy: statistical methods and linguistic atlas data. Journal of English Linguistics 22. Pp. 129–141.
17. Szmrecsanyi, Benedikt 2011. Corpus-based dialectometry: a methodological sketch. Corpora 6. Pp. 45–76.
18. Tobler, Waldo R. 1970. A computer movie simulating urban growth in the Detroit region. Economic Geography 46. Pp. 234–240.
19. Trudgill, Peter 1983. On dialect: social and geographical perspectives. New York: New York University Press. P. 98.
20. Wagner, Philip L. 1958. Remarks on the geography of language. Geographical Review 48. Pp. 86–97.