

In search of patterns of historical language variation and user interaction (or: Who used what linguistic features with whom, when, where, why, and how?)

Mark Richard Lauersdorf

18th Annual Meeting of the Slavic Linguistics Society (SLS-18)
24-26 August 2023 — Bratislava, Slovakia

The field of historical sociolinguistics studies the interaction of language and society in historical periods and from historical perspectives in order to elucidate the social factors at play in language variation and change over time. The traditional research paradigm in the investigation of language history has long been based on the equation “language history = standard language history = history of national identity”, but in recent decades it has been recognized that to truly understand the course of historical language change, one must investigate all varieties of a language in play in a given socio-cultural context (not just a superposed “standard variety”) and one must take a multifaceted approach to questions of language and identity (not just a superposed “national identity”) (cf. Milroy 2001; Crowley 2003; Elspaß 2020; Ayres-Bennett and Bellamy 2021). Taking this stance to its most granular interpretation, it becomes necessary to explore in full detail “who used what linguistic features with whom, when, where, why, and how” (paraphrasing Fishman 1965) in order to fully understand the social history of language varieties and their users.

The research paradigm proposed in this talk seeks to advance the field in this direction by investigating historical language variation and change in all its breadth and in its intimate connection to language users and their interactions. **To do so, it is essential to adopt methods that pursue socio-historical analysis of language variation and change through data-driven/inductive pattern identification in large datasets encompassing both linguistic and socio-cultural data.** This methodological approach can be summarized as follows:

Data-driven corpus-based sociolinguistic investigation of historical language variation and user interaction in their complex socio-political and socio-cultural environments, using statistical and visualization methods of data analysis to identify and correlate salient patterns in the linguistic and socio-historical data.

In order to pursue a historical sociolinguistic research agenda of discovering “who used what linguistic features with whom, when, where, why, and how”, deploying a data-driven/inductive methodological approach, I call for an overarching *modus operandi* that “Use[s] all the data!”. This “Use all the data!” approach involves the following principles (Lauersdorf 2018b: 211–212):

- (1) Identify all possible sources of language data – data may be “hiding” where you don’t expect it, in unexplored physical locations and in unexplored textual locations.
- (2) Consult the entirety of the language data available to you – avoid selective sampling (inclusion or exclusion) of language data on the basis of *a priori* notions of what kind of data you need, how much data you need, where it should come from, etc.
- (3) Language data isn’t the only data – use all the socio-historical data!
 - (a) Identify and use all possible sources of socio-historical data (again being on the lookout for socio-historical data “hidden” in unexpected places and using the entirety of the socio-historical data available to you).
 - (b) We only have the language data that history leaves us (what has “survived” through time), so wrap the language data in all possible socio-historical datasets to help complete the picture.

These principles also include a corollary set of guidelines, formulated originally in regard to data visualization (Lauersdorf 2018a: 112), but proposed here as necessary for all data analysis in historical sociolinguistics:

- Corollary 1: If you use all the data, view all the data.
- (a) If you view all the data, view all the combinations.
 - (b) If you view all the data, view all the angles.
 - (c) If you view all the data, use all the techniques.

This call to “Use all the data!” in historical sociolinguistic investigation derives, in part, from the fact that historical data is inherently “incomplete” (e.g., limited, fragmentary, unbalanced) in ways that the researcher has no control over. Thus it is imperative to identify and gather as much of it as possible for a given investigation, from all interrelated sources, both linguistic and socio-historical, and to consider especially nontraditional data

sources (Lauersdorf 2021: 218-219), if one hopes to be able to assemble a sufficient dataset for data-driven/inductive analysis. If one follows the principle of gathering “all the data” in data collection, the extant historical record can often produce richly layered datasets containing linguistic features of language users in their socio-cultural interactional contexts.

Additionally, this call to “Use all the data!” and “View all the data!” also derives from the conviction that using only a selective sample and/or selective methods of analysis of the available data (generally based on *a priori* assumptions about the features and categories that one should expect to find in the data) limits what one is actually able to discover, given that portions of the data are not being considered and that only certain analytical viewpoints are being entertained. Thus it is imperative to assemble and interrogate the data in a way that facilitates data-driven examination of all possible combinations of all linguistic and socio-cultural information contained in the rich data layers.

From a methodological perspective, this approach clearly implies a scale of operations that requires the assistance of *computational processing*, and *statistical and visualization methods of data analysis*. It has become commonplace in historical sociolinguistics to work with digital corpora of historical source texts, and there is also a growing awareness of the need for digital preparation of the socio-historical data in a way that it can be analyzed interactively with the textual corpora. However, with the approach espoused here, these digital datasets, both linguistic and socio-cultural, grow in size and sophistication; and the computational complexity involved in using and viewing “all the data” increases significantly, benefitting from “big-data” methods, models, and tools to explore the correlations between language users and linguistic features (Lauersdorf 2021: 219).

From a theoretical perspective, this approach allows us to engage fully with notions and models previously unavailable (or only limitedly available), enabling exploration of historical language variation and change within the theoretical frames of mundane mobility, social networks, and communities of practice. *Mundane mobility*, the “mundane movements we engage in in everyday life” (Britain 2013: 165), is “small-scale, less dramatic in distance, and perhaps in life impact at the level of the individual, [but] their scale, intensity and pervasiveness at the level of the community as a whole mean they cannot be ignored as a source of rather striking dialect contact” (Britain 2013: 168). The contact patterns formed by such mobility can be modeled as *social networks* that examine the interactions of individuals in their communities and in their networks of contacts as potential determinants in historical language variation and change (cf. Bergs 2005; Conde-Silvestre 2012). And “[t]he *community of practice* takes us away from the community defined by a location or by a population. Instead, it focuses on a community defined by social engagement – after all, it is this engagement that language serves, not the place and not the people as a bunch of individuals” (Eckert and McConnell-Ginet 1992: 95; emphasis mine).

Data-driven use-all-the-data exploration of the correlation of mundane mobility, social networks, and communities of practice with language variation and change among language users in their interactions in historical periods is expressly the investigation of “who uses what linguistic features with whom, when, where, why, and how”, a research paradigm that holds the potential to paint a more detailed picture of the social history of language varieties and their users.

References:

- Ayres-Bennett, Wendy and John Bellamy, eds. 2021. *The Cambridge Handbook of Language Standardization*. Cambridge: Cambridge University Press.
- Bergs, Alexander. 2005. *Social Networks and Historical Sociolinguistics: Studies in Morphosyntactic Variation in the Paston letters (1421–1503)*. Berlin and New York: Mouton de Gruyter.
- Britain, David. 2013. *The Role of Mundane Mobility and Contact in Dialect Death and Dialect Birth*. In: Daniel Schreier and Marianne Hundt, eds. *English as a Contact Language*. Cambridge: Cambridge University Press, 165-181.
- Conde-Silvestre, Juan Camilo. 2012. The Role of Social Networks and Mobility in Diachronic Sociolinguistics. In: Juan Manuel Hernández-Campoy and Juan Camilo Conde-Silvestre, eds. *The Handbook of Historical Sociolinguistics*. West Sussex: Wiley-Blackwell, 80–98.
- Crowley, Tony. 2003. *Standard English and the Politics of Language*. London: Palgrave Macmillan.
- Eckert, Penelope and Sally McConnell-Ginet. 1992. *Communities of Practice: Where Language, Gender, and Power All Live*. In: Kira Hall, Mary Bucholtz, and Birch Moonwomon, eds. *Locating Power, Proceedings of the 1992 Berkeley Women and Language Conference*. Berkeley: Berkeley Women and Language Group, 89-

- Elspaß, Stephan. 2020. Alternative Sources of Data for Alternative Histories of Standardisation. *Language Policy* 19(2): 281–299. (Special issue: *Revisiting Haugen. Alternative histories of standardization*, Gijbert Rutten and Rik Vosters, eds.).
- Fishman, Joshua A. 1965. Who speaks what language to whom and when? *La linguistique* 1(2): 67–88.
- Lauersdorf, Mark Richard. 2018a. Linguistic Visualizations as *objets d'art*? In: Noah Bubenhofer and Marc Kupietz, eds. *Visualisierung sprachlicher Daten*. Heidelberg: Heidelberg University Publishing, 91–122.
- Lauersdorf, Mark Richard. 2018b. Historical (Standard) Language Development and the Writing of Historical Identities: A Plaidoyer for a Data-Driven Approach to the Investigation of the Sociolinguistic History of (not only) Slovak. In: Stephen M. Dickey and Mark Richard Lauersdorf, eds. *V zeleni drželi zeleni breg: Studies in honor of Marc L. Greenberg*. Bloomington, Indiana: Slavica Publishers (appeared 2019), 199–218.
- Lauersdorf, Mark Richard. 2021. Historical Sociolinguistics and the Necessity of Interdisciplinary Collaboration. In: Allison Burkette and Tamara Warhol, eds. *Crossing Borders, Making Connections: Interdisciplinarity in Linguistics*. Berlin: De Gruyter Mouton, 207–230.
- Milroy, James. 2001. Language Ideologies and the Consequences of Standardization. *Journal of Sociolinguistics* 5(4): 530–55.