Texts as data sources for historical sociolinguistics

Joshua Bousquette, University of Georgia
Rob Howell, University of Wisconsin-Madison
Mark Richard Lauersdorf, University of Kentucky

Historical sociolinguists are largely dependent on textual data in their pursuit of sociolinguistic explanations of language variation and change in periods before the current generation of speakers. The use of such data sources requires an understanding of the nature of textual data – what it can tell us and what it can't. This, in turn, requires an understanding of the nature of writing practices and text types, and an understanding of the socio-historical context(s) in which the writing occurs.

Using a "presentation and open discussion" format, this workshop will consider the interpretation of textual data for sociolinguistic investigation along three lines:

• Texts as representations of language use (Rob Howell).

While we often find variation in textual data, the question immediately arises: "what does this variation tell us?" We will discuss issues with variation in textual data in the Early Modern period, an era of expanding literacy prior to codified written standards, and then move on to the more modern era, roughly 1700-1920, in which standardization can partly or totally suppress or mask variation present in spoken varieties. Questions of text type will be treated as well as issues arising in the interpretation of actual variation in textual data.

• Texts as sources of sociolinguistic context (Joshua Bousquette).

Historical sociolinguists, lacking access to living speakers in field observations, must reconstruct the social component of linguistic community by employing so-called extra-linguistic texts, e.g., government records (census, immigration, judicial, surveyor), church records (baptism, marriage, death), minutes of government and non-government institutions. We will discuss methodologies for mapping historical language contact in social domains, in geographic distribution, and across time, with emphasis on social network theory and community theory (Salmons 2005a, b; Warren 1963), and on social institutions, and the institution of the household, as a locus of language acquisition, a model of cross-generational change, and a dynamic linguistic community with the continuity of subject (Balibar and Wallerstein 2011; Wallerstein 2004).

• Texts as datasets for historical corpora (Mark Richard Lauersdorf).

Critical concepts in constructing linguistic corpora include "representativeness", "balance", "size" – does your corpus adequately represent the language variety(s) being investigated, in a way that is balanced in the language samples it includes across the variables being considered, including enough data to yield statistically viable results? (cf. Sinclair 2005). In a historical context, the overarching concern is that "[h]istorical documents survive by chance, not by design, and the selection that is available is the product of an unpredictable series of historical accidents" (Labov 1994). We will consider issues of representativeness, balance, and size in constructing textual corpora for sociolinguistic work within the limitations of historical textual data.

Central to these discussions of textual data will be the concept of "informational maximalism": "the utilization of all reasonable means to extend our knowledge of what might have been going on in the past...wish[ing] to gain a maximum of information from a maximum of potential sources" (Janda and Joseph 2003). We will consider both methodological concerns and practical implementation, with hands-on examination of sample data to illustrate the concepts.

Balibar, Étienne and Immanuel Wallerstein. 2011 [1991]. *Race, Nation, Class: Ambiguous Identities*. 2nd ed. London: Verso.

Janda, Richard D. and Brian D. Joseph. 2003. On Language, Change, and Language Change – Or, Of History, Linguistics, and Historical Linguistics. In Richard D. Janda and Brian D. Joseph, eds. *The Handbook of Historical Linguistics*. Oxford: Blackwell Publishing, 3-180.

Labov, William. 1994. *Principles of Linguistic Change. Volume 1: Internal Factors*. Oxford: Blackwell Publishing.

Salmons, Joseph C. 2005a. The Role of Community and Regional Structure in Language Shift. In Lothar Hönnighausen, Marc Frey, James Peacock & Niklaus Steiner, eds. *Regionalism in the Age of Globalism Volume 1: Concepts of Regionalism*. Madison: Center for the Study of Upper Midwestern Cultures, University of Wisconsin-Madison, 129-138.

Salmons, Joseph C. 2005b. Community, Region, and Language Shift in German-Speaking Wisconsin. In Lothar Hönnighausen, Anke Ortlepp, James Peacock & Niclaus Steiner, eds. *Regionalism in the Age of Globalism Volume 2: Forms of Regionalism*. Madison: Center for the Study of Upper Midwestern Cultures, University of Wisconsin-Madison, 133-144.

Sinclair, John. 2005. Corpus and Text – Basic Principles. Martin Wynne, ed. *Developing Linguistic Corpora: a Guide to Good Practice*. Oxford: Oxbow Books, 1-16. http://ota.ox.ac.uk/documents/creating/dlc/ Accessed 2017-05-31.

Wallerstein, Immanuel. 2004. *World Systems Analysis: An Introduction*. 5th ed. Durham, North Carolina: Duke University Press.

Warren, Roland. 1963. *The Community in America*. Chicago: Rand McNally & Company.