

## A Multilingual Corpus to Explore Geographic Variation

Naomi Nagy, University of Toronto  
CLAVIER IX Conference Talk

Given that over half of the world's population is multilingual from childhood (Tucker 1999), it's strange that in quantitative variationist studies the trend is decidedly to examine one language at a time, essentially treating speakers as monolingual (Nagy & Meyerhoff 2008)—although we recognize that significant exceptions exist (e.g. Poplack 1980, Poplack & Meechan 1998). Even in Toronto, touted as the “most multilingual city in the world,” two major projects examining ethnic effects on language focus exclusively on English (Tagliamonte 2007, Walker & Hoffman 2008). Complete understanding of how linguistic variation is used to construct identity requires examining multilingual speakers' full repertoires, building on what we have learned from studying speakers' monolingual facets and patterns of code-mixing. To remedy this, we have initiated the **Heritage Language Variation and Change Project** which complements English-focused corpus-development projects in Toronto by examining variation and inter-generational change in 7 (of the ~100) heritage languages spoken in the city: Cantonese, Faetar (an endangered Apulian Francoprovençal variety), Italian, Korean, Russian, Ukrainian, and Urdu.

This project addresses questions such as:

- Which features, structures, rules or constraints are cross-linguistically relevant to borrowing? Which are borrowed earlier and more often?
- Which social/demographic factors are cross-linguistically relevant to borrowing?
- Do the same (types of) speakers lead changes in HLs and in English? Is leadership in language change inherent, or do leaders choose to use one language for this social “work”?

The purpose of this paper is to describe our goals and the methods involved in constructing a large multilingual corpus for the purpose of understanding contact-induced language change. Specifically, we illustrate how our methodology ensures comparability and continuity across communities, languages, and fieldworkers in the data-collection stage and uses ELAN ([www.la-mpi.neu/tools/elan](http://www.la-mpi.neu/tools/elan)), a tool for creating and manipulating time-aligned tiered annotations that eliminates the need for narrow transcription. Our data collection strategy includes hiring fieldworkers from within each community, recording a Labovian-style (Labov 1984) sociolinguistic interview of 60-90 minutes during which participants describe their background, language experiences, demographic characteristics, and linguistic attitudes, and recount personal narratives; complete an Ethnic Orientation Questionnaire (Keffe & Padilla 1987); and complete a picture-identification task to ensure inter-speaker comparability for a set of lexical items.

We will illustrate these methods via description of some cross-generational patterns of change that have been noted in the Italian and Faetar components of our corpus. In our initial fieldwork season (summer 2009) we are conducting 40 interviews in each HL, including first-, second- and third-generation speakers, balanced for a variety of demographic factors. These speakers' variable patterns are compared to published descriptions of the relevant homeland varieties (e.g. Loporcarlo 1997; Nagy 2000; Trumper 1997) to illustrate geographic variation.

An additional goal of this presentation is to encourage further collaboration with scholars interested in using these data.

## References

- Keefe, S. & A. Padilla. 1987. *Chicano Ethnicity*. Albuquerque, NM: University of New Mexico Press.
- Labov, W. 1984. Field methods of the Project on Linguistic Change and Variation. In J. Baugh & J. Sherzer (eds). *Language in Use*. Englewood Cliffs: Prentice Hall. 28-54.
- Loporcarlo, M. 1997. Puglia and Salento. In M. Maiden & M. Parry, (eds). *The Dialects of Italy*. London & NY: Routledge. 338-54.
- Nagy, N. 2000. *Faetar*. Munich. Lincom Europa.
- Nagy, N. & M. Meyerhoff. 2008. The social life of sociolinguistics. In M. Meyerhoff & N. Nagy (eds). *Social Lives in Language: Sociolinguistics and Multilingual Speech Communities*. Amsterdam: John Benjamins.
- Trumper, J. 1997. Calabria and southern Basilicata. In M. Maiden & M. Parry, (eds). *The Dialects of Italy*. London & NY: Routledge. 355-75.
- Poplack, S. 1980. Sometimes I'll start a sentence in Spanish Y TERMINO EN ESPAÑOL: Toward a typology of code-switching. *Linguistics* 18.7/8: 581-618.
- Poplack, S. & M. Meechan, eds. 1998. *Instant loans, easy conditions: The productivity of bilingual borrowing*. Special issue of the *International Journal of Bilingualism* 2.2.
- Tagliamonte, S. 2007. Directions of Change in Canadian English.  
<http://individual.utoronto.ca/tagliamonte/directionsofchange.html>.
- Tucker, G.R. 1999. *A Global Perspective on Bilingualism and Bilingual Education*. Washington, D.C.: Center for Applied Linguistics.
- Walker, J. & M. Hoffman. 2008. Language Contact, Linguistic Variation and Ethnic Identity in Toronto English 2008. <http://www.yorku.ca/jamesw/projects.htm>.