

## *A Multilingual Corpus to Explore Variation in Language Contact Situations*

Naomi Nagy  
University of Toronto

### **Abstract**

*Numerous claims exist about types and ordering of linguistic change in language contact situations. I describe a project to develop a multilingual corpus to allow inter-generational, cross-linguistic, and diatopic comparison to determine what generalizations are possible about the types of variable features, structures or rules that are borrowed earlier and more often. It investigates how social factors (speaker status, network membership, language use, ethnic identity, and linguistic attitude at the individual level, and demographic characteristics and institutional support at the community level) relate to type and degree of language change. The corpus contains data from 40 speakers of each of six heritage languages (Cantonese, Faetar, Korean, Italian, Russian, Ukrainian) spoken in Toronto. Collaboration with experts in these languages is sought.<sup>1</sup>*

### **Introduction**

Given that over half of the world's population is multilingual from childhood (Tucker 1999), it's strange that in the field of variationist sociolinguistics, the trend is decidedly to examine one language at a time, essentially treating speakers as monolingual (Nagy & Meyerhoff 2008a). This contrasts with formal linguists' efforts to detail a Universal Grammar of rules/constraints/parameters that are cross-linguistically relevant. Even in Toronto, touted as the "most multilingual city in the world" (Berkowitz 2003; Toronto Public Library 2006), two major data collection projects within the variationist framework, focus exclusively on English (Tagliamonte 2008; Walker & Hoffman 2008).

Yet, to fully understand how language is used to construct identity, it is essential to examine speakers' full repertoires, and not treat them as monolingual entities. In order to address this, the **Heritage Language Variation and Change in Toronto Project (HLVC)** has been developed. Its goal is to develop a multilingual corpus of naturally-occurring speech in lesser-studied languages, that allows us to examine linguistic variation and change across languages, across locales (homeland vs. transplanted area), and across speakers (different ages, generations, ethnic orientations, etc.). This project complements two English-focused corpus-development projects in Toronto, the Toronto English Archive, or *TEA* (Tagliamonte 2008), and the Ethnicity and Language Project, or *ELP* (Walker & Hoffman 2008), focusing on several of the same speaker groups, by examining variation and inter-generational change in several heritage languages (HL) spoken in the city. This collection of high-quality recordings of naturally-occurring speech in six languages, is being digitally archived and is available to researchers interested in collaboration.

Of course, many studies have been conducted which report quantitatively and accountably on contact-induced language change. And specifically, numerous publications have described HL variation in Canada (*cf.* Budzhak-Jones 1994; Danesi 1985; Fortier 1991; Guardado 2002; Renaud *et al.* 2001; Vizmuller-Zocco 1993), yet little progress in our theories of how languages vary and evolve can be made due to disparate methodologies. The inconsistencies among collection and analysis methods, among contact situations, and among language sets compared, severely limits the possibilities to contribute to theoretical developments.

As set forth in Nagy (1996; 1997), there is a need to increase comparability across studies of different multilingual communities to gain greater understanding of contact-induced language change. To redress this situation, the project reported on here employs a consistent set of methods and examines contact induced language change in a fixed context (the Greater Toronto Area, or the GTA), while studying languages which differ along a number of continua, including their histories in the GTA, their typological distance from English, and the status of their community of speakers, both in the GTA and the homeland.

This cohesive program combines elements of code-switching theory which looks at when each language is used, but not at which forms of the language are selected (Myers-Scotton 1993a; 1993b; Poplack 1980; Sankoff & Poplack 1981) with the variationist approach, which quantifies the effects of various contextual forces on the selection of possible forms within one language (Tagliamonte 2006). This further testing of the Labovian or variationist paradigm with languages other than English will also help move beyond our field's large-language bias (Nagy & Meyerhoff 2008b).

It is, of course, more difficult to establish a valid description of the envelope of variation when more than one language is compared. Many scholars have grappled with this problem (*cf.*, Weinreich 1966; Myers-Scotton 1993a; 1993b; Mahootian 2006; Sanchez 2008). The collaborative HLVC project will make progress by developing a standardized framework, based on investigations of six distinct contact situations, to resolve questions of two types:

(1) Research questions

- LINGUISTIC: Are cross-linguistic generalizations possible about the types of features, structures, rules or constraints that are borrowed earlier and more often in this type of contact situation? If so, what do they include? (How) are social factors deterministic of the type and/or kind of language change?
- SOCIAL: Do the same (types of) speakers lead changes in their HL and in English? That is, is the propensity to be a "leader" in language change (Labov 2001) an inherent trait, or do speakers choose to use one language or the other for this sort of social "work"? The very choice of which language to use is important to consider in this sort of identity-marking, and cannot be left out of the equation.

This paper describes the methods employed in the construction of the HLVC corpus and illustrates how they are meant to lead us to better understanding of these issues.

### ***Heritage Languages***

Because the term "heritage language" has been used in a number of ways, I begin by defining how it is used in this paper. There are three non-overlapping categories of languages in Canada: 1) indigenous; 2) official (French and English); and 3) *heritage languages*, spoken by immigrant groups more recently arrived than the original colonisers (the French and British). Anyone who is a mother tongue speaker of a language identified with their heritage, other than French or British, is thus a HL speaker. I do *not* use the term "heritage language" in Polinsky & Kagan's (2007) sense: "Heritage speakers are people raised in a home where one language is spoken who subsequently switch to another dominant language," or with any implications of linguistic deficit. A primary goal of this project is, in fact, to learn about the paths by which speakers maintain or switch home languages. Because there are many HL speakers in Toronto, documentation of language usage patterns are important for planning and pedagogical purposes.

The HLs examined are shown in Table 1. "MT speakers" is the number of mother tongue speakers reported in Statistics Canada (2007). Information on ethnic origin is from Statistics Canada (2009). The "Established" date is when the first known church operating in each language was established in the GTA, as a readily-available indicator of the existence of a community of speakers (City of Toronto 2010; Gregorovic 1984; Handera 1984; Harney 1984; Kim 1984). "City/region of origin" shows the place we have targeted as the homeland of our first generation speakers, in order to control the amount of (homeland) regional variation being brought into the HL sample.

*Table 1: Heritage Languages examined (Numbers are approximate)*

| Language  | MT speakers | Ethnic Origin        | Established       | City/region of origin    |
|-----------|-------------|----------------------|-------------------|--------------------------|
| Cantonese | 170,000     | 537,000 <sup>2</sup> | 1951 <sup>3</sup> | Hong Kong                |
| Faetar    | <100        | <100                 | 1950              | Faeto, Celle di St. Vito |
| Korean    | 49,000      | 55,000               | 1967              | Seoul                    |
| Italian   | 194,000     | 466,000              | 1908              | Calabria                 |
| Russian   | 66,000      | 58,505               | 1916              | St. Petersburg, Moscow   |
| Ukrainian | 27,000      | 122,000              | 1913              | Lviv                     |

The framework of variationist sociolinguistics (*e.g.*, Labov 1972; Tagliamonte 2006) was developed for the analysis of English, and then extended to examine other, usually widely-spoken, languages. It is ripe for further expansion. To learn whether this approach produces similar results across languages, it is useful to examine other languages in a venue where the results are comparable. Thus, this study is intentionally comparable to Tagliamonte's 1.8 million word *TEA* and Walker & Hoffman's *ELP*. By matching samples and methods to these two English corpora, results will be cross-linguistically comparable within the community, and "whole-speaker" patterns of language use observable.

### **Targeted languages**

The six languages with which we begin the development of this corpus have been strategically selected for contrasting inherent traits, differing degrees of divergence from English, and available resources. For example, contrasting with Italian, for which there are vigorous long-standing communities in Toronto (but only 16% use of Italian at home by ethnic Italians, and only 42% are MT speakers), Korean represents a recent immigrant population where 66% speak Korean, and 89% are MT speakers. Russian is of intermediate standing between them, with waves of immigration since the 1980's and 49% Russian use at home (all demographic facts from Statistics Canada 2007; 2009). In the future, additional languages will be investigated, as researcher interests allow, to develop a series of studies exploring both newer and older communities and further typological diversity.

In addition to the five widely known HLs, this project includes a study of Faetar. Faetar is an endangered Romance variety spoken by fewer than 1,000 people in two mountaintop villages in southern Italy (Apulia): Faeto and Celle di St. Vito. Faetar, a term I use as shorthand for "Faetar and Cellese" only because my own research has been centered in Faeto, is a dialect of Francoprovençal (FP), a language which has died out in France, but due to a migration from France some 600 years ago, survives in these two isolated villages, as well as in several émigré pockets in North America. Perhaps the largest of these pockets is in the GTA, where some speakers estimate there are more than 2,000 people whose ancestors come from Celle. It is

certainly still fluently spoken by at least the 20 people who have so far been recorded for this project. Description of the variety is available in Heap & Nagy 1998; Nagy (1994; 1995; 1996; 2000; 2001; Nagy & Reynolds 1997). A brief example, using Faetar data, of the type of comparisons we will be conducting between HL and homeland varieties concludes this paper.

## Methods

This project involves six stages, the first four of which are already well underway.

### (2) Project stages

1. Establish the communities of interest.
2. Interview and record speakers for approximately 1 hour, beginning with members of their own social networks.
3. Transcribe all interview material broadly in a time-aligned, digital archive.
4. Analyze sociolinguistic variables in each language.
5. Compare trends within and across languages.
6. Develop a generalized model of contact-induced change in a multilingual metropolis.

### **Speaker samples**

In the first stage, research was conducted to find HL communities which are both well established in Toronto and represented by native speakers among the students and faculty members studying linguistics at the University of Toronto. Six languages were selected. In each community, we aim to interview 40 speakers, divided evenly among three generations of speakers. We define *first generation* as speakers who grew up in the homeland (until at least age 18) and then migrated directly to the GTA at least 20 years ago. Thus, we are not dealing with speakers likely to be actively acquiring either the HL or English. *Second generation* speakers are defined as people whose parents (at least one) are first generation speakers. *Third generation* speakers are defined as people whose parents (at least one) are second generation speakers. Within each generation, we seek representatives from three age groups. Within each age group, in each generation, there will be four speakers, two male and two female. Table 2 is an example of the speaker distribution, at a certain stage during data collection, of the Italian sample. The code in each filled box indicates the language, generation, sex (M or F) and age of the representative speaker. Empty boxes indicate categories where speaker recruitment was still in progress. Age, generation, sex, long-time, residency in the GTA, and the self-defined ability to converse for an hour in the HL, are the only selection criteria for speakers. Thus they will vary greatly in degree of fluency (in both the HL and other languages spoken), length and frequency of usage of the HL and other languages, and their ethnic orientation. Such factors will be included as independent variables in our analyses. While the goal is to fill this distribution table for all languages, we come up against the lack of availability of speakers in some of these cells, e.g., third generation Koreans (because the first generation did not arrive long enough ago for grandchildren old enough to participate in the study to exist). We also aim to match our sample with those of *TEA* and *ELP* in order to construct a complete picture of language usage in each community.

Table 2: Italian speaker sample (collection still in progress)

| ITALIAN                              | Age   | Male   |        | Female |        |
|--------------------------------------|---|--------|--------|--------|--------|
| <b>1<sup>st</sup><br/>generation</b> | >60 years   | I1M75A | I1M66A | I1F83A | I1F73A |
|                                      | 39-60   | I1M58A | I1M35A | I1F57A | I1F53A |
|                                      | <i>no 1<sup>st</sup> generation speakers &lt;19 or 19-38 exist, by definition</i> |        |        |        |        |
| <b>2<sup>nd</sup><br/>generation</b> | >60   |        |        |        |        |
|                                      | 39-60   | I2M42A | I2M53A | I2F44A | I2F44B |
|                                      | 19-38   | I2M22A | I2M34A | I2F34A | I2F34B |
|                                      | 12-18   | I2M19A | I2M19B |        |        |
| <b>3<sup>rd</sup><br/>generation</b> | >60   |        |        |        |        |
|                                      | 39-60   |        |        | I3F59A | I3F58A |
|                                      | 19-38   | I3M25A | I3M20A | I3F23A | I3F33A |
|                                      | 12-18   | I3M18A | I3M18B | I3F18A |        |

### Data collection

It is essential that this type of work be based on naturally occurring speech, rather than the small samples of elicited or translated forms that are often the basis for theoretical linguistic description. To this end, data is gathered via three tasks. The first is a sociolinguistic interview (Labov 1984), a method designed to elicit casual speech in the HL, such as might occur within the community. Interviews are about one hour in length, covering a broad range of topics relating to the speaker's background, social networks, and experiences with language. We include questions about immigration history within the speaker's family and their sense of immigration trends in their neighborhood. Questions about what aspects of the language seems to have changed (compared to grandparents, to the standard variety, etc.) are useful for their elicitation of linguistic attitudes as well. Interviewers guide the speakers to discuss a range of these topics, following the speakers' interests rather than forcing a particular conversational structure. The goal is to minimize the distancing effects of the researcher, possibly a stranger, with a microphone and recorder asking questions. The questions are adapted from Labov (1984), and are loosely structured in topic categories such as those given in (3). While the speakers provide valuable information about their background, history, language usage, etc., they also produce extended stretch of naturalistic speech which are analyzed for linguistic forms and variation.

#### (3) Family History Module in the Sociolinguistic Interview (IV)

Why did your family move here?

Because of work?

Because of community roots?

To be close to other Italians? Close to relatives?

Do you know where your family came from?

When did they come here? Why did they come?

Do you remember hearing stories about how your family came to Toronto? ...

Was it hard for them to get set up here?

In addition to the conversational portion of the interview, an Ethnic Orientation Questionnaire (EOQ), also administered orally in the target HL, elicits the "perceived degree of

orientation to the relevant ethnic group.” This instrument is parallel to Walker & Hoffman's (2008) instrument in order to provide comparability across the HL and English samples, and is based on an instrument developed by Keefe & Padilla (1987). An English-language version of the EOQ and the guiding questions for the IV are available on the project website (Nagy 2009). A sample of the questions in the EOQ is given in (4) and other topics are listed in (5).

(4) Sample EOQ Questions

A. Ethnic identification

1. Do you think of yourself as Italian, Canadian or Italian-Canadian?
2. Are most of your friends Italian?
3. Are people in your neighborhood Italian?

B. Language

1. Do you speak Italian? How well? How often?
2. Where did you learn Italian? At home? In school?
3. Do you prefer to speak Italian or English?
4. Do you prefer to read and write in Italian or English?

C. Language choice

1. What language does your family speak when you get together?
2. What language do you speak with your friends?

(5) Other EOQ topics

- D. Cultural heritage
- E. Parents
- F. Partner
- G. Italian culture
- H. Discrimination
- I. Italian culture
- J. Discrimination

The third task is the First Words Task (FW), a picture-naming and story-telling task developed in my Faetar fieldwork (Nagy 1994). It provides easily comparable samples of basic vocabulary and structures by asking speakers to first name items in individual pictures and then describe a scene in which the items appear. This tends to provide first a more carefully-produced "citation" form and then more casually-produced repetitions in the descriptions. Ten pages of the book *First 100 Words* (Amery & Cartwright 1987) are used. Data from some 80 speakers in Faeto engaged in this task exists.

### **Transcription**

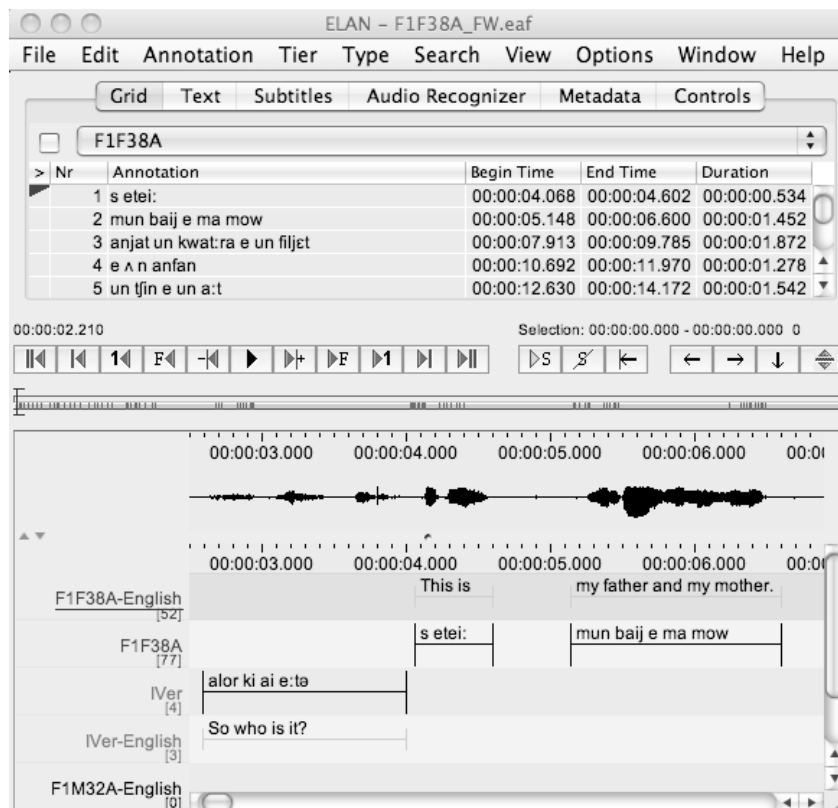
Before recorded speech can be used to investigate a wide range of variables, it must be transcribed. This task is time-consuming and must be done by fluent carefully-trained speakers, and checked by a second fluent speaker. Transcription conventions have been developed and adapted for each language (available on the project's website). Transcriptions are done using the time-aligned transcription system ELAN ([www.lat-mpi.eu/tools/elan](http://www.lat-mpi.eu/tools/elan)). This program synchronizes transcription text files, digital audio files, and other levels of mark-up, so that chronological points in each are linked to the matching points in the others. Several advantages of this method exist:

## (6) Advantages of a time-aligned transcription system

- Text and/or regular expression searches can be used to search for particular segments of the recording, once transcribed.
- At any point downstream in the analysis, the researcher has immediate access to as much context as desired for each token examined, both in the transcription and the recording.
- ELAN allows for interaction with a number of other transcription and analysis systems, with its wide variety of file import and export functions.
- Researchers can (repeatedly) revise analysis codes in ELAN and quickly recreate data files ("on the fly").
- Narrow transcription of the entire recording is not necessary. A broad transcription that serves mainly to "label" each phrase and show who uttered it, is more quickly produced. Narrow transcription of only the relevant features in only the relevant segments can later be conducted on an as-needed basis.
- Separate tiers can be added (and time-linked) to code/define/mark-up information relevant to the analysis of each variable considered.
- Transcriptions, speaker information, and coding of each linguistic variable can be exported in a format ready for statistical analysis by a number of different programs; preliminary distributional statistics are available within ELAN.

ELAN creates displays that look like this:

Figure 1: Transcription in ELAN



Each transcribed element appears aligned with the relevant portion of the sound wave and is searchable, both in the list at the top of the window and the transcription tiers at the bottom.

Different tiers are created for each speaker. Additional tiers may be created, with time-linked fields to existing tiers, and used to code or mark up linguistic forms and structure and variable patterns. When the audio file is played, the cursor indicates the current position in both the recording and the transcription tiers.

### **Data extraction**

Transcribed interviews yield two complementary types of data. The first is what the speakers say: the demographic and background information about each speaker and the community in which they reside. The second is examples (tokens) of selected linguistic variables, culled from the transcriptions. One variable at a time, a sufficient number of tokens is harvested and coded. Coding marks the linguistic context of the token and the social characteristics of the speaker. Codes are initially marked in ELAN and then may be exported as a table for further analysis in other software packages. In the case of pronunciation variables, relevant acoustic measures are also placed in the table, after measurements are conducted using ELAN's built-in link to the acoustic analysis program *Praat* (Boersma & Weenink 2010).

### **Variables**

A selection of variables from the phonetic, phonological, morphological and syntactic domains will ensure a representative description of language contact effects and allow for systematic responses to the questions regarding which levels of language are more susceptible to transfer. Aspects of each language which are distinctive and most likely to arouse comments in terms of “accent” will be the first variables investigated, but the selection must be tempered by availability of corresponding descriptions in the homeland variety. (7) lists candidate variables that have been identified for analysis across all (or many) of the HLs.

#### (7) Candidate cross-linguistic variables

##### *Phonetic*

Voice Onset Time

Korean (Park *in prep.*)

##### *Phonological*

Word-final C deletion & devoicing

##### *Morphological*

Pro-drop (Variable surface subject presence)      Russian (Hollett *in prep.*)

Cantonese, Faetar, Italian, Russian (Nagy *et al.* 2010)

Case and gender marking

Classifier usage

##### *Syntactic*

Word order (major constituents, modifiers)

##### *Lexical*

Borrowings      Korean (Chung 2010), Faetar (Nagy 2010)

Use of (home country) archaisms      Faetar (Nagy 2010)

### **Analysis Stage 1: Monolingual analysis**

The first analytic step is to describe variables at each level of grammar in each language individually, reorganizing formal linguistic theory in order to incorporate variable as well as



categorical patterns (Guy;1991; 2007; Guy & Boberg 1996; Heap & Nagy 1998; Nagy & Reynolds 1997). For phonetic variables, appropriate phonetic cues are measured in Praat. For other variables, discrete variants are coded. Distributional frequencies of the variants are calculated and subjected to multivariate analysis to see how the forms differ across generations and in correlation with both linguistic and social factors. We expect to see variation dependent on aspects of speakers' language history, usage, and attitudes (as measured by the EOQ). In particular, we will be looking at differences in the distribution patterns across the three generation and between speakers with strong vs. weak ethnic affinity. This will necessitate developing an equation of “similarity” that incorporates three types of quantitative comparison (factors, weights, and ranges) of sociolinguistic multivariate analysis (see Meyerhoff 2009; Buchstaller & D'Arcy 2009 for similar approaches), we will measure how much each language changes in each generation.

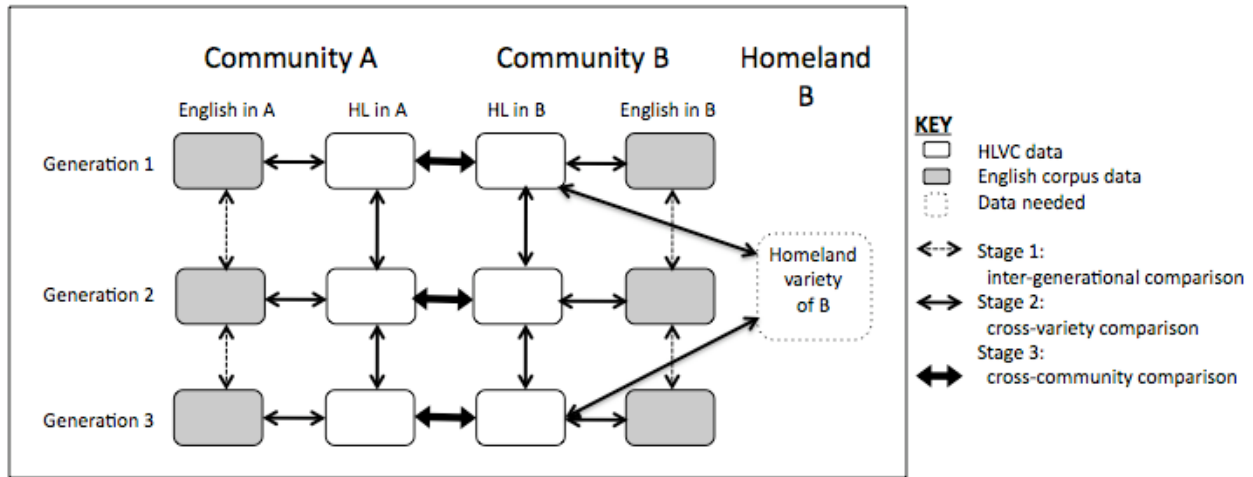
### ***Analysis Stage 2: Diatopic comparison***

Where possible comparison to similar studies of the homeland varieties will be conducted. This part of the project is the most dependent on collaborators studying the homeland varieties, or on future extension of our own work. It would make the project considerably more robust to establish a homeland "baseline" for each variable. In the interim, comparisons of the amount of shift toward English, the locally dominant language, will be made (using *TEA* and *ELP*).

### ***Analysis Stage 3: Cross-linguistic and cross-community comparison***

Several types of comparison will then be pursued to establish general principles of contact-induced language change in a highly multilingual metropolis. The first step will be to see whether the size of the cross-generational differences in each language corresponds to aspects of each language community, such as the size of the community, the proportion of it that continues to speak the HL (see Table 1), and the level of institutional support for each HL. We will then turn to drawing cross-linguistic generalizations about the features of language that change over time, via cross-linguistic comparisons across generations since immigration. Finally, we will seek a better understanding of the social patterns of who leads in linguistic changes. This will be done by comparing the sociolinguistic patterns of each community in their HL and in English. A primary issue is whether the same or complementary groups of speakers act as innovators in their two languages. Figure 2 summarizes the types of comparisons to be conducted.

Figure 2: Cross-language, cross-generation, and cross-communities comparisons



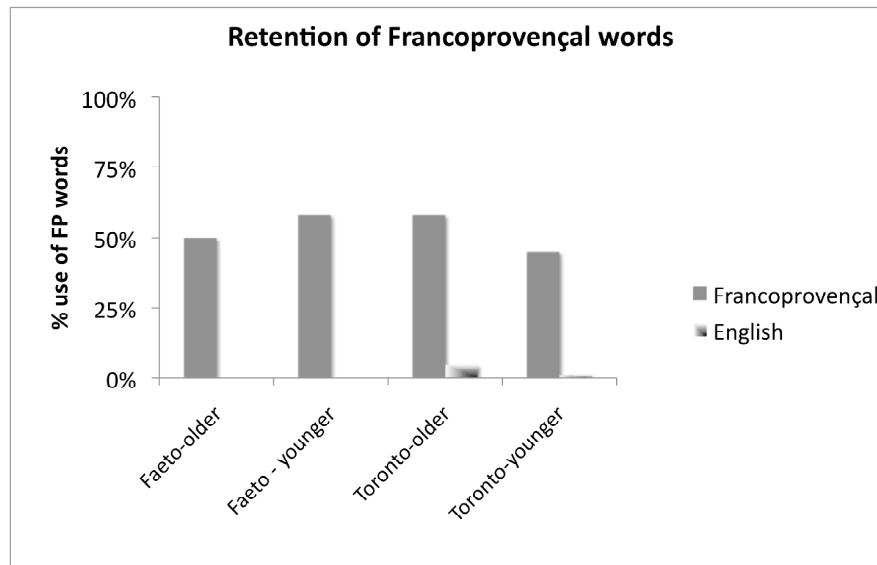
### Homeland vs. transplanted home comparisons

While we are not yet ready to conduct these comparisons in full, Faetar data is available to provide an illustration of the types of findings that this study will produce, focusing on comparison between generations and between homeland and transplanted varieties. I use simple lexical data to illustrate. Many Faetar speakers claim that the younger generation has lost almost all native vocabulary due to Italian influence, so that Faetar is on the verge of disappearing (Nagy 1994:118). The data, however, indicate that lexical change, although real, is not occurring quite as the community members describe. For this illustration, FW data is taken from Nagy (1994) for 31 homeland speakers and contrasted with HLVC data for 13 HL speakers. Of 26 words elicited in the FW task, 13 were invariably produced as Italian (cognate) words (e.g., [la vɛst] < Italian *la vesta* 'dress'), four were invariably produced with FP source words (e.g., [la kuɫiə] < FP [la kuɫi], but Italian *il cucchiaino* 'spoon'), and nine varied (e.g. both [buɸettə] < FP [la buɸe(t)] and [la tawolə] < Italian *la tavola* 'table' were produced). FP forms are from Durrafour (1969).

Considering just this distribution in the homeland, one would indeed be pessimistic about the future of Faetar, regarding its ability to maintain a distinction from Italian, as it suggests that, in basic vocabulary, 48% is now Italian and only 40% of FP origin. The 12% ambiguous forms are likely to be regarded as Italian by all but the most optimistic. However, when we look at the cross-community and cross-generational distributions, we find support for the optimistic view that, at the present time, Faetar is holding its own against Italian. Nagy (1994:26) showed a complete lack of difference between younger and older speakers in their rate of use of the FP-source vs. Italian-source synonyms in Faeto. Figure 3, from Nagy (2010), shows no significant difference in the rate of FP-source words used between Faeto (where the language remains quite isolated) and the GTA, where it is in intense contact with both English and Italian. Nor is there any evidence of a different rate of retention between first generation (average age 75) and second generation (average age 40) speakers. Thus there is no apparent time evidence (Bailey, Wikle, Tillery & Sand 1991) of a change in progress within this robustly-Faetar portion of the vocabulary. Nor is there real time evidence, given the 15 years between the homeland and HL data collections. Furthermore, the miniscule amount of English creeping into HL Faetar is, perhaps surprisingly, from an older, rather than younger speaker. Perhaps future investigation of

linguistic attitude and usage patterns will provide an explanation for this unexpected blip. However, at this point, the data is included only to illustrate the comparative methods to be exploited in our project as we seek to better understand language change in HL communities in the GTA, and more generally.

Figure 3: Retention of FP-source vocabulary in Faetar ( $N = 31$  in Faeto, 13 in Toronto)



## Bibliography

- Amery, H. & S. Cartwright, 1987, *First 100 Words*, Usborne, London.
- Bailey, G., T. Wikle, J. Tillery & L Sand, 1991, "The apparent time construct", *Language Variation and Change*, 3, 3, 21-65.
- Berkowitz, J., 2003, "Quantum Writing", *Science Communication*, 3,1, [www.quantumwriting.com/newsletter/2003\\_n3-1.html](http://www.quantumwriting.com/newsletter/2003_n3-1.html). Accessed Mar. 12, 2010.
- Boersma, P. & D. Weenink, 2010, "Praat: doing phonetics by computer (Version 5.1.29)", [www.praat.org](http://www.praat.org), Accessed Mar. 11, 2010.
- Buchstaller, I. & A. D'Arcy, 2009, "Localized globalization: A multi-local, multivariate investigation of quotative *be like*", *Journal of Sociolinguistics*, 13,3, 291–331.
- Budzhak-Jones, S. 1994, "Variable rule analysis of V/U alternation constraints in Canadian Ukrainian", *Journal of Quantitative Linguistics*, 1, 3, 202-211.
- Chung, S., 2010, "Code-switching as a means of cultural identity among Koreans in Toronto", TULCON '10 conference, Toronto.
- City of Toronto, 2010, "Local Chinese History at the City of Toronto Archives", [www.toronto.ca/archives/chinesehistory.htm](http://www.toronto.ca/archives/chinesehistory.htm), Accessed Mar. 12, 2010.
- Danesi, M., 1985, "Ethnic Language and Acculturation: The Case of Italian Canadians", *Canadian Ethnic Studies/Études ethniques au Canada* 17, 1, 48-103.

- Duraffour, A., 1969, *Glossaire des patois francoprovençaux*, CNRS, Paris.
- Fortier, A.-M., 1991, *Langue et rapports sociaux: analyse des langues d'usage chez des Italiens de deuxième génération*, Centre international de recherche en aménagement linguistique, Québec.
- Gregorovich, A., 1984, "The Ukrainian Community in Toronto from World War One to 1971", *Polyphony: Bulletin of the Multicultural History Society of Ontario: Toronto's People*, Spring/Summer 6,1, 123-126.
- Guardado, M., 2002, "Loss and maintenance of first language skills: Case studies of Hispanic families in Vancouver", *Canadian Modern Language Review*, 58, 3, 341-363.
- Guy, G. & C. Boberg, 1996, "Inherent variability and the OCP", *Language Variation and Change*, 9, 2, 149-164.
- Guy, G., 1991, "Explanation in variable phonology: An exponential model of morphological constraints", *Language Variation and Change*, 3, 1, 1-22.
- Guy, G., 2007, "Variation and phonological theory", In *Sociolinguistic Variation: Theories, Methods and Applications*, R. Bayley & C. Lucas (eds.), Cambridge, NY, 5-23.
- Handera, V.I., 1984, "The Russian Orthodox Church in Toronto", *Polyphony: Bulletin of the Multicultural History Society of Ontario: Toronto's People*, Spring/Summer 6, 1, 83-85.
- Harney, R.F., 1984, "Chiaroscuro: Italians in Toronto, 1815-1915", *Polyphony: Bulletin of the Multicultural History Society of Ontario: Toronto's People*, Spring/Summer 6, 1, 44-49.
- Heap, D. & N. Nagy, 1998, "Francoprovençal null subject and constraint interaction", *CLS 34: The Panels*, M. Gruber, D. Higgins, K. Olsen & T. Wysocki (eds.), Chicago Linguistic Society, Chicago, 151-66.
- Hollett, M., in prep., "Variation in null-subject pronoun use in Russian in Toronto", MA Paper, University of Toronto.
- Keefe, S. & A. Padilla, 1987, *Chicano Ethnicity*, University of New Mexico Press, Albuquerque.
- Kim, J., 1984, "The Russian Orthodox Church in Toronto", *Polyphony: Bulletin of the Multicultural History Society of Ontario: Toronto's People*, Spring/Summer 6, 1, 83-85.
- Labov, W., 1972, *Sociolinguistic Patterns*, University of Pennsylvania Press, Philadelphia.
- Labov, W., 1984, "Field methods of the Project on Linguistic Change and Variation", In *Language in Use*, J. Baugh & J. Sherzer (eds.), Prentice Hall, Englewood Cliffs.
- Labov, W., 2001, *Social factors, Vol. 2 of Principles of linguistic change*, Oxford, Blackwell.
- Language Archiving Technology, 2008, "ELAN", [www.lat-mpi.eu/tools/elan](http://www.lat-mpi.eu/tools/elan), Accessed Mar. 12, 2010, Last modified March 1, 2010.
- Mahootian, S. 2006. "Code Switching and Mixing." In *International Encyclopedia of Language and Linguistics*. Elsevier Ltd., Oxford, 511-527.
- Meyerhoff, M., 2009, "Replication, transfer, and calquing: Using variation as a tool in the study of language contact", *Language Variation and Change* 21, 297-317.

- Myers-Scotton, C., 1993a, *Duelling languages: grammatical structure in codeswitching*, Clarendon.
- Myers-Scotton, C., 1993b, *Social Motivations for Codeswitching: Evidence from Africa*, Oxford University Press.
- Nagy, N. & B. Reynolds, 1997, "Optimality theory and variable word-final deletion in Faetar", *Language Variation and Change* 9, 1, 37-56.
- Nagy, N. & M. Meyerhoff, 2008a, "The social life of sociolinguistics", In *Social Lives in Language: Sociolinguistics and Multilingual Speech Communities*, M. Meyerhoff & N. Nagy (eds.), John Benjamins, Amsterdam.
- Nagy, N. & M. Meyerhoff, 2008b, "The love that dare not speak its name: The monolingual bias in sociolinguistics", *NWAV* 37, Houston.
- Nagy, N., 1994, "Lexical change and language contact", *Penn Review of Linguistics* 18, 117-132.
- Nagy, N., 1995, "Double or Nothing: Romance Alignment Strategies", *University of Pennsylvania Working Papers in Linguistics* 2, 2, 93-102.
- Nagy, N., 1996, *Language contact and language change in the Faetar speech community*, PhD, Linguistics, University of Pennsylvania, Philadelphia, IRCS Technical Report 96-08.
- Nagy, N., 1997, "Modeling contact-induced language change", *University of Pennsylvania Working Papers in Linguistics: A Selection of Papers from NWAVE* 25, 4, 1, 399-418.
- Nagy, N. 2000. *Faetar*. Lincom Europa, Munich.
- Nagy, N., 2001, "Stress and schwa in Faetar", In *Italian Dialects and Phonological Theory*, Repetti, L. (ed.), John Benjamins, Philadelphia, 239-254.
- Nagy, N., 2009, "Heritage Language Variation and Change in Toronto", Project homepage, individual.utoronto.ca/ngn/research/heritage\_lgs.htm, Accessed Mar. 12, 2010, Last updated Feb. 15, 2010.
- Nagy, N. 2010. "Heritage Language Variation and Change Project in Toronto: Bilinguals' competence in their Heritage Language." Paper presented in the Insights into the Development of Sociolinguistic Competence Colloquium, *American Association of Applied Linguistics*, Atlanta.
- Nagy, N., N. Aghdasi, D. Denis, A. Motut, & D. Uscher. 2010. "Pro-drop in Heritage Languages: A cross-linguistic study of contact-induced change." *NWAV* 34, San Antonio, TX.
- Park, S., in prep., "Variation in the phonetics of VOT (voice onset time) in Korean in Toronto", MA Paper, University of Toronto.
- Polinsky, M. & O. Kagan, 2007, "Heritage Languages: In the 'Wild' and in the Classroom", *Language and Linguistics Compass* 1, 5, 368-395.
- Poplack, S., 1980, "Sometimes I'll start a sentence in Spanish y termino en español: toward a typology of code-switching", *Linguistics*, 18, 581-618.
- Renaud, J., L. Gingras, S. Vachon, C. Blaser, J.-F. Godin & B. Gagné, 2001, *Ils sont maintenant d'ici! Les dix premières années au Québec des immigrants admis en 1989*, Collection Études, recherches et statistiques, les Publications du Québec, Québec.

Sanchez, T., 2008, "Accountability in morphological borrowing: Analyzing a linguistic subsystem as a sociolinguistic variable", *Language Variation and Change*, 20, 2, 225-254.

Sankoff, D. & S. Poplack, 1981, "A formal grammar for code-mixing", *Papers in Linguistics*, 14, 1, 3-46.

Statistics Canada, 2007, "Population by mother tongue, by census metropolitan area (2006 Census)", [www40.statcan.gc.ca/101/cst01/demo12c-eng.htm](http://www40.statcan.gc.ca/101/cst01/demo12c-eng.htm), Updated Dec. 12, 2007, Accessed Mar. 12, 2010.

Statistics Canada, 2009, "Visible minority population, by census metropolitan areas (2006 Census)", [www40.statcan.gc.ca/101/cst01/demo53c-eng.htm](http://www40.statcan.gc.ca/101/cst01/demo53c-eng.htm), Updated Nov. 6, 2009, Accessed Mar. 12, 2010.

Tagliamonte, S., 2008, *Toronto English*, [individual.utoronto.ca/tagliamonte/TE.html](http://individual.utoronto.ca/tagliamonte/TE.html), Accessed Sept. 1, 2008.

Toronto Public Library, 2006. *Languages Centre*. [www.tpl.toronto.on.ca/uni\\_lan\\_index.jsp](http://www.tpl.toronto.on.ca/uni_lan_index.jsp), Last updated Jan. 25, 2006, Accessed Sept. 20, 2008.

Tucker, G.R., 1999, *A Global Perspective on Bilingualism and Bilingual Education*, Center for Applied Linguistics, Washington, D.C.

Vizmuller-Zocco, J., 1993, "Sociolinguistics of a Minority Language: Italians in Toronto and Vicinity", *Romance Languages Annual*, 324-8.

Walker, J. & M. Hoffman, 2008, *Ethnolinguistic variation in Toronto English*, [www.yorku.ca/jamesw/projects.htm](http://www.yorku.ca/jamesw/projects.htm), Accessed Mar. 12, 2010.

Weinreich, U., 1966, *Languages in Contact*, Mouton, The Hague.

## Endnotes

---

<sup>1</sup> I wish to acknowledge the support of SSHRC grant # 410-2009-2330, the contributions of my collaborators Yoonjung Kang, Alexei Kochetov, and James Walker, in the development of the project, productive conversations with H el ene Blondeau, Miriam Meyerhoff and Katherine Rehner, the research assistants (listed online in Nagy (2009), who found the speakers, interviewed them, and transcribed their speech, and especially all of the speakers who have contributed their time to this project.

<sup>2</sup> This category is referred to as "Chinese," and thus includes a number of people who speak languages other than Cantonese, but is the most comparable statistic available.

<sup>3</sup> Our research suggests that a Chinese-language church wasn't established until 1972, but there were already some 3,000 speakers by 1951 (City of Toronto).

25 August 2011, version 3

## Author contact information

Naomi Nagy  
Department of Linguistics  
Sidney Smith Hall, 4<sup>th</sup> floor  
100 St. George Street  
Toronto, ON M5S 3G3 CANADA

naomi.nagy@utoronto.ca