Heritage Language Variation & Change in Toronto

A Multilingual Corpus to Explore Geographic Variation

Naomi Nagy

University of Toronto

CLAVIER IX - Modena

Collaborators
    Yoonjung Kang, University of Toronto
    Alexei Kochetov, University of Toronto
    James Walker, York University

Funding source:
    Social Science & Humanities Research Council of Canada

Intro

- I'm a sociolinguist. I study language variation in language contact situations, in a quantitative perspective. One thing that's been bugging me in my field is this:

- 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.

- 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 6 (of the ~100) heritage languages spoken in the city.

11/5/09   Nagy HLVC / CLAVIER '09   2

% Multilingual studies *vs.*
% multilingual people in the world
(North American sample)

Nagy & Meyerhoff 2009

Nagy HLVC / CLAVIER '09   3

Problem

Studies of contact-induced language variation vary widely in terms of methods & contexts, prohibiting generalizable findings

Solution

Consistent methods and context, while varying pairs of languages in contact

11/5/09   Nagy HLVC / CLAVIER '09   4

Long-range questions

LINGUISTIC:

• Are cross-linguistic generalizations possible about the types of features, structures, rules or constraints that are borrowed earlier and more often ?

• If so, what ?

Long-range questions

SOCIOLINGUISTIC:

• How are social factors relevant ?

• Do the same (types of) speakers lead changes in both/all their languages ?

• Or do speakers choose to use one language or the other for this social "work" ?

General framework

Contrasting languages
Consistent methodology

Collect
Collaborate
Compare
Quantify
Conclude

Toronto Star 30 Dec. 2007

Plan

1. **Establish communities of interest**　　2009 √
2. **Interview & record speakers (~ 1 hour each)**　Almost done
3. **Transcribe broadly**　　This year
4. **Analyze variables in each language**　2010-2012
5. **Compare trends across languages**　2011-2012
6. **Develop generalized framework for linguistic change in a multilingual metropolis**　2013

Data & organization methods

Corpus design

6 languages
- 3 generations / language
- 3 age groups / generation
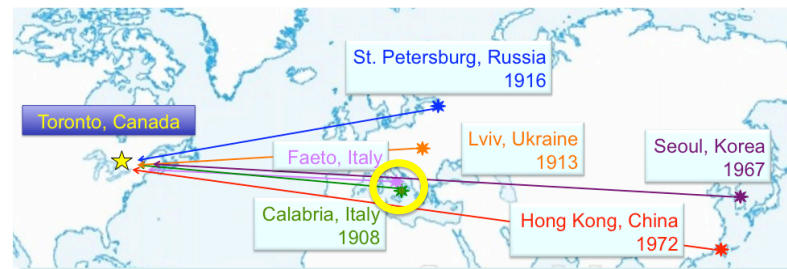- 4 speakers / age group

= 240 speakers
- Balanced for sex
- Varying in fluency, usage, and ethnic orientation

* 3 tasks
- Sociolinguistic interview / Conversation
- Ethnic Orientation Questionnaire
- Picture description task (or Reading Passage)

Languages

St. Petersburg, Russia 1916
Toronto, Canada
Faeto, Italy
Lviv, Ukraine 1913
Seoul, Korea 1967
Calabria, Italy 1908
Hong Kong, China 1972

Years
= founding of first Toronto church for each group, indicative of the establishment of a "community."

Insider interviewers

- Native speakers
- Local
- Outgoing, friendly, careful
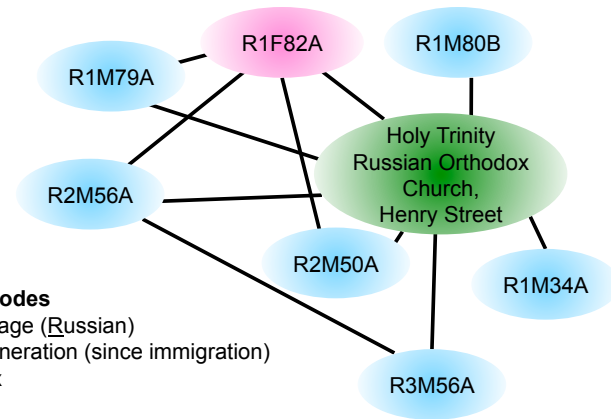- Working in pairs

A big thank you to the RAs:

Tiina Rebane
Olivia Yu
Ka-man Wong
Karen Chan
Joyce Fok
Melania Hrycyna
Taisa Hewka
Vanessa Bertone
Awet Tekeste

Rosanna Calla
Marco Covi
Iryna Kulyk
Yiqing Sim
Anna Shalaginova
Dongkeun Han
Minji Cha
Sheila Chung
Janyce Kim

Community networks

R1F82A
R1M80B
R1M79A
R2M56A
Holy Trinity Russian Orthodox Church, Henry Street
R2M50A
R1M34A
R3M56A

Speaker codes
R = Language (Russian)
1,2,3 = Generation (since immigration)
M, F = Sex
= Age
A,B – just to distinguish multiple speakers in same category

Local distribution

Chinese | All generations | Show
Mean address: Toronto, ON M1W 3G6, Canada
Root mean square deviation: 6.3 (km).

Map created by Anna Shalaginova

Data collection methods

Sociolinguistic Interview

- "Guided conversation"
- Designed to elicit relaxed, conversational speech
- Variety of topics to find speaker's interests
- Minimize the effects of a person (stranger) with a tape recorder and microphone asking questions

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?

Adapted from Labov 1984

Ethnic Orientation Questionnaire

- Identità etnica
    - Ti identifichi come Italiano? Canadese? Italo-Canadese?
    - La maggioranza dei tuoi amici sono italiani?
    - La gente nel tuo quartiere e italiana?
    - Quando eri piccolo/a i tuoi compagni di scuola erano italiani? I tuoi amici?...
- Lingua
    - Parli italiano? Parli bene? A Che livello diresti? Parli italiano spesso? Quante volte per giorno/settimana/mese?
    - Dove hai imparato l'italiano? A casa? A scuola?
    Preferisci parlare italiano o in Inglese? …
- Scelta delle lingue
    - Che lingua parla la tua famiglia quando siete tutti insieme ?
    - Che lingua parli con i tuoi amici?
    - Che lingua usi quando parli di cose personali? Quando sei arrabbiato/a?…
- Cultura
- Genitori
- Moglie/Marito/fidanzato/a
- La cultura italiana
- Discriminazione

Adapted from Keefe & Padilla 1987, Walker & Hoffman 2008

Ethnic Orientation Questionnaire

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 neighbourhood 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?

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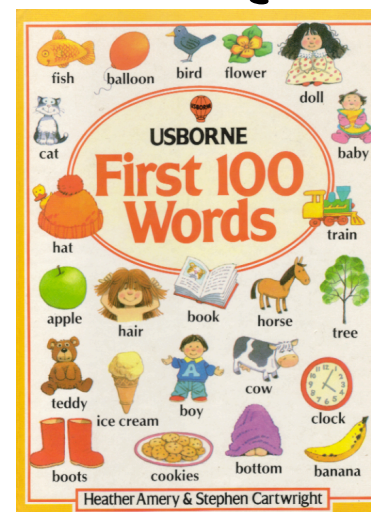H. Discrimination

I.   Italian culture
J.   Discrimination

Adapted from Keefe & Padilla 1987, Walker & Hoffman 2008

Picture description task

(More info ~ Slide #29)

Slide 1

Data analysis methods

Slide 2

Time-aligned Transcription with ELAN

ELAN – F1F38A_FW.eaf

File Edit Annotation Tier Type Search View Options Window Help

Grid | Text | Subtitles | Audio Recognizer | Metadata | Controls

F1F38A

>	Nr	Annotation	Begin Time	End Time	Duration
	1	s etei:	00:00:04.068	00:00:04.602	00:00:00.534
	2	mun baij e ma mow	00:00:05.148	00:00:06.600	00:00:01.452
	3	anjat un kwat:ra e un filjet	00:00:07.913	00:00:09.785	00:00:01.872
	4	e ʌ n anfan	00:00:10.692	00:00:11.970	00:00:01.278
	5	un tʃin e un a:t	00:00:12.630	00:00:14.172	00:00:01.542

00:00:02.210   Selection: 00:00:00.000 - 00:00:00.000 0

www.lat-mpi.eu/tools/elan
Initial transcription is broad
Additional tiers added for details.

Slide 3

Candidate cross-linguistic variables

* **Phonetic**
  * Voice Onset Time   √ Korean
* **Phonological**
  * Word-final C deletion & devoicing
* **Morphological**
  * Case and gender marking
  * Classifiers
  * Pro-drop (Variable subject presence)   √ Russian
* **Syntactic**
  * Word order
* **Lexical**
  * Borrowings   √ Korean,
                                                              √ Faetar
  * Use of (home country) archaisms   √ Faetar
* Other ideas ?

Slide 4

Sample Variable

Null Subject Presence / Absence

I1M75A
Ø Avevo 14 anni e mia moglie ce ne aveva 13.
Ø (I) was 14 and my wife was only 13…

Io ho detto "Ok, ce la faccio, io."
I said, "OK, I'll do it."

R3F25A
Я очень   медленно читаю, …
I read very slowly,…

Ø Начала   немножко   такой research делать.
Ø (I) gradually started to do some research on this.

K2F22B
할머니가 (내 이름을) 지으신것 같아요.
My grandmother named (me).

Ø 동생이랑 오빠 이름이랑 다 지으셨어요.
Ø (she) also named my younger and old brothers.
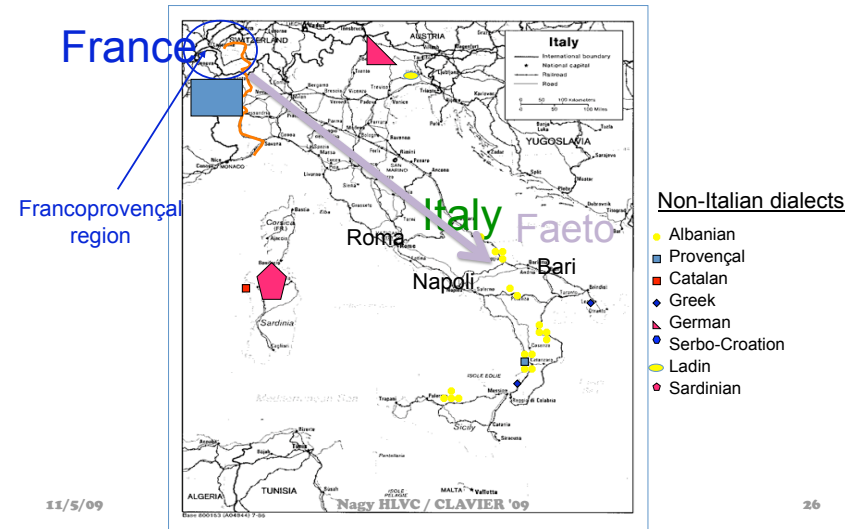
Slide 25

Sample of findings:

Lexical change in

Faeto vs. **Toronto**

http://assets.espn.go.com/photo/
2008/1115/travel_g_toronto_580.jpg

Slide 26

Map of Italy and France

France

Francoprovençal region

Italy

Faeto

Roma

Napoli

Bari

Non-Italian dialects
- Albanian
- Provençal
- Catalan
- Greek
- German
- Serbo-Croation
- Ladin
- Sardinian

Slide 27
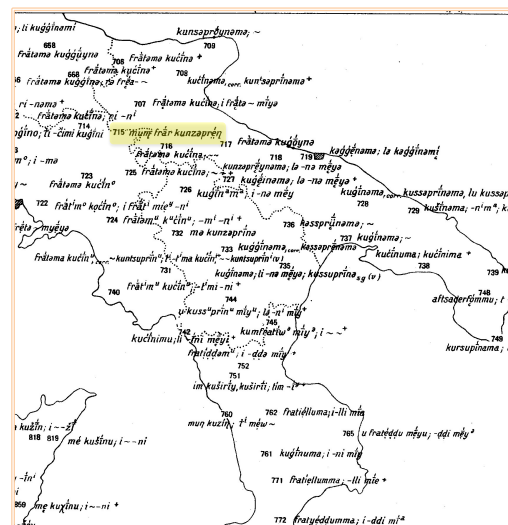
Faeto's location: #715

Jaberg, K. & J. Jud. (eds.) 1928-40.
Sprach- und Sachatlas Italiens und der südschweiz. Zofingen: Ringier.

Slide 28

Saluti a Maria

Cecile
[dʒi me səsíl tu tə kwannáj ʎo ja sun túttə parɛnn]
I am Cecile. You know. Yes, yeah. We are all family.

Peppine
[tok und a diʃ a mari undə diʃ i at maná ənə bɛllə fiʎa a faít]
What to say to Maria? We should say she sent a nice girl to Faeto.
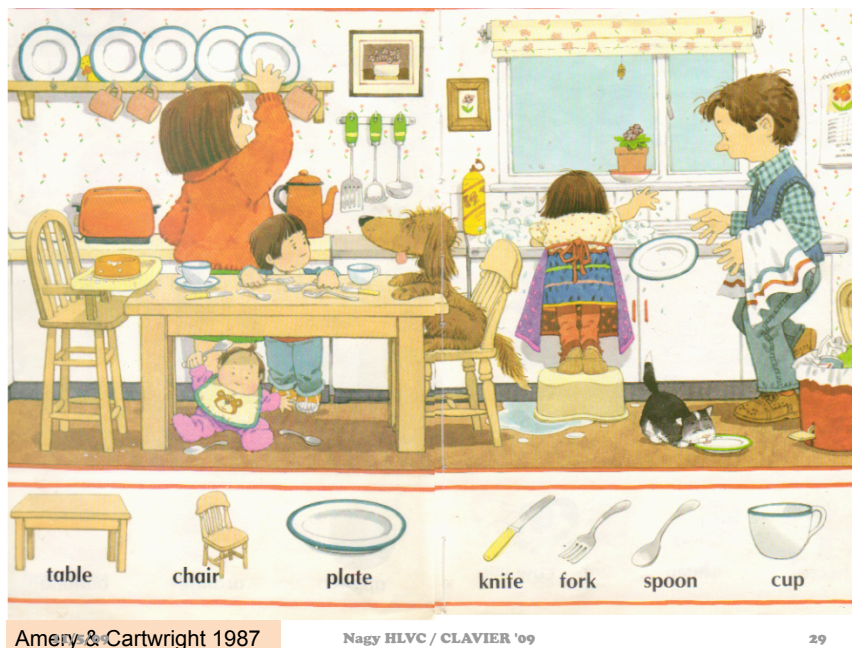
Concettine
[kəmár kuntʃətín kumə t a dir meɣ]
[t a dir sta bunn]   [və sta bunn]
"Godmother Concettine. What more should I say?"
Say, "Be well."   "You are well."
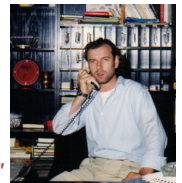
Invariably Italian words
(N = 13)

English	Faetar	Italian
cup	tatzə	tazza
sheep	pɛgɔrə	pecora
chicken	dʒəlinnə	gallina
train	trɛn	treno
dice	daddə	dado
slipper	pantuoflə,	pantofola
	ʃkarpə	scarpe
dress	vɛst	veste
jacket	dʒakkɛttə	giacchetta
house	kas, kaz	casa
(car	magənə	macchina)
(bicycle	bitʃiklɛttə	bicicletta)
(truck	kamiə	camione)

Invariably Francoprov. words
(N = 4)

English	Faetar	Francoprov.	Italian
chair	sɛdʒ	sɛʒ	sedia
spoon	kuʎijə	kuʎi	cucchiaio
cow	vatʃə	vats, vaʃ	vacca
pig	kaijunnə	?	maiale

Italian and FP forms used
(N = 9)

English	Faetar-F	Francoprovençal	Faetar-I	Italian
table	bufɛttə	bufe(t)	tawolə	tavola
	tab	tabla	tawolinə	tavolina
plate	kartolɛh		piat	piatto
knife	kuttej	kute	kutɛl	cutello
fork	brokələ	bigorə ?	fɔrket	forchetta
Grandma	mowə		nonnə	nonna
Grandpa	biaran		nonnə	nonno
hat	tʃappej	ʃape	kapəlinə	cappello
airplane	parekiə		arioplan	aereoplano
donkey /	tʃuwa	ʃuva	tʃutʃ	ciucio
horse	dʒument	ʒumà	dʒument	giumenta

Language Indices

- **Francoprovençal index**
  - the ratio of Francoprovençal-based words to the number of responses given
  - Tall blue bars indicate Faetar maintenance
- **Italian index**
  - the ratio of Italian-based words to the number of responses given
  - Taller red bars = language loss
- Indices are the combined values for the 9 words for which the lexical choices varied.

Overall results (Faeto)

48%	Francoprovençal
40%	Italian
12%	Speaker produced both, or an ambiguous form

How are these distributed?

Predicted pattern
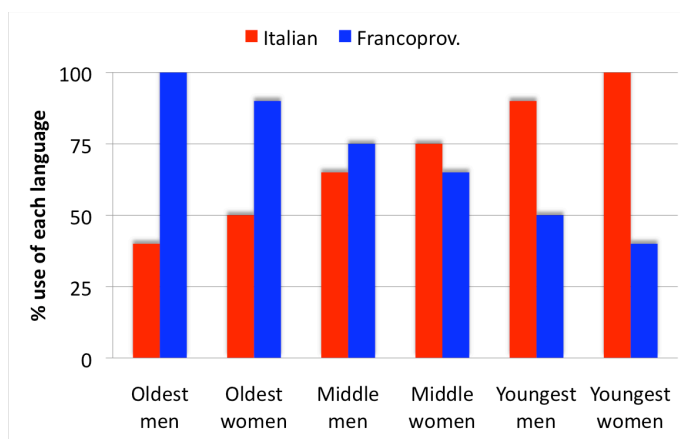
Generational effect on a phonological variable (final deletion)
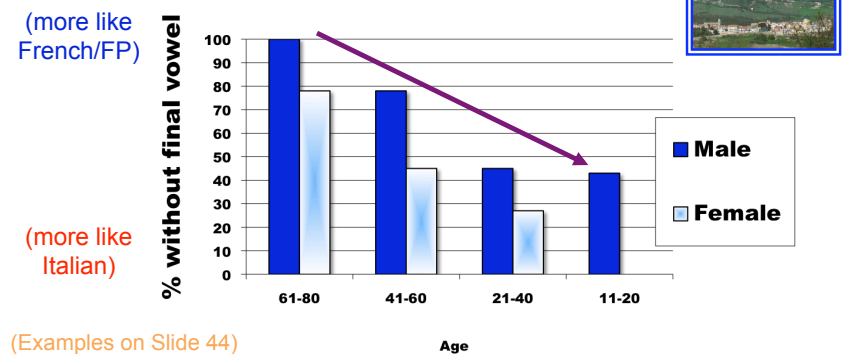
(more like French/FP)

(more like Italian)

(Examples on Slide 44)

Actual (lack of) pattern
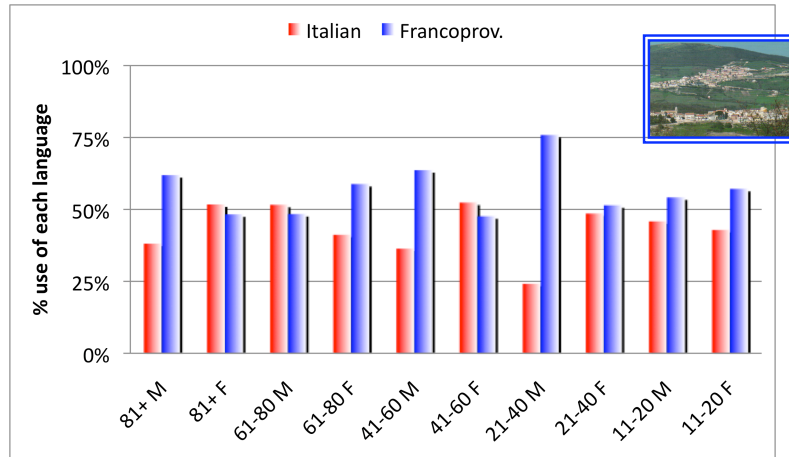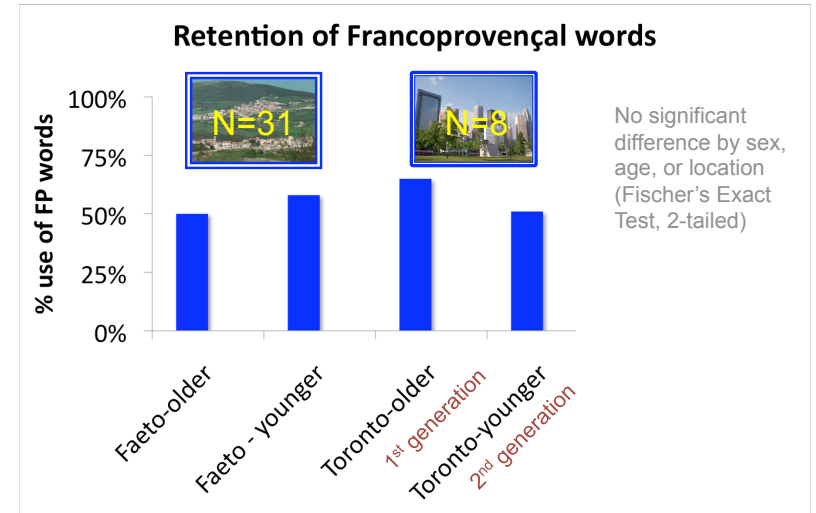
Legend: ■ Italian ■ Francoprov.

Y-axis: % use of each language (0%, 25%, 50%, 75%, 100%)

X-axis: 81+ M, 81+ F, 61-80 M, 61-80 F, 41-60 M, 41-60 F, 21-40 M, 21-40 F, 11-20 M, 11-20 F

Nagy 1994

Faeto vs. Toronto

Retention of Francoprovençal words

N=31   N=8

Y-axis: % use of FP words (0%, 25%, 50%, 75%, 100%)

X-axis: Faeto-older, Faeto - younger, Toronto-older 1st generation, Toronto-younger 2nd generation

No significant difference by sex, age, or location (Fischer's Exact Test, 2-tailed)

Faeto vs. Toronto

Retention of Francoprovençal words

N=31   N=8

Y-axis: % use of FP words (0%, 25%, 50%, 75%, 100%)

X-axis: Faeto-older, Faeto - younger, Toronto-older, Toronto-younger

Legend: ■ Francoprovençal   ■ English
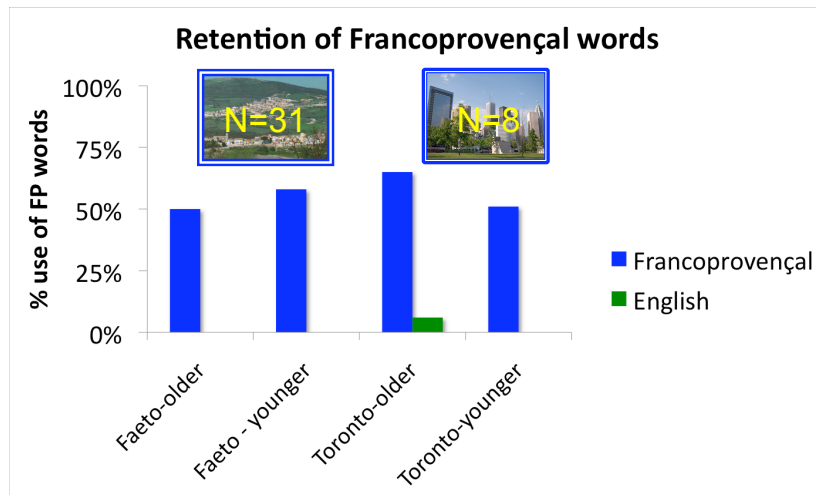
Differentiation by social class

N=31

Effect of social class on word choice

Legend: ■ Italian   ■ Francoprovençal

Y-axis: 0, 0.2, 0.4, 0.6

X-axis: Upper, Middle, Lower

Depth axis: Italian, Francoprovençal
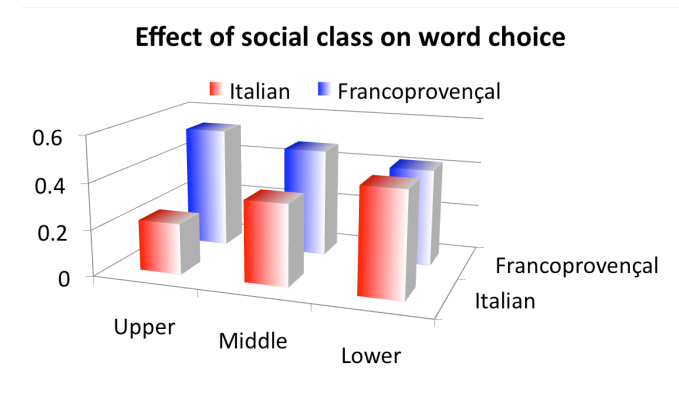
Nagy 1994

It's a big project ...

- … and you're invited to collaborate
- We'd be delighted to have suggestions of others who might have:
  - "home-country" corpora of conversational speech
    - Audio files, transcriptions & annotations
  - Quantitative sociolinguistic descriptions of home-country or other immigrant community varieties

Naomi.Nagy@utoronto.ca

References

Duraffour, A. 1969. *Glossaire des patois francoprovençaux.* Paris: CNRS.

Jaberg, K. & J. Jud. (eds.) 1928-40. *Sprach- und Sachatlas Italiens und der südschweiz*, Zofingen: Ringier.

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.

Nagy, N. 2009. Heritage Language Variation and Change in Toronto. http://individual.utoronto.ca/ngn/heritage_lgs.htm.

Nagy, N. 2000. *Faetar.* Munich. Lincom Europa.

Nagy, N. 1994. Lexical change and language contact. *Penn Review of Linguistics* 18: 117-132. Philadelphia: University of Pennsylvania Department of Linguistics.

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.

Nagy, N. & B. Reynolds. 1997. Optimality theory and variable word-final deletion in Faetar. *Language Variation and Change* 9.1:37-56.

Nagy, N., J. Walker, A. Kochetov and Y. Kang. 2009. Heritage language change and variation in Toronto. NWAV, Ottawa.

Toronto Star. The language quilt. Dec. 30, 2007.

Walker, J. & M. Hoffman. 2008. Language Contact, Linguistic Variation and Ethnic Identity in Toronto English 2008. http://www.yorku.ca/jamesw/projects.htm.

Abstract

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 acomplements 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.lat-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.

Pronunciation variation

76 year old male	14 year old female
lu tawulín	lu tawulínə
la sɛdʒ	la sédʒə
lu kartəlléh	lu kartəlléh
lu kutte'ɣ	lu kuttéjə
la brókk(əl)	la brókkələ
la kəʎi'ɣ	la kəʎíjə
la tátz	la tátzə