

*Outline of talk for:  
A Discussion Workshop on Preparation of Sociolinguistic Corpus Archives*

## **Coding (social) attitudes in Toronto**

Naomi Nagy, University of Toronto

ABSTRACT: We expect bilingual speakers' patterns of language use and attitudes toward each language (and their speakers) to influence their sociolinguistic patterns. To investigate, we use an Ethnic Orientation (EO) Questionnaire adapted from Keefe & Padilla (1987) in our studies of English (Walker & Hoffman 2010) and Heritage Language (Nagy 2009) use in Toronto.\* As part of a sociolinguistic interview conducted by community member research assistants, participants are asked a series of questions regarding their attitudes toward Canadian culture, the culture of their ethnic group, the English language, their heritage language, and their use of both languages. Currently, responses are coded categorically and average scores over sets of questions are compared. In spite of consistent methods of data collection, disparate patterns of correlation emerge between linguistic variants and EO measures (cf. the Ethnolects workshop later in the day and Nagy (2012).

### 1. Context of our study

1.1. 8 Heritage Language (HL) communities in Greater Toronto Area (GTA)

1.2. variation and change in the HLs (Hoffman & Walker look at their English, which may be a co-L1 or L2)

### 2. Who are the participants?

2.1. Because we are trying to quickly build a large-corpus of little-studied languages, and we don't really know what factors will be relevant to linguistic patterning, we are not very picky.

2.2. We are looking for a broad sample, balanced with ease of collection.

2.3. Therefore, we have strict definitions for **location** (in the GTA) and **community membership** (according to definitions of the 3 generations w.r.t. homeland), and **being a HL speaker** (HL was L1; still fluent enough to participate in a one-hour conversation – self-defined)

2.4. community members (RA students) begin by interviewing members of their personal networks, in order to have the most relaxed in-group context.

---

\* Many thanks to SSHRC for funding this research and to my collaborators who have worked on different incarnations of this material, Joanna Chociej and Michol Hoffman in particular. Other collaborators are recognized at [http://individual.utoronto.ca/ngn/research/HLVC\\_personnel.htm](http://individual.utoronto.ca/ngn/research/HLVC_personnel.htm).

- 2.5. When those resources are exhausted, they extend to other members of those members' communities, something through flyers and emails posted in organizations of which they/their networks are members.

**how to code more accurately and appropriately**

- 2.5.1. It'd be good to code for this difference in how close the participant and interviewer are, but we haven't developed that yet.

**coding for 'social situation'**

- 2.5.2. This could be partially reconstructed from the IV catalog, a spreadsheet with columns for where the IV was conducted (e.g., participant's home vs. university conference room) and "notes" about the participant (which might say whether the IVer knows them, etc.)

3. Who asks the questions?

- 3.1. community members (RA students) begin by interviewing members of their personal networks

**how demographics were elicited & why ; relative merits of likert scales,vs. question modules**

4. How do we ask the questions?

- 4.1. We are interested in the **participants' attitudes**.

- 4.2. To some extent, we can also see effects of the wider **community's (GTA) attitudes** by the degree of institutional support and status that the languages have in the GTA (Ethnolinguistic Vitality Model, Giles, Bourhis & Taylor 1977)

- 4.3. 2 formats – we could compare the outcomes – they seem complementary and each useful.

- 4.4. We ask everything in the HL.

- 4.4.1. NB: Sankoff & Thibault's Montreal Anglophones study found that people gave different answers when the same questions were asked in French vs. English, but this has not been systematically explored.

5. sociolinguistic interview (= **question modules**, RAs work from a mostly-memorized script.)

- 5.1. really and truly using the questions from Labov 1984 -Phila LCV project, with only the necessary adaptations to reflect the present circumstances.

- 5.2. asked in a conversational, open-ended fashion, jumping from topic to topic in search of the interests of the participant that will elicit uninhibited speech, especially narratives.

5.3. here, collection of EO information is incidental and disorganized. It requires much searching through transcripts for specific topics. very labor-intensive. CA-type coding would be possible, but we don't typically do that.

6. Ethnic Orientation Questionnaire (RAs work quite strictly from a script.)

6.1. adapted from social psychologists Keefe & Padilla (1987) study of Chicano ethnicity in California, because we want to compare our findings to Hoffman & Walker (2010) study of the English spoken by the same (types of) speakers.

6.2. again, conversational and (meant to be) open-ended fashion, where any sort of answer is acceptable for most of the questions

6.3. questions various measures of "contact" w/ English and Heritage lg. & Canadian and heritage culture, as well as preferences for same

6.4. 35 Questions about:

<b>Whose?</b> <b>(Reference Group;</b> developed in Boyd <i>et</i>	<b>What?</b> <b>(Topic;</b> this was Keefe & Padilla's
Participant's	Language use
Their family's	Language preference
Their network's	Language learning
	Cultural attitude
	Discrimination

6.5. For example (full text on HLVC website),

Ethnic identity

Do you think of yourself as Italian, Canadian or Italian-Canadian?

Are most of your friends Italian?

Are people in your neighbourhood Italian?...

Language use

Do you speak Italian? How well? How often?

Where did you learn Italian? At home? In school?

Do you prefer to speak Italian or English?

Do you prefer to read and write in Italian or English? ...

Family language choice

What language does your family speak when you get together?

What language do your parents prefer to speak? ...

Cultural heritage...

Parents...

Partner...

Culture...

Discrimination experience...

## 7. How do you determine how fine grained you need demographic coding to be?

- 7.1. We have some basic demographics that structure our corpus: generation, age and sex.
- 7.2. Then we want to find out what other factors are relevant, but we can't afford to interview enough speakers to balance it for every possible measure of language use and community membership. And we don't yet know what they are.
  - 7.2.1. We are exploring which of these are relevant in our study, but don't necessarily see that they will be relevant in other communities.
- 7.3. We want to compare w/ Hoffman & Walker to see the degree of commonality across the 2 languages used in each community, as a starter.
- 7.4. Not seeing any particular benefit of a pre-determined Likert scale (how would we know, *a priori*, how many points to include? whether to make it odd- or even-numbered?), we ask in an open-ended way and we record and eventually transcribe the responses so that they could be converted to a forced-choice scoring system.
- 7.5. We then code these into a 3-way distinction (which may or may not form a continuum)
  - 7.5.1. 0=preference for English language or Canadian culture
  - 7.5.2. 2=preference for HL or Heritage culture
  - 7.5.3. 1= no preference or mixed
  - 7.5.4. X=didn't answer question (possibly because it wasn't asked)
  - 7.5.5. There are 2 ways to think about these scores
    - 7.5.5.1. *Language Continuum*: most Canadian <--> most Heritage
    - 7.5.5.2. *Language Mixing*: one situation-one language vs. one situation-two languages. This separates the mixers from the non-mixers, which might be relevant to contact-induced language change.

- 7.5.5.3. The continuum method mirrors generation differences; first generation speakers are always more heritage-oriented than second generation speakers. The language mixing method gives different results depending on speaker group. The mixing method gives more “new” information (i.e., it does not just restate what generation the speaker belongs to).
- 7.6. some interviewers, having learned about the coding schema, adapt a more forced choice approach into question-asking technique. We could check the recordings and distinguish these to see if it has an effect. What hypothesis should we test?
- 7.7. But we already don’t get quite enough data when we do these at the end of a long IV. Many people don’t answer all questions. Multiple visits would be better, but hard to organize.
  - 7.7.1. We exclude from EOQ analysis participants who answer < 50% of the questions, for whatever reason.
  - 7.7.2. We exclude questions that were answered by <60% of the participants.
- 7.8. Some researchers also feel that a more fine-grained division would be better, but specific proposal are not forth-coming (unless Walker & Hoffman have something to say on this matter here)
- 8. How do we use the data?
  - 8.1. We want to find correlations between linguistic patterns (rates of use or constraint-rankings) and EOQ scores.
  - 8.2. It’s unmanageable to use 35 indep. variables in most lx. multivariate analyses, so we tried a number of other approaches.
    - 8.2.1. correlation tests among the 35 questions
    - 8.2.2. Principle Components Analysis for the 35 questions
      - 8.2.2.1. “In principal components analysis (PCA) [...] one wishes to extract from a set of  $p$  variables a reduced set of  $m$  components or factors that accounts for most of the variance in the  $p$  variables” (Wuench 2009).
    - 8.2.3. ≈ Keefe & Padilla’s method of Factor Analysis
    - 8.2.4. grouped and averaged
      - 8.2.4.1. by Reference Group (Boyd, Walker & Hoffman 2011)
      - 8.2.4.2. by Topic (Keefe & Padilla 1987)
      - 8.2.4.3. by Language Use (Chociej 2010)
    - 8.2.5. Then we can use a smaller set of EOQ factors in MV analysis.



Significant components	(t,d)					(E)					(æ)				
	All	CAN	ITA	2 <sup>nd</sup>	2 <sup>nd</sup>	All	C.	I.	2 <sup>nd</sup>	2 <sup>nd</sup>	All	C.	I.	2 <sup>nd</sup>	2 <sup>nd</sup>
Average of all 35 Qs	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
Topic method	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
Reference group meth.	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
Network ethnicity, Grandparents' AoA	ns	ns	ns	ns	-	ns	ns	ns	ns	ns	ns	ns	ns	ns	+
Family language choice	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	+
Language use meth.	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	-	ns	ns	ns	ns
Language Mixing	ns	ns	+*	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
Ethnic Continuum	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	-	+

9.1.5. Multivariate regression analyses (Mixed Effects Models) fare better than correlations, but that's for another talk.

10. how you would carry out specific tasks {IRB, demographic, situational or attitude coding} to permit YOU YOURSELF to return to the same corpus later, and carry out further comparative study.

11. IRB and sharing issues (not officially my topic, but I can contribute to those discussions just a tiny bit)

11.1. 3-level consent to use/share data (show consent form)

11.2. Corpora in the Classroom – specific tool for sharing data with students

11.2.1.1. consent to follow ethics protocol is built-in

11.2.1.2. anyone can browse what is available, only U of T community members (with UTORid) can access files.

## References

- Boyd, S., J. Walker & M. Hoffman. 2011. Sociolinguistic practice among multilingual youth in Sweden and Canada. [International Symposium on Bilingualism](#). Oslo.
- Chocie, Joanna. 2010. Quantifying Degree of Contact: Determining the Factors Significant for Heritage Language Speakers. *Bilingual Workshop in Theoretical Linguistics*, Toronto.
- Giles, Howard, Richard Y. Bourhis, and Donald M. Taylor. 1977. Towards a theory of language in ethnic group relations. In *Language, ethnicity, and intergroup relations*. Howard Giles, ed. New York: Academic Press. 307-48.
- Hoffman, M. & J. Walker. 2010. Ethnolects and the city: Ethnic orientation and linguistic variation in Toronto English. *LVC* 22:37-67.
- Keefe, S. & A. Padilla. 1987. *Chicano Ethnicity*. Albuquerque, NM: UNM Press.
- Nagy, N. 2009. *Heritage Language Variation and Change*.  
[http://individual.utoronto.ca/ngn/research/heritage\\_lgs.htm](http://individual.utoronto.ca/ngn/research/heritage_lgs.htm).
- Nagy, N. 2010. *Corpora in the Classroom-HLVC*. <https://corpora.chass.utoronto.ca>.
- Nagy, N. 2012. *Looking for contact-induced change in Heritage Languages*. LSA Annual meeting, Portland, OR.
- Nagy, N. & A. Kochetov. 2011. VOT across the Generations: A cross-linguistic study of contact-induced change. [ICLaVE 6](#), Freiburg, Germany.
- Nagy, N., N. Aghdasi, D. Denis, & A. Motut. 2011. Pro-drop in [Heritage Languages](#): A cross-linguistic study of contact-induced change. [Penn Working Papers in Linguistics](#) 17.2, Article 16.
- Wuench, Karl. 2009. Principal Components Analysis - SPSS.