What kind of data is it? Situating sociolinguistic corpora in context

Workshop on sociolinguistic archive preparation, LSA 2012

Sali A. Tagliamonte

University of Toronto

http://individual.utoronto.ca/tagliamonte

Sociolinguistic Corpora

❖ How can we document and maximize access to the context of the data collection setting?

❖ Demographic information is critical, but so is the context

  E *What kind of data are we dealing with?*

  E *How can we situate it for interpretation?*

❖ Crucial for the comparative endeavour

Typical data sample

Sociolinguistic Corpora Differences

❖ Research projects

❖ Types of communities,

❖ Eras

❖ Data types, written/spoken

❖ Dyad types, friends/strangers

❖ Etc.

❖ *But how far can we go?*

❖ *What difference does it make?*

Goals for this presentation

❖ Outline my 'best practice'

❖ Highlight some issues and problems

❖ Build on the foundations of earlier corpus-building projects

  E Canada

    ❖ 1970-1990 (Sankoff & Sankoff, 1973; Sankoff & Cedergren, 1971; Thibault & Vincent, 1990, Poplack, 1989)

  E Britain and Northern Ireland, Canada

    ❖ 1995-2011 (Tagliamonte, 1996-1998, 1999-2001, 2001-2003, 2003-2006, 2007-2010, 2010-2013).

How comparable?

- ❖ What were the original research goals and practice?
  - ᴇ If language contact is the goal then you need to contrast the relevant dimension, e.g. a border
  - ᴇ If a study of quotatives (or the historical present) is the goal, then story-telling is imperative.
  - ᴇ If a study of future temporal reference is the goal, then the fieldworkers should be instructed to ask questions about future plans and intentions.

Situational Information

E Situational information can be recorded in field notes and post-fieldwork observations that are documented in the meta-data files:

- ❖ **Time/date/place of interview**
- ❖ **Interviewing technique**
- ❖ **Interviewer(s)**
- ❖ **Participant(s)**
- ❖ **Nature of the interview situation**
  - E e.g. *what was going on, what was it like, what happened?*

E *But no black and white list to check off here!*

Does it make a difference?

❖ A trained sociolinguistic researcher, male early 20's interviewd a 13 year old girl from same neighbourhood

  E Gillian Simatovic, Toronto, Canada, 2004

❖ *Interviewer comments:*

❖ "It was tough to think of things to ask a 13-year-old girl"

❖ "Her and her cousin would talk and laugh about the questions <u>and I let them</u>"

Does it make a difference?

❖ *Beginning of IV:*

❖ [4] What grade are you in?

❖ [009] Eight.

❖ [4] Grade Eight. What school do you go to? [009] St. Brendan's.

❖ [4] Do you like it there?

❖ [009] Yeah.

Does it make a difference?

❖ *Later in the IV [25.01]*

❖ I was downtown and uh- my friend A who lives there and B was sitting on a car- a white car that was in front of A's house. And then I come out from behind the fence when her mom came out. And she 's like "B you were sitting on the car" and B was like "yes, I 'm sorry." And then she 's like "G, I saw you sitting there." And I 'm like "No I wasn 't." And she 's like, "hey I seen you!"

Time/date/place

❖ Situate the data

  E Time and space are particularly important in recent years as researchers are beginning to conduct large-scale cross-variety studies

    ❖ Buchstaller & D'Arcy, 2009; Tagliamonte, to appear; Tagliamonte, Durham & Smith, 2009

❖ 1550 vs. 2011; 1995 vs. 2001

❖ Old vs. young; pre-adolescents vs. adolescents

❖ *Make a difference to frequency and patterning*

A fast-moving change

* In Canadian English, for example, the frequency of *be like* increased from 13% to 63% [awithin the same sector of the population] between 1995 and 2001 – a 6 year time span!
  * E ([Tagliamonte & D'Arcy, 2007](#)).

* Date of birth! The change is diffusing so quickly a 12 year old's grammar of *be like* will be entirely unlike a 35 year old's.

Quotative *be like*

- ❖ Female
- ❖ 13 years of age
- ❖ Urban, Toronto, Canada
- ❖ Date of recording 26.09.04
- ❖ Interviewed with several female friends

- ❖ *What stage was* be like *at in 2004 among pre-adolescents in a large city in Canada?*
    - E Dialogue in 1ˢᵗ and 3ʳᵈ person

Patterns reflect demography

❖ I was downtown and uh- my friend A who lives there and B was sitting on a car- a white car that was in front of A's house. And then I come out from behind the fence when her mom came out. And she *'s like* "B you were sitting on the car" and B *was like* "yes, I 'm sorry." And then she *'s like* "G, I saw you sitting there." And I *'m like* "No I wasn 't." And she *'s like*, "hey I seen you!"

Nature of the data

❖ A wide range of sociolinguistic corpora in the current literature was not collected using standard sociolinguistic interviews.

> E Oral histories, interviews which were recorded for a broadcast to a much larger TV or radio audience, e.g. the 7-UP series, public speaking [Van de Velde …., Kemp and Yaeger-Dror 1991, recent work of Hall-Lew and others…].

❖ *Discuss and document the nature of the data!*

Documentation

❖ Detailed description of the project and can be a key component of the interpretation of the results.

   ᴇ African American English in the diaspora: Evidence from old-line Nova Scotians

      ❖ Poplack and Tagliamonte (1991:307-315)

   ᴇ Fieldwork and data collection practices comprise nearly 30% of the published paper

   ᴇ A critical background and foundation for the analysis and interpretation of the results that follow.

Interview Technique

* Sociolinguistic Interview?
    * E ([Labov, 1971](#), [1972b](#), [1984](#))
* Types of questions
    * E Were specific types of questions used for specific purposes?
* Who did each interview and how?
* How successful?

Mike O'Leary, *Cullybackey, NI, 2001*

❖ (CLB, Mike O'Leary, 53, MO 013, EM 3. Tommy 7, Claire 8, Tape 013)

❖

❖ *{Interviewer comment: speaker lives outside of Cullybackey and to her, 'does not sound as Scottish as those in Cullybackey district." Some technical problems. Speed of recording changes at times on side A}*

Discourse Styles

❖ Conversational interaction

❖ Story-telling

❖ Soapbox speech

  ᴇ ([Labov, 1972a](#))

❖ Performance

  ᴇ (Schilling-Estes 1998)

❖ The contrasts among these evince entirely different linguistic behavior

  ᴇ ([e.g. Paradis, 1996](#)).

Brian Whiting, *Maryport, UK c. 2001*

- (MPT, Brian Whiting, 82, BW 014, GW 2, Tapes 12 & 13)

-

- (Tape 12, Side A)
- *(NB: Speaker is quite standard)*

John Abbott, *Portavogie, NI, 2001*

❖ (John Abbott 67, JA 007, Michael Adair 008, Sheila Adair 009, HL 1.)

❖ *Has fished from various ports in the UK, but not resident anywhere other than PVG. Fantastic interview - watch out for the ghost stories! Also particularly sad story of crewmates being washed overboard.*

Study of Stutterers

❖ An uncommon population — only 1% of the population
  - E Ritter ([2008](#))
❖ Difficulties in recruiting participants.
❖ Search the Toronto English Corpus meta-data
  - E Field notes included comments such as "stuttering a lot"
❖ Examination of these audio files exposed a bona fide stutterer thus providing an invaluable informant for the research study.

Interlocutors

❖ An individual will express him or herself quite differently depending on the interlocutor(s)

  ᴇ (Cukor-Avila & Bailey, 2001; Douglas-Cowie, 1978; Watt, Llamas & Johnson, 2009, 2010).

❖ One-on-one with an out-group interviewer will produce a different type of interaction than one who is interviewed with a local 'facilitator', and both differ from the interactions between actual friends.

Relative pronoun *who*

- We discovered a high correlation of *who* with highly educated middle aged women in Toronto. Why?
  - E (D'Arcy & Tagliamonte, 2010)
- A quick search of the database (where info about the interviewer was recorded) enabled us to re-code the data file according to the nature of the interview dyad
- The frequency of *who* increased when the interviewer was a woman

When demography and situation collide

❖ The relevant factor was not only the demographics of the speaker, but the interaction of the two interlocutors' demographics, an aspect of the social interactive situation

❖ An innovative new perspective on relative pronoun variation.

❖ Interview participants' relative age, sex, age and ethnicity play into the nature of the interaction.

What to code?

❖ Type of surroundings (living room vs. front porch; grandfather clock, bird, aquarium, etc.)

❖ Particularly successful parts of the interview.

❖ Any outstanding features of the context
  - E a person who stutters a lot (noted above),
  - E someone's whose beard interferes with the mike,
  - E an interview where alcohol was involved, etc.

How to document this?

* Planning stages of research,
  * E Document rationale, goals and strategies
* Fieldwork
  * E Record anthropological observation, add to research metadata
* Transcription
  * E Add additional meta-comments and record interesting features
* Analysis
  * E Access fieldnotes to situate data and results

Typical data sample

E Male

E Born in 1939

E Northern Ireland

E Small village, Cullybackey in County Antrim [Ulster Scots community]

E Interviewed by local interviewer in 2001

E Female, same generation, known to speaker

Typical data sample

❖ And you were sort of fascinated with this whistling and crackling and chirruping and going on at this thing. And then everybody was an expert and they would prod at her and turn her a bit. (laughter) You know, but then there were a lump of copper wire that got throwed out the window. Now, you might as well have spit you know it done it nae good but some of the old boys thought it done it good.

E [Sandy Milroy, 60, CLB, 019]

Take home message

❖ Sociolinguistic data is only interpretable *in context*

❖ The researcher must make it a priority to record key aspects of the fieldwork situation

   ❖ Time/date/place of interview

   ❖ Interviewing technique

   ❖ Interviewer(s) and Participant(s)

   ❖ Nature of the interview situation

❖ This begins in the planning stage, but continues throughout a project or projects,