Large Scale Sampling of Spoken Language Across Space and Time with the Penn Sociolinguistic Archive

Bill Labov
University of Pennsylvania

The reliability of monolingual eliciting and the efficiency and dispatch of bilingual elicitation can be obtained by combining magnetic recording with bilingual field work.

NAGRA III Tape Recorder
SLX Corpus of Classic Sociolinguistic Interviews

Item Name: SLX Corpus of Classic Sociolinguistic Interviews
Author(s): Stephanie Strassel, Jeffrey Conn, Suzanne Evans, Christopher Cieri, William Kazuaki Maeda
LDC Catalog No.: LDC2003T15
ISLRN: 034-299-958-433-1
Release Date: November 25, 2003
Member Year(s): 2003
DCMI Type(s): Sound
Sample Type: pcm
Sample Rate: 22050
Data Source(s): field recordings
<table>
<thead>
<tr>
<th>Speaker</th>
<th>Age</th>
<th>Speech Community</th>
<th>Occupation</th>
<th>Ethnicity</th>
<th>Education</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adolphus H.</td>
<td>81</td>
<td>Near Hillsboro, NC</td>
<td>Farmer</td>
<td>African American</td>
<td>Very little</td>
</tr>
<tr>
<td>Bobbie A.</td>
<td>22</td>
<td>Ayr, Scotland</td>
<td>Saw Doctor</td>
<td>Scottish/Italian</td>
<td>Some technical college</td>
</tr>
<tr>
<td>Henry G.</td>
<td>60</td>
<td>E. Atlanta, GA (Dekalb Co.)</td>
<td>Railroad foreman</td>
<td>European American</td>
<td>High school graduate</td>
</tr>
<tr>
<td>Jerry T.</td>
<td>19</td>
<td>Near Leakey, Texas</td>
<td>Gas station attendant</td>
<td>European American</td>
<td>Some high school</td>
</tr>
<tr>
<td>Joe D. (interviewed with</td>
<td>21</td>
<td>Liverpool, England</td>
<td>Docker</td>
<td>English</td>
<td>Some high school</td>
</tr>
<tr>
<td>Eddie M. (Interviewed with</td>
<td>19</td>
<td>Liverpool, England</td>
<td>Docker</td>
<td>English</td>
<td>Some high school</td>
</tr>
<tr>
<td>Joe D.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kathy D.</td>
<td>15</td>
<td>Rochester, NY</td>
<td>Student</td>
<td>European American</td>
<td>In 11th grade</td>
</tr>
<tr>
<td>Louise A.</td>
<td>53</td>
<td>Knoxville, TN</td>
<td>Mother</td>
<td>European American</td>
<td>Unknown</td>
</tr>
<tr>
<td>Rose B.</td>
<td>43</td>
<td>New York, NY (Lower East Side)</td>
<td>Factory seamstress</td>
<td>Italian American</td>
<td>Sixth Grade</td>
</tr>
</tbody>
</table>
Social Motivation of a Sound Change
Martha’s Vineyard,  N=84  1961-2
The Social Stratification of English in New York City
N=321    1962-4
African-American English in Harlem
N = 436, 1965-8
Sound change in Progress, U.S.-U.K.

N = 856, 1971-98
Exploratory Interviews, 1971-1998
Linguistic change and Variation in Philadelphia
N=976, 1972-79
Atlas of North American English
N=1017, 1992-1999
Philadelphia Neighborhood Corpus
N=1280, 1973-2012
McBee Keysort Cards, 1961-64
Sorting McBee Cards, 1961-64
Kay Sonograph, 1969
Drawing formant centers on sonograms
Hardware Spectrum Analyzer, 1975
Kay wave form and spectrum
Linear predictive coding (B. S. Atal)
The FAVE web site fave.ling.upenn.edu


Forced Alignment & Vowel Extraction (FAVE)
An online suite for automatic vowel analysis

What is FAVE?
The FAVE (Forced Alignment & Vowel Extraction) program suite allows you to automatically align and extract large quantities of vowel formant measurements from sociolinguistic interviews or other bodies of orthographically transcribed data.

FAVE is a set of two programs: FAVE-align and FAVE-extract.
Kay Speech Lab measurements of a Philadelphia ANAE subject, N= 214

Dawn Torres, 29, Philadelphia, PA TS 520
FAVE analysis of Jean H., 60, Interviewed by J. Fruehwald in 2006 [N=7,101]
A natural misunderstanding (/ey/ -> /iy/)  

Rosemarie D.: All right, come to dinner! [carrying out the food on a tray]  

WL: You run a tight ship.  

Tom D. [husband]: She makes us slave.  "makes us leave"  

Rosemarie: Why would I want you to leave?  

Tom D.: One day, we'll explain it all to Rosemarie
Raising of /ey/ for Jean M., 60, Philadelphia [2006]
The raising of short-a in Philadelphia

• Anne Bower interviewing Celeste Sullivan on Daley St., 1973

[æ] → [iːə] in *bad, dance*
But I used to go *dance* at the *canteen*. (The canteen?) Yeah they had a *canteen* that– it was all service men. And you couldn't get in if you didn't have a date. And you'd have to be a service *man*-- to take you in. But once you got in you could leave this guy, you know, and *dance* with everybody else. And that's what we used to do, *stand* on the corner. And wait for the fellas, and they’d-- I'd say to one of the sailors, "Are you going in there?" And he'd say "Yeah", "Would you take me in?" "Sure!" And they would take you-- and nothing *bad* would happen in there. And– the Salvation Army *ran* a beautiful thing. It was only coffee and doughnuts, there was no drinking, and soda, and music. All night you'd *dance, dance, dance*. 
Philadelphia development of new verb of quotation *be like* by year and age

Year of interview

Proportion of *be like*

Age
- 9-12
- 13-19
- 20-29
- 30-39
- 40-59
- 60-93
This number kept on coming up on his pager when he was in this--he had to drop off papers, in Center City
so I was like checking his beeper.
And I'm like “OK, it came up once.” I'm like, "Alright."
Came up again two seconds later and I'm like . . . “Alright!!”
And it came up five times total within ten minutes.
So I wrote the number down and called it.
And I said you know, “How do you know [X]?”
And she's like, "Because . . uh . . he's my brother-in-law."
And I was like— and she was like, "How do you know him?
You one of his sluts?"
And I was like, "Oh my god!” I hung up the phone, I called him.
He's like, ”Why did you call?” whatever her name was.
And I was like, "I'm sorry."
He was like, "She told my wife.”
PRAAT: Boersma and Weenink
Defining the variable (eyC)

\[\begin{align*}
\textit{babe} & \quad \text{?} \\
\text{CVC} & \quad \text{CV}
\end{align*}\]

\[\begin{align*}
\textit{baby} & \quad \text{?} \\
\text{CVCV} & \quad \text{CV}
\end{align*}\]
Syllable context for the raising of /eyC/: open syllable (*baby*), closed syllable (*babe*), final (*bay*) & hiatus (*mayor*)—from Fruehwald 2012
Social and stylistic stratification of (ing) in the random sample of the Lower East Side of New York City adults [N=81/321]

The discovery of orderly heterogeneity

workin’
ceilin’
somethin’

Percent /in/

Casual speech  Careful speech  Reading style

Lower working class
Upper working class
Lower middle class
Upper middle Class
General principles of chain shifting

In chain shifts,

I. Tense nuclei rise along a peripheral track
II. Lax nuclei fall along a non-peripheral track
Peripherality is defined for mid and high vowels only
Apparent time changes in the vowel system of Philadelphia

Figure 3.6. Movement of Philadelphia vowels in apparent time. Circles: mean values for 116 speakers in the Neighborhood Study. Vectors connect values for groups 25 yrs older & younger than mean. _F = free vowel; _C = checked vowel. _0 = before voiceless finals.

<table>
<thead>
<tr>
<th>P &lt; .10</th>
<th>P &lt; .05</th>
<th>P &lt; .01</th>
<th>P &lt; .001</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The Northern Cities Shift in phonological space. Means for Inland North dialect highlighted.

/i/  
/e/  
/o/  
/æ/  
/oh/
What can be learned from 7,101 vowel measurements in 2013 that could not be learned from 214 in 1995?
What can be learned from 344 speakers in 2013 that could not be learned from 112 speakers in 1975?
• More rapid and more accurate means of measurement
• New questions raised by previous finding
Lexically regular variation: coronal stop deletion in English.

Two basic constraints on the simplification of /t,d/ clusters in English: the effect of a following consonant vs. a following vowel (a vs. b) and the effect of grammatical boundary (c vs. d).

Monomorphemic

hold

Past tense

held

holed
DEC PDP-11, 1970s

28K Words
Distribution by age and year of interview for 164 Philadelphia Neighborhood Corpus speakers analyzed for verbs of quotation (N > 4)