The Study of Language Variation and Change

Robert Bayley
Linguistics 152
2019 LSA Summer Institute

Ways of Studying Language and Society

<table>
<thead>
<tr>
<th>Subfield</th>
<th>Some Important Figures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language variation and change</td>
<td>William Labov, Walt Wolfram, John Rickford, Penelope Eckert, John Baugh</td>
</tr>
<tr>
<td>Ethnography of communication</td>
<td>Dell Hymes</td>
</tr>
<tr>
<td>Language socialization</td>
<td>Shirley Brice Heath, Bambi Schieffelin, Elinor Ochs, Ana Celia Zentella</td>
</tr>
<tr>
<td>Language contact</td>
<td>Gillian Sankoff, Don Winford, Carmen Silva-Conlalan, Shana Poplack</td>
</tr>
<tr>
<td>Sociology of language</td>
<td>Joshua Fishman</td>
</tr>
<tr>
<td>Language planning and language policy</td>
<td>Bernard Spolsky, Joel Walters, Charles Ferguson</td>
</tr>
<tr>
<td>Language attitudes</td>
<td>Dennis Preston</td>
</tr>
</tbody>
</table>

Sociolinguistics and the Study of Variation and Change

- The study of variation is at the core of the study of language and society.
- The field has its own conference, New Ways of Analyzing Variation, now in its 48th year, with offshoots in Asia and the Pacific, Canada, Europe, and the UK.
- The field also has its own journal, Language Variation and Change, now in its 31st year.
- Variationists have been prominent in the profession. For example, variationists who have been named as fellows of Linguistic Society of America include William Labov, John Baugh, John Rickford, Walt Wolfram, Shana Poplack, Dennis Preston, Roger Shuy, and Penny Eckert, among others.

The Scope of LVC: Key Questions 1

- It’s obvious that most people have more than one way of saying the same thing. Thus, sometimes a person will say, “I’ve been workin’ real hard,” and at other times “I have been working extremely hard.” Sometimes a person will say, “The guy who lives across the hall is a pain” and at other times “The guy that lives across the hall is a pain.”
- To what extent is the variation that we can observe every day systematic? That is, can we model in a probabilistic way, which people in which circumstances are more likely to use variant X and which people in which circumstances are more likely to use variant Y.

The Scope of LVC: Key Questions 2

- What are the social influences on language variation?
  - Social class
  - Gender
  - Ethnicity
  - Style
  - Age
  - Regional origin
  - Individual identity
- What are the linguistic influences on language variation?
  - Phonological environment
  - Grammatical class
  - Discourse type

Key Problems in the Study of Language Variation and Change (from Weinreich, Labov, & Herzog 1968)
1. The Constraint Problem
- What is a possible or an impossible change? How might a universal grammar constrain the range of change types?
- What conditions must exist for change to occur or to begin and then proceed to completion?
- What factors, linguistic and social, may constrain the type, direction, or structural characteristics of linguistic change?

2. The Transition Problem
- Given two states of a language, call them A and B, which differ in certain ways, how did the transition occur from A to B? Was the transition sudden or gradual? If gradual, what were the stages of change?
- How does change proceed within the linguistic system?
- If a sound change is involved, does the change proceed one word at a time or does it systematically apply to the phoneme wherever it occurs?

3. The Embedding Problem
The Embedding Problem entails the twin concerns of embedding with the structures of language and within the social structures of the community. Thus:
- How is the change embedded in the structures of language itself?
- In other words, how does linguistic context influence variation between innovative and conservative variants of a variable in the process of change?
- With respect to social categories of people, who promotes the change?
- Who doesn’t? Why?

4. The Evaluation Problem
The evaluation problem confronts issue of language attitudes and of communicative efficiency as change unfolds. Questions that emerge include:
- In the process of change, how does the community subjectively evaluate the change, if at all? Moreover, is such evaluation uniform from start to finish or does it differ at different points in the time course of change?
- How can one evaluate the impact of change on such issues as effectiveness of communication? Does change, for instance, restructure functional load?

5. The Actuation Problem
This is the most difficult to address:
- Why does a given change occur when it does and where it does?
- Why doesn’t it occur elsewhere?

A Key Concept: The Sociolinguistic Variable
- **Sociolinguistic variable**: Two (or more) ways of saying the same thing, e.g. *workin’/working*, or expressing the same function, e.g. introducing a quotation with *say, go, be all, be like*, etc. This criterion is fairly easy to satisfy at the levels of phonology and morphology. It becomes more difficult when we move to syntax. For example, are “Vivian closed the door” and “The door was closed by Vivian” really two ways of saying the same thing?
- The sociolinguistic variable is an abstract entity, much like a phoneme. It includes all the realizations of a variable form. Thus, *pas/(O)/ and past are tokens of the sociolinguistic variable (*-t,d*).
The Sociolinguistic Variable 2

- Variation may be observed at all linguistic levels. Thus, sometimes speakers will pronounce the /l/ in past and sometimes they won’t. Sometimes speakers will use more than one negative in an English sentence and sometimes they won’t. Sometimes speakers will introduce a quotation with “say” and other times they will use “be like”. In bilingual communities, speakers will sometimes choose one language and then switch to another in mid-sentence. We observe these phenomena in all languages, including signed languages.
- The following slides show examples of variation above the level of phonology.

Variation above phonology: Code-switching 1

“José Díaz”, San Antonio, Texas

- “uh...there was a place called El Barrillito, y más ant...más antes, bailaban El Shottz...have you heard about that?...El Shottz...es un baile...it’s kinda like a hillbilly dance...pero en español...you know...el, el El Shottz had, had its own way of...you really had to know how...to dance, you know..."y mi "huelito...daban este, to the best dancer...te, te daban seis...six pack of Lone Star...that was the prize, you know...
- "ah...there was a place called El Barrillito, and before, before, they used to dance the Shotiz...it’s a dance and it’s kinda like a hillbilly dance, but in Spanish...the the Shotiz ... and my grandfather...they gave this...they gave a six pack of Lone Star..."

Variation above phonology: Quotatives

- Innovative quotatives in Ontario English:
  - Then some girl goes "uh they jumped you right?" And I was like, "Oh, my god, you had to say that!" And I was like, "No they didn’t." And she [the speaker’s mother] was all “what, what happened?” I was like, “uh nothing.” She’s like, “J., you better tell me”. And I had to tell her. And I go "well don’t, don’t go to my school. If I have to fight then I’ll take care of it, I’ll fight them by myself.” And she goes, “Well they gave you a ticket J.” And I, she goes, “Does Miss A. (the principal) know?” So I was like, "Yes." (SA, f, 12)
  --From Bayley & Santa Ana 2004

Variation above phonology: Code-switching in the media (KXTN San Antonio) (from Bayley & Zapata 1992)

Di Talk
KXTN Tejano one-o-seven FM, the official hit station for Jesse James. Keep it right where it is to win your free tickets to the Jesse James Loja fight... December third, huh, mañana, veista? At the Arena? (Esuche y gane! Con sus amigos de Tejano one-o-seven FM. Good morning)

News and Sports
Good morning, it’s sixty-three-three, I’m Alex Cruz with the KXTN one hundred and seven second news update. City Manager Alex Briseño y su familia fueron víctimas de robo anoche. San Antonio Police dice que Briseño llegó a su casa en el fifteen thousand Peblee Díez, fueron confrontados por dos hombres armados que dijeron ser del DPS, los obligaron a que se tiraran al suelo y les robaron la Ford Explorer, las joyas de la esposa. In sports los tres equipos de basketball de Texas perdieron – Spurs one-o-six, one-nineteen a los Washington Bullets, Denver gano a los Houston Rockets one-twelve, one-o-five, Utah venció a Dallas, one-sixteen, one-o-one.

Variation above phonology: Code-switching in the media (KXTN San Antonio) (from Bayley & Zapata 1992)

Du Talk
KXTN Tejano one-o-seven FM, the official hit station for Jesse James. Keep it right where it is to win your free tickets to the Jesse James Loja fight... December third, huh, mañana, veista? At the Arena? (Esuche y gane! Con sus amigos de Tejano one-o-seven FM. Good morning)

News and Sports
Good morning, it’s sixty-three-three, I’m Alex Cruz with the KXTN one hundred and seven second news update. City Manager Alex Briseño y su familia fueron víctimas de robo anoche. San Antonio Police dice que Briseño llegó a su casa en el fifteen thousand Peblee Díez, fueron confrontados por dos hombres armados que dijeron ser del DPS, los obligaron a que se tiraran al suelo y les robaron la Ford Explorer, las joyas de la esposa. In sports los tres equipos de basketball de Texas perdieron – Spurs one-o-six, one-nineteen a los Washington Bullets, Denver gano a los Houston Rockets one-twelve, one-o-five, Utah venció a Dallas, one-sixteen, one-o-one.

Other questions motivating the study of linguistic variation: Background

- For many years, the alternation of variable forms, e.g. presence or absence of post-vocalic r or alternation of r/ and ing was considered either random (i.e. “free variation”) or unimportant (by Chomskyans).
- Non-standard dialects, especially dialects spoken by racial or ethnic minorities, were regarded by many as imperfect or debased versions of the standard language. Examples include African American Vernacular English (AAVE), Quebec French, popular Brazilian Portuguese, Puerto Rican and Southwest Spanish, among many others.
- To return to a key question – to what extent is the variation that we observe in language systematic? That is, can we predict, in a probabilistic sense, in which environments speakers are likely to use variant X and in which environments they are likely to use variant Y?
- How can we use the knowledge gained by studying disadvantaged communities to improve conditions in those communities, especially with regard to education?
How do linguists study variation in speech communities?

- Our goal is to find out how people talk when they are NOT being observed. That is, we want to know how people use language in their day to day encounters, especially in informal situations.
- However, a problem arises. To obtain the kind of data we need for analysis, we need to record people in reasonably quiet environments. We are faced with what William Labov referred to as “the OBSERVER’S PARADOX.” That is, we can only make inferences about unmonitored language use by wiring people for sound (and sometimes video) and recording them.
- Sociolinguists have developed a number of ways to overcome the observer’s paradox. The rapid and anonymous survey described in Labov’s NYC department store study is one way.

The Sociolinguistic Interview

- The sociolinguistic interview consists of a number of modules about topics such as neighborhood concerns, childhood experiences, etc. Modules are designed to elicit the maximum about of interviewee speech. Questions are short, e.g. “What are the girls/guys like around here?” or “Did you ever get blamed for something you didn’t do?” The researcher is not interested in the responses per se, but in the language in which the responses are expressed.
- Sociolinguistic interviews often include a number of sections designed to promote style shifting, e.g. a reading passage containing examples of the variables under study, a word list, and a list of minimal pairs.

Other Issues: Who are the participants?

- In most sociolinguistic studies, the interest is in the group, whether the group is the population of an entire city or region or whether the group is a small subgroup (e.g. the “nerd girls” in Bucholtz 1999). This raises an obvious problem – how do we know that the people we recruit to participate in our studies are representative of the group we are studying? Several methods have been tried:
  - **Random sampling.** Certainly the most rigorous and probably the most reliable way to be sure that the results of the study really do reflect the population. However, there are problems with this approach:
    - If we tried to sample even 1% of the population of a small city, we would have more data than we could possibly handle. In Davis, for example, we would have interviews with approximately 600 people.
    - Groups that constitute only a small percentage of the population might be missed entirely.

Judgment Samples and Social Networks

- As an alternative to the random sample, a number of sociolinguists later used judgment samples. That is, based on their knowledge of the community, they sought to include an appropriate number of speakers from each major group, e.g., men & women, speakers of different generations, ethnicities, etc.
- Beginning in the late 1970s James and Lesley Milroy used a social network approach in their studies of Belfast, N. Ireland. A similar method had been used by William Labov in the 1960s in his study of the speech of African American youth in Harlem.
- So, what did people find out?

Sociolinguistic Findings 1

- Language is characterized by “orderly heterogeneity.” That is, most of the variation that we observe in language is systematic and can be predicted in a probabilistic sense if we know enough about the linguistic environment and the social structure of the community to which the speaker belongs.
- People who belong to the same speech community share similar patterns of variation, although they may use particular variants at very different rates. For example, all of the New Yorkers in Labov’s study pronounced /r/ more frequently in formal than in informal styles, regardless of their overall rate of /r/-usage.

(r) Indices for Three Social Classes and Five Styles in NYC

<table>
<thead>
<tr>
<th>Class</th>
<th>Minimal pairs</th>
<th>Word list</th>
<th>Reading passage</th>
<th>Interview style</th>
<th>Casual speech</th>
</tr>
</thead>
<tbody>
<tr>
<td>Style</td>
<td>Lower Working Middle</td>
<td>50.5</td>
<td>45</td>
<td>30</td>
<td>76.5</td>
</tr>
</tbody>
</table>

Source: Labov 1966.
Note that a smaller number = less frequent use of the r-less variant.
Social stratification of /r/ in New York City

IN’ vs. ING in New York City by class and style

Negative Concord in Detroit AAVE by Gender and Social Class

Percentages of –t,d deletion in Detroit African American speech

Relative pronoun choice in American English

Sociolinguistic Findings 2

- In the data in the previous tables and graphs, we have illustrations of the following findings, which have since been replicated in many languages and many different communities:
  - There are no single style speakers.
  - In general, speakers tend to shift toward the more standard variant when they are paying greater attention to speech.
  - Women usually produce fewer socially stigmatized variants (e.g. negative concord) than men of the same social class and ethnic background.
  - Variation is constrained by social factors such as class, gender, and ethnicity, by speech style, and by linguistic factors such as morphological class and the surrounding phonological environment.

<table>
<thead>
<tr>
<th>Gender</th>
<th>Upper Middle</th>
<th>Lower Middle</th>
<th>Upper Working</th>
<th>Lower Working</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>10.4</td>
<td>22.4</td>
<td>68.2</td>
<td>81.3</td>
</tr>
<tr>
<td>Female</td>
<td>6.0</td>
<td>2.4</td>
<td>41.2</td>
<td>74.3</td>
</tr>
</tbody>
</table>

Source: Wolfram 1969

<table>
<thead>
<tr>
<th>Social Class</th>
<th>Following vowel:</th>
<th>Following consonant:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-t,d is past morpheme (e.g. ‘missed in’)</td>
<td>-t,d is not past morpheme (e.g. ‘mist in’)</td>
</tr>
<tr>
<td>Upper Middle</td>
<td>7.0</td>
<td>28.0</td>
</tr>
<tr>
<td>Lower Middle</td>
<td>13.0</td>
<td>43.0</td>
</tr>
<tr>
<td>Upper Working</td>
<td>24.0</td>
<td>65.0</td>
</tr>
<tr>
<td>Lower Working</td>
<td>34.0</td>
<td>72.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gender</th>
<th>Following vowel:</th>
<th>Following consonant:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>49.0</td>
<td>79.0</td>
</tr>
<tr>
<td>Female</td>
<td>79.0</td>
<td>97.0</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>Relative pronoun</th>
<th>Kikai et al. 1987</th>
<th>Guy &amp; Bayley 1995</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>%</td>
</tr>
<tr>
<td>wh</td>
<td>598</td>
<td>34</td>
</tr>
<tr>
<td>that</td>
<td>782</td>
<td>45</td>
</tr>
<tr>
<td>Ø</td>
<td>300</td>
<td>21</td>
</tr>
<tr>
<td>Total</td>
<td>1740</td>
<td>827</td>
</tr>
</tbody>
</table>
Sociolinguistic Findings 3

- On the basis of results such as those we have seen, sociolinguistic analysis observes two main principles:
  - The principle of quantitative modeling, which states that given sufficient data we can establish correlations between various linguistic (internal) and social (external) factors or influences on variation;
  - The principle of multiple causes, which states that it is unlikely that the variation we see at all levels of language can be attributed to a single cause.
- As we will see, these principles give rise to the need for multivariate analysis.

Other core principles

- The principle of accountability: we are responsible for all the data in our corpus. We cannot arbitrarily exclude data that do not fit our hypothesis. We need to account for it all.
- The principle of falsifiability: As in any area of scientific endeavor, our hypotheses must be falsifiable.

Consistency of findings

- Five decades of sociolinguistic research have resulted in a remarkably consistent body of findings. For example, speakers of English always delete coronal stops more frequently from monomorphemes than from inflected forms. Speakers and signers of a variety of languages always use overt subject pronouns more frequently when there is a switch in reference than when there is continuity of subject. We could easily multiply examples.

What about the individual?

- Although sociolinguistic studies have produced remarkably consistent findings, we are left with a basic question: if the community (and various subcategories within the community) is the unit of analysis, how do we know that our findings apply to individual speakers? For example, if half the speakers in our study always used the alveolar form of (ing) and half always used the apical form, we could report a rate of alveolarization of (ing) of 50%, although the rate would not describe any individual.
- In practice, the scenario just described doesn’t happen. Guy (1980), for example, demonstrated in fine detail how speakers from New York and Philadelphia replicated in fine detail the constraints on coronal stop deletion, given sufficient data.

What about the individual? An Example

- As we will see in this course, recent advances in statistics, particularly the development of mixed models of logistic regression, have enabled us to easily consider the influence of individuals.
- Final /z/-devoicing, e.g. boy[z] → boy[s], is a well-attested feature of Chicano English as well as of a number of other English dialects (e.g. Maori English in New Zealand).
- In a 2014 study, Cory Holland and I examined data that I had collected earlier from young people in a public housing project in a San Antonio, Texas barrio.
- We coded more than 1,800 examples of the variable and found that /z/-devoicing was subject to a wide array of constraints, both linguistic and social.
- One important predictor of the extent of /z/-devoicing was the desire of the speaker to escape the rather dismal conditions of the housing project (“the courts”). Results for four individuals are shown in the following table.

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Orientation</th>
<th>% Devoiced</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alicia</td>
<td>Courts</td>
<td>57.0</td>
</tr>
<tr>
<td>Don</td>
<td>Courts</td>
<td>69.8</td>
</tr>
<tr>
<td>Bettina</td>
<td>External</td>
<td>25.0</td>
</tr>
<tr>
<td>Keith</td>
<td>External</td>
<td>22.8</td>
</tr>
<tr>
<td>All speakers</td>
<td></td>
<td>39.1</td>
</tr>
</tbody>
</table>

These results illustrate that even when they live in the same environment, individuals differ greatly in their rate of use of individual variants. However, all individuals are subject to the same linguistic constraints, regardless of their overall rate of use.

Aside from Adamson and Regan’s (1991) study of (ING) and Bayley’s (1996) study of CCR, most early studies concerned Type 1, or horizontal variation, that is variation in features that are usually considered invariant in the target language.

Beginning with Adamson and Regan’s (1991) study of SE Asian learners and Regan’s study of Irish learners acquisition of patterns of ne deletion in French, researchers also began to examine Type 2, or vertical, variation.

Much of the work on Type 2 variation focused on L2 French (see Mougeon et al. 2010 and Regan et al. 2009), but other languages were also studied including L2 Chinese (Li 2010, 2012, 2014).

What are some of the communities studied? Here is a small sample:

- Anglo (mostly) HS students in Michigan (Otheguy, Zentella, Livert 2007; Otheguy & Zentella 2012).
- Guyanese (Riddiford 1987) and Jamaican Creoles (Patrick 1998).
- Indian English (Chand 2009).
- Maltese English (Bonnici 2010).
- Tai zhong Mandarin (Li ao 2010).