SOCIOLINGUISTIC VARIATION AND CHANGE

Miriam Meyerhoff
Linguistics and English Language, University of Edinburgh, UK

Keywords: Language variation, linguistic variability, language change, linguistic structure, speech community, age, gender, social class, constraints, style, register, stereotypes, markers, awareness, sociolinguistic interviews, quantitative.

Contents

Summary

All language is subject to variability: some is completely predictable, but some is not. What appears to be free and unconstrained variation between and within individuals often proves to be subject to non-categorical linguistic and/or social constraints. This structured variability may be an indication of changes taking place in the language, thus diachronic studies of change may be complemented by synchronic studies of variation. The research methods and the quantitative analysis of such variation are reviewed. Key findings in variationist linguistics are discussed: the systematic distribution of variants in different styles and across different social groups (age, class or social network, gender). Striking parallels between the distribution of variants across different social categories emerge: for example, changes below conscious awareness are favored in more careful styles, by speakers from the middle (or lower middle) class and women. The meaning of these distributional parallels are considered. Finally, recent directions in the study of language variation and change are discussed. These include links with qualitative methods for studying language variation and the use of variable data to illuminate studies of language contact.

1. Sociolinguistics and the Study of Variation and Change

When you start to study any language, it immediately becomes clear that it contains a lot of inconsistencies and irregularities. Sometimes different people in a group of

speakers use one pronunciation for a word, and sometimes they use another without changing its meaning. Sometimes speakers use different word orders without that difference contributing anything substantive to what the sentence means. Often where there is this kind of variation *between* speakers (interspeaker variation), we also find the same variation *within* speakers (intraspeaker variation), in other words, the same person may alternate between different pronunciations of a word, or different ways of ordering elements in a sentence.

Variation like this is central to linguistics. Indeed, there would be very little for linguists to study if, for instance, a computer had designed language and made everything completely orderly. But languages aren't designed as perfectly regular systems. They take their shape in the way speakers use them in social and interpersonal contexts. The variation within and between speakers that we observe is partly the result of interactions between linguistic factors (that is, aspects of the grammar and phonology of the language) and partly the result of interactions between social factors and language (e.g. who the speaker and addressee are, whether the talk occurs in a formal or informal context). A close study variation involves taking all the social and linguistic factors into account, and for this reason the study of language variation is generally described as a form of *socio*linguistics.

What does it mean to say a language takes its shape in social and interpersonal contexts? It means that speakers use it to explore and reflect their attitudes to themselves and others. Language can be a tool for negotiating and reflecting the relative importance of different social and personal qualities. In this article, we will consider some of the more important qualities that sociolinguists have found to constrain language variation. These include a speaker's identification with different social groups, such as social class, friendship networks, gender and age. It will also review some of the more interpersonal factors that have been found to affect variation, such as the relative formality in which they are speaking, or their familiarity with others present.

We will also see that variation may be studied in two ways. One way is synchronically (that is, at a single point in time), and this provides a snapshot of the social and linguistic features that show the most robust correlations with the variation observed. A second way is diachronically, that is, over a period of time. In many cases (but not all), the variation we observe in today's speech turns out to be the seeds of tomorrow's change.

1.1. Is Variation "Free" or Structured?

Before the 1960s, the general feeling in Western linguistics was that some of the variation observed in language was "free" and "unconstrained". Some forms of variation in language are entirely predictable. These are known as allophones, when they involve elements from the sound system alone, or allomorphs, when they involve elements in a language's morphology as well. For example, in Spanish, the sounds [b] (a voiced labial stop) and [ß] (a voiced bilabial fricative— imagine a sound halfway between *b* and *v*) are completely predictable allophones: [b] occurs at the start of a syllable and [ß] everywhere else (example 1). Similarly, in English, the prefix *in-* (meaning 'not') has allomorphs that vary in form depending on the following segment (example 2).

1. Example of allophonic variation (Spanish)
   [bino] wine [ußu] grape
   [imbierno] winter [sußteraneo] subterranean

2. Example of allomorphic variation (English)
   in + alienable > *in*alienable
   in + sufficient > *in*sufficient
   in + possible > *im*possible
   in + legible > *il*legible

But some examples of variation are not predictable in this way. For example, speakers of Latvian may delete vowels in many words, but there are no simple linguistic rules telling us when they will do so (example 3).

3. Vowel deletion in Latvian
   [kuplu] [kupl] full (masculine accusative singular)
   [spilgtas] [spilgts] dazzling (feminine nominative plural)
   [sakne] [sakn] root (feminine nominative singular)

Both forms are heard in Latvian and even the same speaker may use both. Another good example is the alternation between *–in* and *-ing* in English words like *finding*, *running*. Sociolinguists call these phenomena *variables* and the way they are realized are called *variants*. For generations, when linguists were unable to find linguistic factors that reliably predicted the form of a word, they would say that variants like these that are found in natural speech were in "free variation".

1.2. The Role of Constraints on Variation

However, in a ground-breaking survey undertaken in 1962, William Labov demonstrated that a close quantitative analysis of the distribution of apparently "free" variants actually shows some internal regularity. He noted that words like *ice* and *night*, *mouth* and *loud* were pronounced in different ways on the island of Martha's Vineyard (in the United States)—sometimes with a raised or sometimes with a lower starting point in the vowel. There was no apparent linguistic basis for this (i.e. they were not allophones) but when he correlated the distribution of different pronunciations with different social or personal attributes of the speakers on Martha's Vineyard, he found that the variation was not completely random. For instance, men in their thirties from fishing villages and who liked living on the island were very likely to use the raised variant.

As we will see in subsequent sections, the correlation with age is often important in inferring diachronic (or on-going) change from synchronic variation. As we will also see, some of the social factors (e.g. gender, social networks) that he identified as relevant on Martha's Vineyard, are relevant to the analysis of variables in other communities of speakers. For instance, the English *–ing* variable is strongly constrained by the social class of the speaker (as a group, middle class speakers use more *–ing* variants than they do *–in* variants), formality (all speakers use more *–ing* forms in careful speech than they do in casual speech), and gender (as a group, women tend to use more *–ing* forms than men do).

Some variables are primarily constrained by linguistic factors. Latvian vowel deletion in (3) is one example. Vowel deletion is more likely if the vowel immediately follows a glide [j], and if the syllable is closed, rather than open (*spilgtas* rather than *kuplu*). However, unlike the linguistic constraints on allophones/allomorphs, these constraints do not predict the form that speakers will produce 100% of the time. They are probabilistic constraints, not deterministic ones.

The result of Labov's startling finding on Martha's Vineyard and subsequent work following his methods (these methods are outlined in the next section) was to redefine the notion of "free" variation. Variation can be restated in terms of *inherent variability* and *structured heterogeneity*. That means that instead of modeling variation as something that is random or unconstrained, variation can be modeled as something that is integral to the linguistic system (i.e. variability is an inherent property of natural language). Furthermore, if variation in inherent to language, Labov postulated that variation could and should be represented as part of the structure of a language, not extraneous to it. That is, because the varied (heterogeneous) forms occurring in natural language are regularly distributed across speakers, this structured heterogeneity must also be one of the things that speakers "know" when we say that they know a language.

We will see further evidence of this structured heterogeneity as this article unfolds. However, it is important to state up front because it defines one of the most fundamental differences between the way linguists can approach inter- and intra-speaker variation. Many formal linguists (especially formal syntacticians) argue that variation involves an alternation between different grammars. In other words, they do not agree that variation is inherent to the grammar. Variation is, instead, seen as the outcome of speakers' alternation between different, competing, grammars.

Many functional linguists, too, reject the notion that variation is an inherent part of the grammar. Functionalists stress the contingent nature of all language and all interaction. Generally speaking, functionalists consider variation to be the outcome of sometimes very subtle differences in how a speaker feels about what they are talking about, or who they are talking to, or the relation between the current utterance and past or planned utterances. In other words, functional linguistics has little use for the notion of linguistic variables realized by variants with the "same meaning". For them, all variation indicates subtle differences in (social or linguistic) meaning.

Most of the data to be reviewed in this article will be quantitative studies following the tradition of Labov.

2. Methods for Studying Variation and Change

In order for a quantitative study of variation to produce reliable results, researchers usually need to draw on quite large corpora of spontaneous or natural data. Speakers may be quite unaware of some variation in their speech, so it is not useful to ask them to provide grammaticality judgments of the kind where they accept or reject sentences or words produced in made-up contexts. In addition, the patterns that underlie the variation may be very subtle, so it may be impossible for even highly-trained descriptive linguists to identify which social or linguistic factors play a statistically significant role (where this article uses the term *significant(ly)* it will always mean "statistical(ly) significant",

that is, something that is not attributable to chance).

The more factors that the linguist intends to investigate, the more data s/he will need to undertake their study. This is because the most commonly-used quantitative tests generally require at least 20 tokens to produce reliable results. If there are many social and linguistics factors to be tested, and you need to have at least 20 tokens for every possible combination of those factors, the researcher must start out with a very large number of tokens overall.

The usual method adopted to satisfy the need for natural, spontaneous speech and many tokens is the sociolinguistic interview. The classic interview involves several hours (often over repeated visits) of speech from every speaker. The interviewer tries to encourage casual speech by discussing topics such as childhood reminiscences and games, and personal anecdotes of dangerous, or spiritual, or supernatural experiences the speaker may have had. With luck, during an extended interview the interviewee will also speak to someone other than the interviewer for a while; this is assumed to capture even more naturalistic speech. The interview may stand alone or it may be complemented with three additional verbal tasks. These very often include reading a prose passage aloud (to provide a more formal register), and the reading of a word list or minimal pairs (minimal pairs differ only in one segment, e.g. *bet~bed* and *merry~Mary*). Sociolinguists recognize that the three reading tasks are qualitatively different from spontaneous speech, but many studies have shown that they are a good way of gradually increasing the relative attention an interviewee may pay to their speech. To this extent they seem to be a good way of replicating speech differences that are attributable to increasing formality of style.

A collection of interviews comprises a *sociolinguistic survey* or a *social dialect survey*. The resulting corpus of speech is similar to a regional dialect survey (in which speakers from a range of locales are asked, e.g., to name objects or "translate" a set of sentences into their local, vernacular). But the data gathered in a sociolinguistic survey is less subject to interference from standardized norms or perceptions about what is or is not correct. This is because the primary data consists of a stream of spontaneous speech. The trade-off associated with this spontaneity is that it is impossible to control the data that is gathered. That is, the researcher cannot be sure that all speakers will produce tokens of a variable at the same rate and in exactly the same linguistic contexts. This is one reason why sociolinguistic interviews are generally very lengthy. This maximizes the chance of obtaining comparable data from all speakers.

Other methods may be used to increase the likelihood of obtaining numerous tokens of a variable. For example, if a researcher is interested in how speakers pronounce the diphthong in *couch*, it is possible to gain a lot of very clear tokens of the sound through the use of so-called *semantic differential* tests. This involves asking interviewees a question like "What is the difference between a *couch* and a *sofa*?" Or the researcher might use a rapid and anonymous survey to quickly obtain a number of tokens of a variable. A famous example of this methodology is Labov's study of the presence or absence of /r/ in the speech of employees in several department stores in New York City. He gathered hundred of tokens by asking for directions to a product that he knew was sold on the fourth floor. Whenever someone directed him to "The fourth floor," he noted carefully whether or not they pronounced the /r/ in either *fourth* or *floor*.

It is customary for sociolinguistic surveys to record speakers from several different age groups. Often the frequency of different variants differs within the same community according to how old the speaker is. Where there are such differences, they can be used as diagnostics of on-going change that is taking place (a point we return to below).

Recently, sociolinguists have begun to explore even more creative methods for exploring the complex relationship between social and linguistic factors and variation and change. The methods of social dialectology focus exclusively on production of language: what variants do speakers use in different (social or linguistic) contexts? Some work has begun to also ask whether social or linguistic factors have an effect on the way speakers *perceive* language, e.g., do people hear something different if they think they are listening to speakers with different social attributes? Work in this area suggests that this may indeed be the case.

3. Linguistic Constraints on Variation

It is almost always the case that where there is linguistic variation, the main constraints on a variable are linguistic, and not social, factors. Indeed, many linguistic variables are only constrained by other linguistic features. For this reason, it is extremely important for sociolinguists to carefully consider the importance of language-internal factors before beginning to make generalizations about the social conditioning of a variable.

Suppose a pronunciation variant is more frequent in one phonological context than in all others—as Latvian vowel deletion is, following /j/. Suppose, furthermore, that in the sociolinguistic interviews making up a researcher's corpus, there happens to be more examples of vowels following /j/ in the speech of young men than there are the interviews with young women or with older men. If the researcher does not take into account the fact that an immediately preceding /j/ favors vowel deletion in all speakers, s/he might mistakenly think that vowel deletion is more frequent in the speech of men than women and in the speech of young men especially. (The need to control for potentially skewed distributions like this is also a reason why sociolinguists prefer to conduct statistical tests which determine the relative frequency of a variant, rather than rely solely on absolute frequencies or percentages.)

The same principle holds for research on syntactic variables. For example, many languages allow speakers to choose between using a pronoun subject and leaving it out. This is true in such typologically different languages as Mandarin, Spanish, Tamambo (spoken in Vanuatu), and Hebrew. The following example is from the English-based creole, Bislama, spoken in Vanuatu. The pronoun subject is underlined in the first sentence, and the place where it has been omitted is underlined and marked with a zero (Ø) in the second.

4. Ale hem i fasem ae blong hem wetem tawel, ale Ø i putum a, olsem maot blong masket ya insaed long maot blong hem…

'So he tied up his eyes with a towel, then [he] put like the mouth of the gun in his mouth.'

In Bislama, the principal constraints on this alternation are linguistic. First, whether the subject is third person or not (in the third person, the agreement on the predicate is more

distinctive and therefore more informative than it is in first and second person). Second, the relationship between subjects across sentences or clauses. As in (4), speakers are more likely to leave out the pronoun when the subject stays the same across clauses, and especially if it has already been omitted before. (Both these constraints are shared in other languages that allow speakers to omit pronouns, but their relative importance differs in different languages.) Researchers must control for these linguistic variables in order to be able to draw valid conclusions about whether or not social factors are important constraints on the use of pronominal subjects in Bislama. For example, in a corpus of conversational Bislama gathered in northern Vanuatu, women omitted 50% of the pronouns, and men 42%. But when the principal—linguistic—constraints were controlled for (person of the subject; subject continuity across clauses), the difference between women and men proved not to be significant.

A striking example of linguistic constraints on a variable can be seen in the predictable distribution of consonant cluster reduction in English. This variable refers to the manner in which word-final consonant clusters ending in /t/ or /d/ may be simplified in all varieties of spoken English. For example, speakers may realize a monomorphemic word like *west* as /wEst/ or as /wEs/. Similarly, a consonant cluster formed by suffixing the past tense morpheme to a verb stem may be simplified, i.e. *push+ ed > pushed* which may be realized as /pUSt/ or /pUS/. Many studies have examined the distribution of this variable in different varieties of English, and generally linguistic constraints have the greatest effect favoring or disfavoring the reduced or unreduced forms. The following segment plays a major role: the final /t/ or /d/ is more likely to be retained if it can combine with the following segment as a syllable onset in English. So for example, a following vowel (*pushed out*) or following /r/ (*west rim*) favors retention of the final /t/.

This seems to be because the phrase can be reanalyzed as /pUS tout/ or /wEs trIm/. On the other hand, if the following segment is /l/ (*pushed left*) or /k/ (*west corner*) then the reduced variant is favoured. This seems to be because syllable-initial /tl/ and /tk/ are not allowed in English (they are in other languages, so this shows that the constraint is specific to the structure of English and not a general one on human languages). In addition, there seems to be evidence that this variable is constrained by the internal structure of the word itself. Monomorphemic words like *west*, *sand*, and *trend* are more likely to have the final cluster reduced than words where the cluster crosses a morpheme boundary, e.g., *changed*, *traced*, *left* (the past form of *leave*). Both these factors argue strongly for representing the variation as part of the structure of the linguistic system, in the same way that the regular variation between the realization of the past tense morpheme *-ed* as /t/, /d/ and /Id/ (depending on the final consonant of the verb stem) is expressed as part of the grammar of English.

-
-
-

Bibliography

Chambers, J.K., P. Trudgill and N. Schilling-Estes (eds) (2001). *The Handbook of Language Variation and Change*. Oxford: Basil Blackwell. [A collection of original articles on a wide range of topics (theory and practice) in variationist sociolinguistics.]

Eckert, P. (2000). *Linguistic Variation as Social Practice*. Oxford: Basil Blackwell. [Thorough exploration of the connections between qualitative and quantitative analysis of variation. The relationship between gender and social and linguistic practices is especially well developed.]

Eckert, P. & J.R. Rickford (eds) (2001). *Style and Sociolinguistic Variation*. Cambridge: Cambridge University Press. [A collection of papers providing different perspectives on style from sociolinguists and anthropologists. Some are quite technical.]

Holmes, J. (2001). *An Introduction to Sociolinguistics* (2nd ed.). London: Longman. [A clear, well-written text outlining the basics findings in language variation studies. Presents variationist sociolinguistics alongside social and cultural aspects of language, e.g. bilingualism. Suitable for readers from secondary level up.]

Labov, W. (1972). *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press. [An edited collection of some early seminal papers on variation and change. Includes clear exposition of methods and key principles and terms. Some papers suitable for advanced undergraduate students. Also available in French as *Sociolinguistique*.]

Labov, W. (1994). *Principles of Linguistic Change: Internal factors*. Oxford: Basil Blackwell.

Labov, W. (2000). *Principles of Linguistic Change: Social factors*. Oxford: Basil Blackwell. [These volumes are indispensable, but not introductory, texts summarizing four decades of the quantitative study of variation and change and setting out methods for undertaking variationist studies. The first emphasizes linguistic constraints but necessarily also mentions factors like style and age. The second includes some painstaking work teasing out the quantitative effects of different social factors.]

Meyerhoff, M. (2006). *Introducing Sociolinguistics*. London & New York: Routledge. [Introductory text for undergraduates. Contains numerous exercises for readers and expands on connections between sociolinguistics, other branches of linguistics and other social sciences.]

Milroy, L. & M. Gordon (2003). *Sociolinguistics: Method and interpretation*. Oxford: Basil Blackwell. [Variationist methods and findings. Especially strong on the study of variation and change in social networks or groups.]

Nevalainen, T. & H. Raumolin-Brunberg (2003). *Historical Sociolinguistics: Language change in Tudor and Stuart England*. London: Longman. [Invaluable book demonstrating the use of variationist methods on historical corpora and the robustness of generalizations about variation across social class and gender.]

Trudgill, P. 2002. *Sociolinguistic Variation and Change*. Edinburgh: Edinburgh University Press/Georgetown: Georgetown University Press. [A collection of papers focusing on variation. Unusual in focusing on the effects of language and social contact on variation and change, also some comments on creolization and processes of language stabilization.]

Biographical Sketch

Miriam Meyerhoff is Professor of Sociolinguistics at the University of Edinburgh. She has researched the linguistic variation found in a number of languages and in a number of different social contexts in the Pacific and Caribbean with a particular focus on creole languages. She is particularly interested in the intersection between linguistic variation and group identities such as gender, as well as the impact of language contact on long-term processes such as grammaticalization. She has taught in New Zealand, the United States, Vanuatu, and Scotland.