

# Analysing Sociolinguistic Variation

Sali A. Tagliamonte

KEY TOPICS IN SOCIOLINGUISTICS

CAMBRIDGE

This page intentionally left blank

## Analysing sociolinguistic variation

The study of how language varies in social context, and how it can be analysed and accounted for, are the two key goals of sociolinguistics. Until now, however, the actual tools and methods have been largely passed on through 'word of mouth' rather than being formally documented. This is the first comprehensive, 'how-to' guide to the formal analysis of sociolinguistic variation. It shows step-by-step how the analysis is carried out, leading the reader through every stage of a research project from start to finish. Topics covered include fieldwork, data organisation and management, analysis and interpretation, presenting research results and writing up a paper. Practical and informal, the book contains all the information needed to conduct a fully fledged sociolinguistic investigation, and includes exercises, checklists, references and insider tips. It is set to become an essential resource for students, researchers and fieldworkers embarking on research projects in sociolinguistics.

SALIA A. TAGLIAMONTE is Associate Professor of Linguistics at the University of Toronto. Her research focuses on variation and change in the evolution of English. She is author of *African American English in the diaspora: tense and aspect* (with Shana Poplack, 2001).

## KEY TOPICS IN SOCIOLINGUISTICS

Series editor:

Rajend Mesthrie, University of Cape Town.

This new series focuses on the main topics of study in sociolinguistics today. It consists of accessible yet challenging accounts of the most important issues to consider when examining the relationship between language and society. Some topics have been the subject of sociolinguistic study for many years, and are here re-examined in the light of new developments in the field; others are issues of growing importance that have not so far been given a sustained treatment. Written by leading experts, the books in the series are designed to be used on courses and in seminars, and include useful suggestions for further reading and a helpful glossary.

Already published in the series:

*Politeness* by Richard J. Watts

*Language Policy* by Bernard Spolsky

*Discourse* by Jan Blommaert

Forthcoming titles:

*World Englishes* by Rakesh Bhatt and Raj Mesthrie

*Language and Ethnicity* by Carmen Fought

*Bilingual Talk* by Peter Auer

# Analysing Sociolinguistic Variation

SALI A. TAGLIAMONTE



CAMBRIDGE UNIVERSITY PRESS

Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

Cambridge University Press

The Edinburgh Building, Cambridge CB2 2RU, UK

Published in the United States of America by Cambridge University Press, New York

[www.cambridge.org](http://www.cambridge.org)

Information on this title: [www.cambridge.org/9780521771153](http://www.cambridge.org/9780521771153)

© Sali A. Tagliamonte 2006

This publication is in copyright. Subject to statutory exception and to the provision of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published in print format 2006

ISBN-13 978-0-511-16646-4 eBook (Adobe Reader)

ISBN-10 0-511-16646-X eBook (Adobe Reader)

ISBN-13 978-0-521-77115-3 hardback

ISBN-10 0-521-77115-3 hardback

ISBN-13 978-0-521-77818-3

ISBN-10 0-521-77818-2

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for external or third-party internet websites referred to in this publication, and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

For Tara, Shaman, Freya and Dazzian  
Love  
*Mom*





# Contents

*Preface* ix

*Notes on codes and abbreviations* xi

1. Introduction 1
2. Data collection 17
3. The sociolinguistic interview 37
4. Data, data and more data 50
5. The linguistic variable 70
6. Formulating hypotheses/operationalising claims 99
7. The variable rule program: theory and practice 128
8. The how-to's of a variationist analysis 158
9. Distributional analysis 191
10. Multivariate analysis 217
11. Interpreting your results 235
12. Finding the story 254

*Glossary of terms* 263

*References* 267

*Index* 277



# Preface

The variationist approach to sociolinguistics began during the 1960s, when Labov, working with Uriel Weinreich, developed a theory of language change (Weinreich et al. 1968). Thereafter, Labov continued to advance the method and analysis of language variation and change, which today is often referred to as variation theory (e.g. Labov 1963, 1966/1982).

In the 1970s, one of Labov's graduate students at the University of Pennsylvania was Shana Poplack. In 1981, Shana became a professor of Sociolinguistics at the University of Ottawa's Department of Linguistics, the same year I entered the MA programme. I was fortunate to be Shana's student until I completed my Ph.D. dissertation in 1991. Everything you will read in this book has come directly from what has been passed on from this lineage – training, techniques, insights, knowledge, and sheer passion for the field. The entire period from 1981 to 1995 was an invaluable apprenticeship through my studies with Shana and our many collaborations (e.g. Tagliamonte and Poplack 1988, Poplack and Tagliamonte 1989, 1991). I also benefited tremendously from the influence of David Sankoff, whose input to my questions of method and analysis was innumerable.

Most knowledge and learning in variation theory has been acquired like this, passed on through word of mouth, from one researcher to the next (see also Guy 1988: 124). In fact, it has often been noted that the practical details of how to actually do variation analysis are arcane, largely unwritten and, for the most part, undocumented (but see Paolillo 2002). This is precisely why this book was conceived and has now been written. The method needed to be recorded, systematically, thoroughly and straightforwardly.

I had originally intended this book to be completed by the mid-1990s, but academic life is unforgiving for time and relentless for energy. The advantage is that I have had that many more years of experience. Between 1995 and 2005, I have trained some of the next generation of variationist sociolinguists and, as has always been

the case, the students have helped the teacher learn a lot more than what she thought she already knew. Yet there is always room for improvement.

In completing this book, I requested ‘no-holds-barred’ comments from the best methodologists of my colleagues and students. I am indebted to Alex D’Arcy, Ann Taylor, Jennifer Smith and James Walker, who came through with the best feedback one could hope for – an intense amount of red ink. Through the Herculean efforts of my assistant Sonja Molfenter, who took on the unenviable role of ‘book bulldog’ during the revision process, I have taken all their comments into account, and then some. My own method has evolved in just this way, incrementally changing from one research project to the next, one student to the next, in my perpetual efforts to do things more usefully, more efficiently and more transparently. At the same time, the basics endure. The original ideas enshrined in Weinreich et al. (1968), built upon and elaborated by William Labov (Labov 1966/1982) herein are fundamental and pervasive. In sum, this book is simply one user’s tried-and-true manual of best practice.

## Notes on codes and abbreviations

Codes in parentheses refer to the community from which the data come (see abbreviations), followed by a single character speaker code which identifies the individual speaker in each community. In some cases additional information may appear, e.g. the audio-tape number. Abbreviations for communities are: BCK = Buckie; CLB = Cullybackey, Northern Ireland; CMK = Cumnock, Scotland; DVN = Devon, Southeast England; ESR = Ex-Slave Recordings; GYE = Guysborough Enclave; GYV = Guysborough Village, Nova Scotia; KID = a corpus of child language acquisition, England; MPT = Maryport, England; NPR = North Preston, Nova Scotia; OTT = Ottawa; ROP2-4 = data collected in Toronto, Canada, in the years 2002-2004 through the Research Opportunities Programs at the University of Toronto; PVG = Portavogie; ROO = Roots Corpus, data collected in remote communities in Northern Ireland, Lowland Scotland and Northwest England; SAM = Samaná, Dominican Republic; TIV = Tiverton, England; TOR = Toronto, Canada; WHL = Wheatley Hill, England; YRK = York, England. Abbreviations for additional corpora include ST1, ST2 and ST3 - all narrative data sets of Canadian English.



# 1 Introduction

This book is about *doing* variation analysis. My goal is to give you a step-by-step guide which will take you through a variationist analysis from beginning to end. Although I will cover the major issues, I will not attempt a full treatment of the theoretical issues nor of the statistical underpinnings. Instead, you will be directed to references where the relevant points are treated fully and in detail. In later chapters, explicit discussion will be made as to how different types of analysis either challenge, contribute to or advance the basic theoretical issues. This is important for demonstrating (and encouraging) evolution in the field and for providing a sense of its ongoing development. Such a synthetic perspective is also critical for evolving our research in the most interesting direction(s). In other words, this book is meant to be a learning resource which can stimulate methodological developments, curriculum development as well as advancements in teaching and transmission of knowledge in variation analysis.

## **WHAT IS VARIATION ANALYSIS?**

---

Variation analysis combines techniques from linguistics, anthropology and statistics to investigate language use and structure (Poplack 1993: 251). For example, a seven-year-old boy answers a teacher's question by saying, 'I don't know nothing about that!' A middle-aged woman asks another, 'You got a big family?' Are these utterances instances of dialect, slang, or simply performance errors, mistakes? Where on the planet were they spoken, why, by people of what background and character, in which sociocultural setting, under what conditions? How might such utterances be contextualised in the history of the language and with respect to its use in society? This book provides an explicit account of a method that can answer these questions, a step-by-step 'user's guide' for the investigation of language use and structure as it is manifested in situ.

At the outset, however, I would like to put variationist sociolinguistics in perspective. First, what is the difference between sociolinguistics and linguistics? Further, how does the variationist tradition fit in with the field of sociolinguistics as a whole?

## LINGUISTICS

---

The enterprise of linguistics is to determine the properties of natural language. Here, the aim is to examine individual languages with the intention of explaining why the whole set of languages are the way they are. This is the search for a theory of universal grammar. In this process, the analyst aims to construct a device, a grammar, which can specify the grammatical strings of one language, say English or Japanese, but which is also relevant for the grammar of any natural language. In this way, linguistics puts its focus on determining what the component parts and inner mechanism of languages are. The goal is to work out 'the rules of language X' – whether that language is English, Welsh, Igbo, Inuktitut, Niuean, or any other human language on the planet.

The type of question a linguist might ask is: 'How do you say X?' For example, if a linguist was studying Welsh, she would try to find a native speaker of Welsh and then she would ask that person, How do you say 'dog' in Welsh? How do you say 'The child calls the dog', 'The dog plays with the children', etc. This type of research has been highly successful in discovering, explaining and accounting for the complex and subtle aspects of linguistic structure. However, in accomplishing this, modern theoretical models of language have had to exclude certain things, consigning them to the lexical, semantic or pragmatic components of languages, or even outside of language altogether. For example, in a recent syntactic account of grammatical change, Roberts and Rousseau (2003: 11) state:

Of course, many social, historical and cultural factors influence speech communities, and hence the transmission of changes (see Labov 1972c, 1994). From the perspective of linguistic theory, though, we abstract away from these factors and attempt, as far [sic] the historical record permits, to focus on change purely as a relation between grammatical systems.

In this way, linguistic theory focuses on the structure of the language. It does not concern itself with the context in which the language is learned and, more importantly, it does not concern itself with



the way the language is used. Only in more recent forays have researchers begun to make the link between variation theory and syntactic theory (e.g. Beals et al. 1994, Meechan and Foley 1994, Cornips and Corrigan 2005).

## **SOCIOLINGUISTICS**

---

Sociolinguistics argues that language exists in context, dependent on the speaker who is using it, and dependent on where it is being used and why. Speakers mark their personal history and identity in their speech as well as their sociocultural, economic and geographical coordinates in time and space. Indeed, some researchers would argue that, since speech is obviously social, to study it without reference to society would be like studying courtship behaviour without relating the behaviour of one partner to that of the other. Two important arguments support this view. First, you cannot take the notion of language X for granted since this in itself is a social notion in so far as it is defined in terms of a group of people who speak X. Therefore, if you want to define the English language you have to define it based on the group of people who speak it. Second, speech has a social function, both as a means of communication and also as a way of identifying social groups.

Standard definitions of sociolinguistics read something like this:

the study of language in its social contexts and the study of social life through linguistics (Coupland and Jaworski 1997: 1)

the relationship between language and society (Trudgill 2000: 21)

the correlation of dependent linguistic variables with independent social variables (Chambers 2003: ix)

However, the many different ways that society can impinge on language make the field of reference extremely broad. Studies of the various ways in which social structure and linguistic structure come together include personal, stylistic, social, sociocultural and sociological aspects. Depending on the purposes of the research, the different orientations of sociolinguistic research have traditionally been subsumed by one of two umbrella terms: 'sociolinguistics' and 'the sociology of language'. A further division could also be made between qualitative (ethnography of communication, discourse analysis, etc.) and quantitative (language variation and change) approaches. Sociolinguistics tends to put emphasis on *language* in social context,

whereas the sociology of language emphasises the *social* interpretation of language. Variation analysis is embedded in sociolinguistics, the area of linguistics which takes as a starting point the rules of grammar and then studies the points at which these rules make contact with society. But then the question becomes: How and to what extent? Methods of analyses, and focus on linguistics or sociology, are what differentiate the strands of sociolinguistics. From this perspective, variation analysis is inherently linguistic, analytic and quantitative.

## ***VARIATIONIST SOCIOLINGUISTICS***

---

Variationist sociolinguistics has evolved over the last nearly four decades as a discipline that integrates social and linguistic aspects of language. Perhaps the foremost motivation for the development of this approach was to present a model of language which could accommodate the paradoxes of language change. Formal theories of language were attempting to determine the structure of language as a fixed set of rules or principles, but at the same time language changes perpetually, so structure must be fluid. How does this happen? The idea that language is structurally sound is difficult to resolve with the fact that languages change over time.

structural theories of language, so fruitful in synchronic investigation, have saddled historical linguistics with a cluster of paradoxes, which have not been fully overcome. (Weinreich et al. 1968: 98)

Unfortunately, because it is such a expansive field of research, sociolinguistics often comes across as either too restricting to social categories such as class, sex, style, geography (the external factors), or too restricting to linguistic categories such as systems, constraints and rate of change (the structural factors). In fact, when sociolinguistic research using variationist methods has shown a focus on the linguistic system, as opposed to the social aspects of the individual and context, it has garnered considerable criticism (e.g. Cameron 1990, Rickford 1999, Eckert 2000). More than anything this highlights the bi-partite underpinnings of the field (Milroy and Gordon 2003: 8). When attempting to synthesise both internal and external aspects of language, the challenge will always be to fully explore both. While this will likely always be tempered by researchers' own predilections, it is also the case that the research questions, data and findings may naturally lead to a focus on one domain over the other. Having said

all this, the variationist enterprise is essentially, and foremost, the study of the interplay between variation, social meaning and the evolution and development of the linguistic system itself.

Indeed, as Weinreich et al. (1968: 188) so well described in their foundational work,

Explanations of language which are confined to one or the other aspect – linguistic or social – no matter how well constructed, will fail to account for the rich body of regularities that can be observed in empirical studies of language behaviour . . .

This ‘duality of focus’ has been fondly described more recently by Guy (1993: 223) as follows:

One of the attractions – and one of the challenges – of dialect research is the Janus-like point-of-view it takes on the problems of human language, looking one way at the organisation of linguistic forms, while simultaneously gazing the other way at their social significance.

In my view, variationist sociolinguistics is most aptly described as the branch of linguistics which studies the foremost characteristics of language in balance with each other – linguistic structure and social structure; grammatical meaning and social meaning – those properties of language which require reference to both external (social) and internal (systemic) factors in their explanation.

Therefore, instead of asking the question: ‘How do you say X?’ as a linguist might, a sociolinguist is more likely not to ask a question at all. The sociolinguist will just let you talk about whatever you want to talk about and listen for all the ways you say X.

### Note

There is a distinct ‘occupational hazard’ to being a sociolinguist. You will be in the middle of a conversation with someone and you will notice something interesting about the way he or she is saying it. You will make note of the form. You will wonder about the context. You may notice a pattern. All of a sudden you will hear that person saying to you, ‘Are you listening to me?’ and you will have to say, ‘I was listening so intently to how you were saying it that I didn’t hear what you said!’

The essence of variationist sociolinguistics depends on three facts about language that are often ignored in the field of linguistics. First, the notion of ‘orderly heterogeneity’ (Weinreich et al. 1968: 100), or what Labov (1982: 17) refers to as ‘normal’ heterogeneity; second, the fact that language changes perpetually; and third, that language

conveys more than simply the meaning of its words. It also communicates abundant non-linguistic information. Let us consider each of these in turn.

## **ORDERLY HETEROGENEITY**

---

Heterogeneity is essentially the observation that language varies. Speakers have more than one way to say more or less the same thing. Variation can be viewed across whole languages, e.g. French, English, Spanish, etc. In this case, variation would be in the choice of one language or the other by bilingual or multilingual speakers. However, linguistic variation also encompasses an entire continuum of choices ranging from the choice between English or French, for example, to the choice between different constructions, different morphological affixes, right down to the minute microlinguistic level where there are subtle differences in the pronunciation of individual vowels and consonants. Importantly, this is the normal state of affairs:

The key to a rational conception of language change – indeed, of language itself – is the possibility of describing orderly differentiation in a language serving a community . . . It is *absence* of structural heterogeneity that would be dysfunctional. (Weinreich et al. 1968: 100–1)

Furthermore, heterogeneity is not random, but patterned. It reflects order and structure within the grammar. Variation analysis aims to characterise the nature of this complex system.

## **LANGUAGE CHANGE**

---

Language is always in flux. The English language today is not the same as it was 100 years ago, or 400 years ago. Things have changed. For example, *ain't* used to be the normal way of doing negation in English, but now it is stigmatised. Another good example is *not*. It used to be placed after the verb, e.g. *I know not*. Now it is placed before the verb, along with a supporting word, *do*, as in *I do not know*. Double negation, e.g. *I don't know nothing*, is ill-regarded in contemporary English. Not so in earlier times. Similarly, use of the ending *-th* for simple present was once the favoured form, e.g. *doth*, not *do*, and pre-verbal periphrastic *do*, e.g. *I do know*, and use of the comparative ending *-er*, e.g. *honester*, not *more honest*, used to be much more frequent. Such examples are

easily found in historical corpora such as the Corpus of Early English Correspondence (Nevalainen and Raumolin-Brunberg 2003).

Variation analysis aims to put linguistic features such as these in the context of where each one has come from and where it is going – how and why.

## ***SOCIAL IDENTITY***

---

Language serves a critical purpose for its users that is just as important as the obvious one. Language is used for transmitting information from one person to another, but at the same time a speaker is using language to make statements about who she is, what her group loyalties are, how she perceives her relationship to her hearers, and what sort of speech event she considers herself to be engaged in. The only way all these things can be carried out at the same time is precisely because language varies. The choices speakers make among alternative linguistic means to communicate the same information often conveys important extralinguistic information. While you can inevitably identify a person's sex from a fragment of their speech, it is often nearly as easy to localise her age and sometimes even her socio-economic class. Further, depending on one's familiarity with the variety, it can be relatively straightforward to identify nationality, locality, community, etc. For example, is the following excerpt from a young person or an old person?

I don't know, it's jus' stuff that really annoys me. And I jus' like stare at him and jus' go ... like, "huh". (YRK98/S014c)

How about the following? Male or female? Old or young?

It was sort-of just grass steps down and where I dare say it had been flower beds and goodness-knows-what ... (YRK/v)

I am willing to bet it was relatively easy to make these decisions and to do so correctly. The first is a young woman, aged eighteen. The second is a female, aged seventy-nine.

## ***KEY CHARACTERISTICS OF VARIATIONIST SOCIOLINGUISTICS***

---

Given these three aspects of language – inherent variation, constant change and pervasive social meaning – variationist sociolinguistics rests its method and analysis on a number of key concepts.

## THE 'VERNACULAR'

---

A specific goal of variationist methodology is to gain access to what is referred to as the 'vernacular'. The vernacular has had many definitions in the field. It was first defined as 'the style in which the minimum attention is given to the monitoring of speech' (Labov 1972c: 208). Later discussions of the vernacular reaffirmed that the ideal target of investigation for variation analysis is 'every day speech' (Sankoff 1974, 1980: 54), 'real language in use' (Milroy 1992: 66) and 'spontaneous speech reserved for intimate or casual situations' (Poplack 1993: 252) – what can simply be described as informal speech.

Access to the vernacular is critical because it is thought to be the most systematic form of speech. Why? First, because it is assumed to be the variety that was acquired first. Second, because it is the variety of speech most free from hypercorrection or style-shifting, both of which are considered to be later overlays on the original linguistic system. Third, the vernacular is the style from which every other style must be calibrated (Labov 1984: 29). As Labov originally argued (1972c: 208), the vernacular provides the 'fundamental relations which determine the course of linguistic evolution'.

The vernacular is positioned maximally distant from the idealised norm (Milroy 1992: 66, Poplack 1993: 252). Once the vernacular baseline is established, the multi-dimensional nature of speech behaviour can be revealed. For example, Bell (1999: 526) argues that performance styles are defined by normative use. Thus, the unmonitored speech behaviour of the vernacular enables us to tap in to the broader dimensions of the speech community. In other words, the vernacular is the foundation from which every other speech behaviour can be understood.

### Note

Many of my students report that their room-mates switch into their vernacular when talking to their mother on the phone. However you will notice it shine through whenever a person is emotionally involved, e.g. excited, scared, angry, moderately drunk, etc. Listen out for it!

## THE SPEECH COMMUNITY

---

In order to 'tap the vernacular' (Sankoff 1988b: 157), a vital component of variation analysis requires that the analyst immerse herself in

the speech community, entering it both as an observer and a participant. In this way, the analyst may record language use in its socio-cultural setting (e.g. Labov et al. 1968, Trudgill 1974, Milroy 1987, Poplack 1993: 252). Due to its focus on unmonitored speech behaviour, this methodology has succeeded in overcoming many of the analytical difficulties associated with intuitive judgements and anecdotal reporting used in other paradigms (Sankoff 1988b). This is crucial in the study of non-standard varieties, as well as ethnic, rural, informal and other less highly regarded forms of language, where normative pressure typically inhibits the use of vernacular forms.

For example, when you hear people use utterances such as (i) 'I ain't gotta tell you anything', certain social judgements will surely arise. Whatever judgements come to mind are based on hypotheses that arise from interpreting the various linguistic features within these utterances. What are those features? Most people, when asked why someone sounds different, will appeal to their 'accent', their 'tone of voice' or their 'way of emphasising words'. However, innumerable linguistic features of language provoke social judgements.

One way to explore this is to contemplate the various ways the utterance in (i) could have been spoken, as in (1). Each possible utterance has its own social value, ranging from the highly vernacular to standard. Notice, too, how each feature of language varies in particular ways. *Ain't* appears to vary with *haven't* and possibly *don't*. *Gotta* appears to vary with *have to* as well as *got to*. *Nothing* varies with *anything*. In this way, each item alternates with a specific set – different ways of saying the same thing.

(1)

- a. I ain't gotta tell you nothing/anything
- b. I haven't gotta tell you nothing/anything
- c. I don't have to tell you nothing/anything

The linguistic items which vary amongst themselves with the same referential meaning are the 'variables' which are the substance of variation analysis. But the next question becomes: How do you determine what *truly* varies with what?

## FORM/FUNCTION ASYMMETRY

---

The identification of 'variables' in language use rests on a fundamental view in variation analysis – the possibility of multiple forms for the same function. Do all the sentences in (1) mean the same thing? Some

linguists might assume that different forms can never have identical function. In variation analysis, however, it is argued that different forms such as these can indeed be used for the same function, particularly in the case of ongoing linguistic change. In other words, there is a basic recognition of instability in linguistic form/function relationships (Poplack 1993: 252) and, further, that differences amongst competing forms may be neutralised in discourse (Sankoff 1988b: 153). Where functional differences are neutralised is always an empirical question. It must first be established what varies with what and how. Notice that you can't say *I ain't haven't to tell you nothing*. Why? The goal of variation analysis is to pinpoint the form/function overlap and explain how this overlap exists and why.

## LINGUISTIC VARIABLES

---

Different ways of saying more or less the same thing may occur at every level of grammar in a language, in every variety of a language, in every style, dialect and register of a language, in every speaker, often even in the same sentence in the same discourse. In fact, variation is everywhere, all the time. Consider the examples in (2) to (10), all of which are taken from the York English Corpus (YRK), which represents the variety spoken in the city of York in north England (Tagliamonte 1998).

Phonology/morphology, variable (t,d):

(2)

I did a college course when I lefØ school actually, but I left it because it was business studies. (YRK/h)

Phonology/morphology, variable (ing):

(3)

We were having a good time out in what we were doin'. (YRK/E)

Morphology, variable (ly):

(4)

You go to Leeds and Castleford, they take it so much more seriously . . . They really are, they take it so seriousØ. (YRK/T)

Tense/aspect, variable future temporal reference forms:



(5)

... I think she's **gonna** be pretty cheeky. I think she **'ll be** cheeky. (YRK/O)

Modal auxiliary system, deontic modality:

(6)

'**I've got to** cycle all the way back and then this afternoon I'll be cycling back up again!' ... You **have to** keep those thoughts err thoughts to yourself. (YRK/X)

Intensifiers:

(7)

I gave him a **right** dirty look ... and I gave him a **really** dirty look. (YRK/O)

Syntax/semantics, variable stative possessive meaning:

(8)

He's **got** bad-breath; he **has** smelly feet. (YRK/i)

Syntax, agreement:

(9)

She **were** a good worker. She **was** a helluva good worker. (YRK/¥)

Discourse/pragmatics, quotative use:

(10)

He just **said** 'Fine, go.'

It **was like** 'It 's gonna cost me a fortune!'

I **thought** 'Ah,' and that was it. (YRK/d)

How can such variability become interpretable? It is necessary to refer to more than just social meaning. Such variation might be explained by external pragmatic factors; however, it is more often the case that variation such as this has complex social, linguistic and historical implications. In the case of variable (ly), stative possessive *have got*, the modal auxiliary system, intensifiers and others, variation amongst forms can be traced back to longitudinal change in the history of the English language. In the case of adverb placement and variable agreement, synchronic patterns may address issues pertaining to the configuration of phrase structure, feature checking and other issues of theoretical importance. Indeed, much of the work on historical syntax has highlighted the complexity of how linguistic structures evolve in the process of grammatical change (e.g. Kroch 1989, Warner 1993, Taylor 1994, Pintzuk 1995).

## **THE QUANTITATIVE METHOD**

---

Perhaps the most important aspect of variation analysis that sets it apart from most other areas of linguistics, and even sociolinguistics, is its quantitative approach. The combination of techniques employed in variation analysis forms part of the ‘descriptive-interpretative’ strand of modern linguistic research (Sankoff 1988b: 142–3). Studies employing this methodology are based on the observation that speakers make choices when they use language and that these choices are discrete alternatives with the same referential value or grammatical function. Furthermore, these choices vary in a systematic way and as such they can be quantitatively modelled (Labov 1969a, Cedergren and Sankoff 1974; see also more recent re-statements in Young and Bayley 1996: 254, Poplack and Tagliamonte 2001: 88). This is perhaps most candidly put by Sankoff (1988b: 151):

whenever a choice can be perceived as having been made in the course of linguistic performance, and where this choice may have been influenced by factors such as the nature of the grammatical context, discursive function of the utterance, topic, style, interactional context or personal or sociodemographic characteristics of the speaker or other participants, then it is difficult to avoid invoking notions and methods of statistical inference, if only as a heuristic tool to attempt to grasp the interaction of the various components in a complex situation.

The advantage of the quantitative approach lies in its ability to model the simultaneous, multi-dimensional factors impacting on speaker choices, to identify even subtle grammatical tendencies and regularities in the data, and to assess their relative strength and significance. These measures provide the basis for comparative linguistic research. However, such sophisticated techniques are only as good as the analytic procedures upon which they are based:

The ultimate goal of any quantitative study . . . is not to produce numbers (i.e. summary statistics), but to identify and explain linguistic phenomena. (Guy 1993: 235)

## **THE PRINCIPLE OF ACCOUNTABILITY**

---

According to Labov (1972c: 72), ‘the most important step in sociolinguistic investigation is the correct analysis of the linguistic variable’. ‘Correct’ in this case means ‘accountable’ to the data. In variation

analysis, accountability is defined by the ‘principle of accountability’, which holds that every variant that is part of the variable context, whether the variants are realised or unrealised elements in the system, must be taken into account. In other words, you cannot simply study the variant forms that are new, interesting, unusual or non-standard – *ain’t*, for example, or *got*. You must also study the forms with which such features vary in all the contexts in which either of them would have been possible. In the case of *ain’t*, this would mean all the cases where *ain’t* is used as well as all other negation variants with the same referential value as *ain’t*, e.g. *I haven’t got nothing* or perhaps even *I don’t got nothing* – whatever occurs in the same context. By definition, an accountable analysis demands of the analyst an exhaustive report for every case in which a variable element occurs out of the total number of environments where the variable element could have occurred, but did not. In Labov’s (1972c: 72) words, ‘report values for every case where the variable element occurs in the relevant environments as we have defined them’. ‘As we have defined them’ is the important point here. What does this mean?

## ***CIRCUMSCRIBING THE VARIABLE CONTEXT***

---

How does the analyst determine the variants of a variable and the contexts in which they vary? This procedure is most accurately characterised as a ‘long series of exploratory manoeuvres’ (Labov 1969a: 728–9):

1. Identify the total population of utterances in which the feature varies. Exclude contexts where one variant is categorical.
2. Decide on how many variants can be reliably identified. Set aside contexts that are indeterminate, neutralised, etc.

These manoeuvres accentuate that variation analysis is not interested in individual occurrences of linguistic features, but requires systematic study of the recurrent choices an individual makes (Poplack and Tagliamonte 2001: 89). Analysis of these recurrent choices enables the analyst to ‘tap in’ to an individual’s use of the targeted forms. In this case, a ‘pattern’ refers to ‘a series of parallel occurrences (established according to structural and/or functional criteria) occurring at a non-negligible rate in a corpus of language use’ (Poplack and Meechan 1998: 129). So, now the question is: How do you find the patterns?

## TESTING HYPOTHESES

---

Labov's (1969a: 729) third exploratory manoeuvre is to 'identify all the sub-categories which would reasonably be relevant in determining the frequency' of forms. These are the underlying patterns, the internal linguistic contexts that are hypothesised to influence the choice of one variant over another. How does one find them? Sometimes these are discovered by scouring the literature, both synchronic and diachronic. Sometimes they 'emerge from the ongoing analysis as a result of various suspicions, inspections, and analogies' (Labov 1969a: 729). Sometimes they are stumbled upon by chance in the midst of analysis and a 'Eureka!' experience unfolds. More often, the very worst days of variation analysis come when you are in the midst of reams of statistical analyses and data and numbers, and you just can't see the forest for the trees! As long as one's practice has been 'carried out with a degree of accuracy and linguistic insight', Labov promises that 'the end result is a set of regular constraints which operate upon every group and almost every individual' (Labov 1969a: 729). Indeed, it never ceases to amaze me what patterns underlie linguistic variables that one has no inkling of in the beginning.

### Note

While I was writing my dissertation, I came to a particularly impassable dead end in my analysis. I could not see any patterns! In desperation, I wailed at one of my mentors, 'There are just no patterns at all.' The response was empathetic, but firm: 'Take it from an old variationist like me – there will be patterns. Keep looking.' And, of course, there were.

Once it can be established that a variable exists in a body of materials, the variationist sociolinguist will embark on the long process of studying the feature: circumscribing the variable context, extracting the relevant data from corpora, coding the material according to reasoned hypotheses gleaned from the diachronic and synchronic literature, and then analysing and interpreting the results *in situ*.

The question inevitably arises: Why use variation analysis? My answer is this. It is an area of the discipline that involves 'real' language as it is being used, so it is inherently hands-on and practical; it employs a methodology that is replicable and 'accountable' to the data; it provides you with the 'tools' to analyse language, not simply on an item-by-item basis, but at the level of the underlying system. Finally, variation analysis puts language in context, socially,

linguistically, synchronically and diachronically. In the end, by conducting a variation analysis you get closer to knowing what language is and what human beings are all about.

## **ORGANISATION AND LOGIC OF THIS BOOK**

---

The book is organised so that the chapters take the reader from the first stages of research right through to the last ones. The chapters build from simple observations, to basic conceptual initiatives, to training procedures. I then move to general research problems and issues and gradually turn to explaining how to resolve complex data handling, computational and linguistic problems. In the final chapters, the focus turns to performing analytic techniques and developing interpretation skills. Examples from my own research demonstrate problem-solving at each stage in the research process. This presents you not only with the techniques of variation analysis, but also with the *process* through which it unfolds. Each chapter ends with an exercise highlighting the topic just covered.

Most of the examples come from two linguistic variables which are the most well known and the most extensively studied in the field – variable (t,d) and variable (ing).

In order to tap into the structured heterogeneity that is rich in living language, it is necessary to gain access to language in use, whether it is in the literature, in the media or, as is most typical of the variationist approach, in the street. For the latter undertaking, it is necessary to go out of the office, beyond the anecdotal, and into the speech community. Fieldwork and data collection will be described in Chapter 2.

### **Exercise 1: Becoming aware of linguistic variation**

The purpose of this exercise is to develop your ability to observe linguistic variability. In the process you will begin to develop a sociolinguist's ear and eye.

Find some language material. Any data will do, e.g. an audio-tape, a video, a TV show, a newspaper, a novel, an email message, an MSN conversational history. Then examine it – carefully.

Consider features from different areas of grammar (e.g. phonology, morphology, syntax, discourse). Notice morphological, syntactic or discursive alternations, e.g. zero plurals, zero possessives, missing prepositions, articles, discourse markers, variation in quotative use,

syntactic structure, etc. Provide an inventory of different variants that occur in the data. Illustrate intra-speaker alternation of forms, as in (i):

- (i) I mean I was **real** small and everything you-know **really** tiny built . . . (YRK/O)

Summarise the nature of the features you have identified. Are there any features that are unfamiliar to you? Are there features that are typical of older rather than younger speakers? Male more than female, or vice versa? Standard vs non-standard, etc.? Do they vary across your sample? Is one variant predisposed to certain sectors of the population over the other? How? Also make note of the types of contexts in which each variant occurs. Can you spot any trends?

When you make a linguistic observation from a data set, always back it up with an example. Further, ensure that the example is referenced to the location in the original data sample (i.e. speaker number and line number), audio-tape (counter number) or whatever is suitable for the data you are examining.

### Note

I will typically take the front page of the newspaper from the day I present this topic to a class and use it to illustrate how normal variation in language really is. There are usually some good examples. Try it!