1 Introduction

This book is about *doing* variation analysis. My goal is to give you a manual 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 theoretical issues. This is important for demonstrating (and encouraging) evolution in the field and for capturing a sense of its ongoing development. Such a synthetic perspective is also critical for evolving our research in the most fruitful direction(s). This book is meant to be a learning resource which can stimulate methodological progression, curriculum development as well as advancements in teaching and transmission of knowledge in variation analysis. With any luck new discoveries will be made.

What Is Variation Analysis?

Variation analysis goes by different names, sometimes it is called 'Labovian sociolinguistics' after its founder William Labov; another term is 'variationist sociolinguistics', yet another is 'language variation and change', as in the name of this subdiscipline's major journal. In this book, I will use these terms somewhat interchangeably; however, the emerging term encompassing worldwide developments is variation linguistics and 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?' An octogenarian might say, 'I did see it.' 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, how does the variationist tradition fit in with the field of sociolinguistics as a whole? What is its relationship to linguistics?

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 unified theory of grammar which can specify the permissible rules 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, they would try to find a fluent speaker of the language, and then they 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"?' and so on. 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 set aside certain aspects, consigning them to the lexical, semantic, or pragmatic components of languages, or even outside of language altogether in the socio-stylistic components of its use. For example, in a 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.

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 is not interested in the way the language is used. However, see an early review of attempts at rapprochement in phonological theory in Coetzee and Pater (2011). Only in the late 1990s and into the 2000s have researchers begun to make the link between variation theory and syntactic theory (e.g. Beals et al., 1994; Meechan & Foley, 1994; Cornips & Corrigan, 2002; Adger & Smith, 2005).

Sociolinguistics

Sociolinguistics argues that language exists in context, influenced by the individual who is using it, and dependent on where it is being used and why. Individuals 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 language 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. As Joseph (2020) argued, the 'evaluation problem' of Weinreich et al. (1968:183–187) entails that 'Someone has to do the evaluating, and someone has to produce a word or an utterance or a piece thereof that can be evaluated, and this means that change is not just something in an isolated individual but involves at least two people. It is inherently social in nature, as a result, and requires contact between speakers.'

Two important arguments support the integral social role of language. First, you cannot take the notion of language X for granted since this is a social notion insofar as it is defined in terms of a group of people who speak X. Therefore, if you want to describe the English language you must define it based on the group of people who speak it. Second, speech has a social function, both as a means of communication and 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 & 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 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 in the 1960s and 1970s was subsumed by one of two umbrella terms: 'sociolinguistics' and 'the sociology of language' (Fasold, 1984; 1990). A further division could also be made between qualitative (ethnography of communication, discourse analysis, and so on) and quantitative (language variation and change) approaches. Sociolinguistics tends to put emphasis on *language* in social context, whereas sociology of language puts emphasis on *society* and 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 their links 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 different subdisciplines of sociolinguistics. From this perspective, variation analysis is inherently linguistic, analytic, and quantitative.

Variationist Sociolinguistics

Variationist sociolinguistics has evolved since the 1960s 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 because language changes perpetually, the structure must be fluid. How does this happen? The idea that language is structurally sound is difficult to reconcile 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).

Change in language does not happen in a linguistic vacuum. Because it is used by human beings in a social world, there is also a need to consider the social world. This interface between language as a system and language as a social phenomenon makes sociolinguistics an unusually expansive field of research, with researchers having a myriad of unique foci. Sociolinguistics often comes across as either too restricting to social categories, using such categorisations as class, gender, style, geography (external, social, factors), or too restricting to linguistic categories, using concepts such as systems or complexity (internal, linguistic, factors). When variationist methods have focused on the linguistic system, as opposed to the social aspects of the individual and context, it has garnered considerable critique (e.g. Cameron, 1990; Rickford, 1999; Eckert, 2000), restating the bipartite underpinnings of the field (Milroy & Gordon, 2003:8). When attempting to synthesise both internal and external aspects of language, the challenge will always be how to explore both without compromising one or the other. 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 the complex 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) stated:

> 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 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 they are saying it. You will make note of the form. You will wonder about the context. You may notice a pattern. Suddenly, 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 rests 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 is always changing; 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 the observation that there is variability in language. Individuals have 'more than one way to say more or less the same thing', that is, accomplish the same function. Variation can be viewed across whole languages, in the choice of one language or the other by bilingual or multilingual individuals, for example French, Tamil, Inuktitut. However, linguistic variation also be observed across an entire continuum of choice types ranging between different word orders, morphological affixes, constructions, right down to the minute microlinguistic level where there are subtle differences in the pronunciation of individual vowels, consonants, intonation contours, and tone. 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., 1968:100–101). Furthermore, heterogeneity is crucially not random, but patterned. It reflects order and structure within the grammar. Variation analysis aims to characterise 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. 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 follow the verb (e.g. *I know not*). Now it precedes the verb, along with a supporting word, *do* (e.g. *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; see studies of historical corpora such as the Corpus of Early English Correspondence (Nevalainen & 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 – with a focus on 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 it is used by individuals to make statements about who they are, what their group loyalties are, how they perceive their relationship to their hearers, and what sort of speech event they consider themselves 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 individuals make among alternative linguistic means to communicate the same information often conveys important extralinguistic information. While you can sometimes identify a person's gender from a fragment of their speech, it is often nearly as easy to identify their age and even their socioeconomic class, but these judgements can be misleading. Further, depending on one's familiarity with the variety, it can be relatively straightforward to identify nationality, locality, community, etc. For example, are the following excerpts from the late twentieth century 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'.

How about the following? Man or woman? 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.

> It was just a fun experience in general, like, the experience of like, you know, debating random problems and stuff.

To a certain point sweeping decisions on social judgements may be accurate. The first is a young woman, aged 30 in 2018 (YRK 2018, spickering, woman, 30).

The second is a woman, aged 79 in 1997 (York, UK, rbaker, woman, 79). The third is a man, aged 19 in 2021 (TOR 2021, rbarman, man, 19).

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 in sociolinguistics as 'the style in which the minimum attention is given to the monitoring of speech' (Labov, 1972d:208). Later characterisations of the vernacular reaffirmed that the ideal target of investigation for variation analysis is 'everyday speech' (Sankoff, G., 1974:54), 'real language in use' (Milroy, 1992:66), and 'spontaneous speech reserved for intimate or casual situations' (Poplack, 1993:252) – what can 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).

The position of the vernacular is pivotal, positioned maximally distant from the idealised norm (Milroy, 1992:66; Poplack, 1993:252; Poplack et al., 2015). Once the vernacular baseline is established, the multi-dimensional nature of speech behaviour can be revealed. Bell (1999:526) argues that performance styles are defined by normative use, making unmonitored speech the focus for taping the dimensions of the speech community. Moreover, as Labov (1972d:208) argued, the vernacular provides the 'fundamental relations which determine the course of linguistic evolution'. The vernacular is the foundation from which every other speech behaviour can be understood, and in which change in progress must be situated.

NOTE

Many of my students report that their roommates switch into their vernacular when talking to their family on the phone. You will also notice it shining through whenever a person is emotionally involved (e.g. excited, scared, angry, moderately drunk). Listen out for it!

The Speech Community

To 'tap the vernacular' (Sankoff, D., 1988b:157), a vital component of variation analysis, analysts are required to immerse themselves in the speech community, entering it both as an observer and as a participant. In this way, analysts may record language use in its sociocultural setting (e.g. Labov et al., 1968; Trudgill, 1974; Milroy, 1987; Poplack, 1993:252). This methodology's focus on unmonitored speech behaviour has allowed it to overcome many of the analytical difficulties associated with intuitive judgements and anecdotal reporting use in other paradigms (Poplack, 1980; Sankoff, D., 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 inhibits the use of vernacular forms.

For example, when you hear people use utterances such as: 'I ain't gotta tell you anything', certain social judgements may 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 cited above *could have been* said, (1).

(1) a. I ain't gotta tell you nothing/anything.
b. I haven't got to tell you nothing/anything.
c. I don't have to tell you nothing/anything.
d. I don't need to tell you nothing/anything.

Each possible utterance has its own social value, ranging from the highly vernacular to standard. Notice, too, how each feature of language varies in its own 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* and *need to*. *Nothing* varies with *anything*. In this way, each item alternates with a specific set – different ways of saying the same thing.

The linguistic items which vary amongst themselves with the same referential meaning constitute the set of linguistic items, the linguistic variable, 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 to achieve the same function. Do all the sentences in (1) mean the same thing? Some linguists might assume that different forms can never have identical meaning. In variation analysis, however, it is argued that different options such as these can indeed be used interchangeably for

the same function, particularly in the case of ongoing linguistic change. There is a basic recognition of instability in linguistic form–function relationships (Poplack, 1993:252; 2018) and, further, that differences amongst competing forms may be neutralised in discourse (Sankoff, D., 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, in every individual, and often even in the same interaction, discourse, and sentence. 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),[1] which documents the variety spoken in the city of York in the north of England (Tagliamonte, 1998).

(2) 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, kdilks, woman, 26)

(3) Phonology/morphology, variable (-ing):
We were *having* a good time out in what we were *doin'*. (YRK, nheath, woman, 20)

(4) Morphology, variable (-ly):
You go to Leeds and Castleford, they take it so much more *seriously* … They really are, they take it so *seriousØ*. (YRK, sdonaldson, *woman*, *41*)

(5) Tense/aspect, variable future temporal reference forms:
… I think she'*s gonna* be pretty cheeky. I think she'*ll be* cheeky. (YRK, kyoung, woman, 31)

(6) Modal auxiliary system, deontic modality:
'*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, rslater, man, 59)

(7) Intensifiers:
I gave him a *right* dirty look … and I gave him a *really* dirty look. (YRK, kyoung, woman, 31)

(8) Syntax/semantics, variable stative possessive meaning:
He'*s got* bad-breath; he *has* smelly feet. (YRK, cbiggs, woman, 33)

[1] Unless otherwise indicated, all examples labelled 'YRK' come from the 1997 corpus (Tagliamonte, 1996–1998). All names in the examples are pseudonyms, except those from the KID corpus.

[2] The use of the template 'variable (x)' for linguistic variables is a labelling practice from my own work. I have used this nomenclature in this book so that readers can identify the linguistic variables under discussion. In some cases, I have not labelled all the potential linguistic variables observed, since they have not been studied yet.

(9) Syntax, agreement:
She *were* a good worker. She *was* a hell of a good worker. (YRK, rfielding, 70 in 1986)

(10) Discourse/pragmatics, quotative use:
a. She *was like*, 'What are they saying?' I *was like*, 'We need to leave now.' And she *goes*, 'Why?' (YRK 2018, vevans, woman, age 20)
b. I *thought*, 'That's – that's my dog.' (YRK, woman, tlaxton, age 48)
c. And we *said*, 'No.' And then Ned *said*, 'Would you like me to go to a cash point?' (YRK 2013, jjubb, man, age 19)

How can such alternation become interpretable? It is necessary to refer to more than just social meaning. Such variation might be explained by external pragmatic factors; however, more often this variation is the reflex of social, linguistic, and historical implications. In the case of variable (-ly), adverb morphology, *have got*, stative possessive meaning, the modal auxiliary system, intensifying adverbs, 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 matters 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 (Labov, 2008). The combination of techniques employed in variation analysis forms part of the 'descriptive-interpretative' strand of modern linguistic research (Sankoff, D., 1988b:142–143). Studies employing this methodology are based on the observation that individuals 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 & Sankoff, D., 1974); see also restatements in Young and Bayley (1996:254), Poplack and Tagliamonte (2001:88), which are affirmed in later textbooks (e.g. Van Herk, 2012; Meyerhoff, 2013). This is perhaps most candidly put by Sankoff, D. (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 individual 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.