6

Language Variation

Key Concepts

Dialect regions

Methodology in dialectology: assumptions and challenges

Linguistic variables and social meaning

Defining social class categories and membership

Data collection: how do we know what we have?

What correlations can tell us

This chapter builds on the discussion of varieties in chapter 2 to present a history of variationist sociolinguistic research which focuses on regional and social dialects. Sociolinguists today are generally more concerned with social variation in language than with regional variation. However, if we are to gain a sound understanding of the various procedures used in studies of social variation, we should look at least briefly at previous work in regional dialectology. That work points the way to understanding how recent investigations have proceeded as they have. Studies of social variation in language grew out of studies of regional variation. It was largely in order to widen the limits and repair the flaws that were perceived to exist in the latter that investigators turned their attention to social variation in language. As we will see, there may still be certain limitations in investigating such variation but they are of

a different kind. It is also important to note that even if there are limitations to this kind of work, many sociolinguists regard it as being essentially what sociolinguistics is – or should be – all about. In this view, the study of language variation tells us important things about languages and how they change. This chapter and the two that follow deal with such matters.

Regional Variation

The mapping of regional dialects has had a long history in linguistics (see Petyt 1980, Chambers and Trudgill 1998, and Wakelin 1977). In fact, it is a well-established part of the study of how languages change over time, that is, of **diachronic** or **historical linguistics**. Traditionally, **dialect geography**, as this area of linguistic study is known, has employed assumptions and methods drawn from historical linguistics, and many of its results have been used to confirm findings drawn from other historical sources, for example, archeological findings, population studies, and written records. In this view, languages differentiate internally as speakers distance themselves from one another over time and space; the changes result in the creation of dialects of the languages. Over sufficient time, the resulting dialects might become new languages as speakers of the resulting varieties become unintelligible to one another. So Latin became French in France, Spanish in Spain, Italian in Italy, and so on.

In this model of language change and dialect differentiation, it should always be possible to relate any variation found within a language to the two factors of time and distance alone; for example, the British and American varieties, or dialects, of English are separated by well over two centuries of political independence and by the Atlantic Ocean; Northumbrian and Cockney English are nearly 300 miles and many centuries apart. In each case, linguists working in this tradition try to explain any differences they find with models familiar to the historical linguist, models which incorporate such concepts as the 'family tree' (Latin has 'branched' into French, Spanish, and Italian), phonemic 'split' (English /f/ and /v/ are now distinctive **phonemes** whereas once they were phonetic variants, or allophones, of a single phoneme) or phonemic 'coalescence' (English *ea* and *ee* spellings, as in *beat* and *beet*, were once designated different pronunciations but they have now coalesced into the same sound), the 'comparative method' of reconstruction (English *knave* and German *Knabe* come from the same source), and 'internal reconstruction' (though *mouse* and *mice* now have different vowel sounds, this was not always the case).

Mapping dialects

Dialect geographers have traditionally attempted to reproduce their findings on maps in what they call **dialect atlases**. They try to show the geographical boundaries of the distribution of a particular linguistic feature by drawing a line on a map. Such a line is called an **isogloss:** on one side of the line people say something one way,

for example, pronounce *bath* with the first vowel of *father*, and on the other side they use some other pronunciation, for example, the vowel of *cat*. Quite often, when the boundaries for different linguistic features are mapped in this way the isoglosses show a considerable amount of criss-crossing. On occasion, though, a number coincide; that is, there is a bundle of **isoglosses**. Such a bundle is often said to mark a **dialect boundary**. One such bundle crosses the south of France from east to west approximately at the 45th parallel (Grenoble to Bordeaux) with words like *chandelle*, *chanter*, and *chaud* beginning with a *sh* sound to the north and a *k* sound to the south. Quite often, that dialect boundary coincides with some geographical or political factor, for example, a mountain ridge, a river, or the boundary of an old principality or diocese. Isoglosses can also show that a particular set of linguistic features appears to be spreading from one location, a **focal area**, into neighboring locations. In the 1930s and 1940s, Boston and Charleston were the two focal areas for the temporary spread of *r*-lessness in the eastern United States. Alternatively, a particular area, a **relic area**, may show characteristics of being unaffected by changes spreading out from one or more neighboring areas. Places like London and Boston are obviously focal areas; places like Martha's Vineyard in New England – it remained *r*-pronouncing in the 1930s and 1940s even as Boston dropped the pronunciation – and Devon in the extreme southwest of England are relic areas. Wolfram (2004) calls the dialect of such an area a **remnant dialect** and, in doing so, reminds us that not everything in such a dialect is a relic of the past for such areas also have their own innovations. Huntley, a rural enclave in Aberdeenshire, Scotland, where Marshall worked (2003, 2004), is also a relic area.

The Rhenish Fan is one of the best-known sets of isoglosses in Europe, setting off Low German to the north from High German to the south. The set comprises the modern reflexes (i.e., results) of the pre-Germanic stop consonants *p, *t, and *k. These have remained **stops** [p,t,k] in Low German but have become the **fricatives** [f,s,x] in High German (i.e., Modern Standard German), giving variant forms for 'make' [makən], [maxən]; 'that' [dat], [das]; 'village' [dorp], [dorf]; and 'I' [ik], [ix]. Across most of Germany these isoglosses run virtually together from just north of Berlin in an east–west direction until they reach the Rhine. At that point they 'fan,' as in figure 6.1. Each area within the fan has a different incidence of stops and fricatives in these words, for example, speakers in region 2 have 'ich,' 'maken,' 'Dorp,' and 'dat,' and speakers in region 4 have 'ich,' 'machen,' 'Dorf,' and [dat]. The boundaries within the fan coincide with old ecclesiastical and political boundaries. The change of stops to fricatives, called the Second German Consonant Shift, appears to have spread along the Rhine from the south of Germany to the north. Political and ecclesiastical frontiers along the Rhine were important in that spread as were centers like Cologne and Trier. The area covered by the fan itself is sometimes called a transition area (in this case, between Low and High German) through which a change is progressing, in contrast to either a focal or relic area.

Very often the isoglosses for individual phonological features do not coincide with one another to give us clearly demarcated dialect areas. As shown in figure 6.2, while the ideal is that isoglosses coincide as in (a), in reality isoglosses may
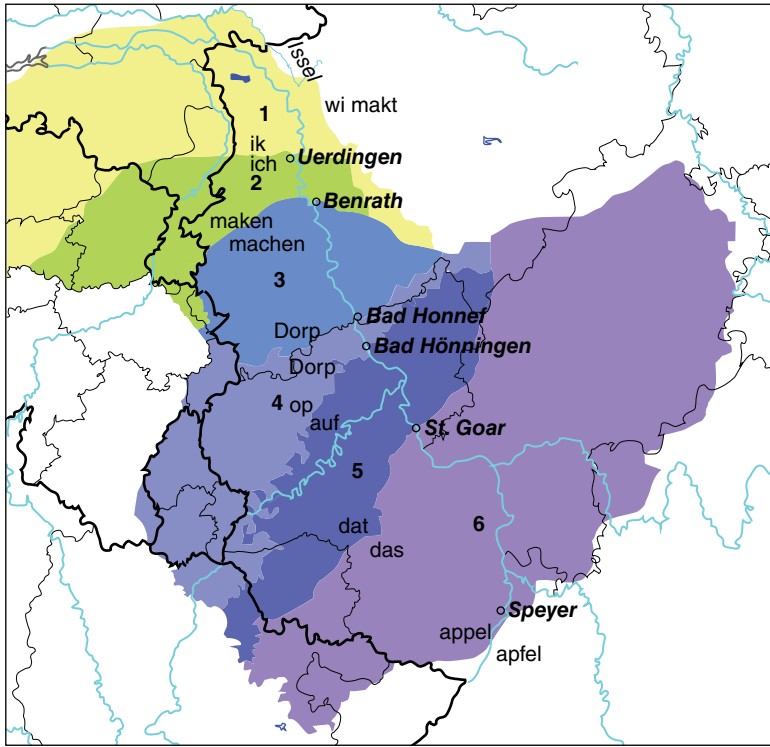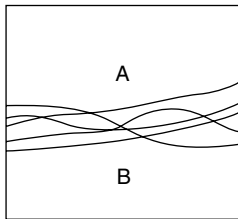
Figure 6.1 The Rhenish Fan

The main kinds of isogloss

| Term	Separates	Examples
isolex	lexical items	*nunch* vs *nuncheon*
isomorph	morphological features	*dived* vs *dove*
isophone	phonological features	*put* / pʊt / vs / pʌt /
isoseme	semantic features	*dinner* (mid-day meal) vs (evening meal)

(a)

The expectation
Isoglosses will form neat
bundles, demarcating
dialect A from dialect B.
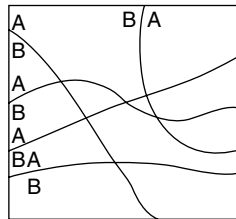
(b)

The reality Isoglosses
criss-cross an area, with
no clear boundary
between A and B.
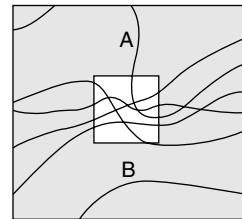
(c)

Focal and transitional On
a larger scale, the isoglosses
are seen to constitute a
transitional area between
the focal areas A and B.

Figure 6.2 Isoglosses

cross-cross as in (b); some examples of how different features of dialects might pattern can be seen in (c). Such patterns are just about impossible to explain using the traditional **family-tree account of language change**. Isoglosses do cross and bundles of them are rare. It is consequently extremely difficult to determine boundaries between dialects in this way and dialectologists acknowledge this fact. The postulated dialect areas show considerable internal variation and the actual areas proposed are often based on only a few key items (or linguistic variables in our terminology). Consequently, as Le Page (1997, 18) says, 'the dialect areas outlined by the isoglosses on the maps were artifacts of the geographer; they had to be matched against such stereotypes as "southern dialect" or "Alemmanic" or "langue d'oc," concepts which often related in the minds of outsiders to just one or two variables characterizing a complete, discrete system.'

Methods in dialectology

There are methodological issues which have caused sociolinguists to question some dialect studies. One of these issues has to do with the **sample** used for the research. First, sampling methods were based on assumptions about who 'representative' speakers of dialects were. For example, the focus was almost exclusively on rural areas, which were regarded as 'conservative' in the sense that they were seen to preserve 'older' forms of the languages under investigation. Urban areas were acknowledged to be innovative, unstable linguistically, and difficult to approach using existing survey techniques. When the occasional approach was made, it was biased toward finding the most conservative variety of urban speech. Ignoring towns and cities may be defensible in an agrarian-based society; however, it is hardly defensible in the heavily urbanizing societies of today's world.

Further, there was a circularity in how social class was addressed; in the data collection for the *Linguistic Atlas of the United States and Canada*, the analysis was partly intended to find out how speech related to social class, but speech was itself used as one of the criteria for assigning membership in a social class. For example, the **informants** chosen for the *Linguistic Atlas of the United States and Canada* were of three types (Kurath 1939, 44), chosen as follows:

Type I: Little formal education, little reading, and restricted social contacts
Type II: Better formal education (usually high school) and/or wider reading and social contacts
Type III: Superior education (usually college), cultured background, wide reading, and/or extensive social contacts

Each of these three types was then sub-categorized as follows:

Type A: Aged, and/or regarded by the field worker as old-fashioned
Type B: Middle-aged or younger, and/or regarded by the field worker as more modern

We should also note that it was the field worker for the *Atlas* who decided exactly where each informant fitted in the above scheme of things. The field worker alone judged whether a particular informant should be used in the study, and Type IA informants were particularly prized as being most representative of local speech.

In England, the Survey of English Dialects carried out between 1950 and 1961 with informants from 313 localities in England and Wales employed similar criteria (Orton et al. 1978, 3):

> The selection of informants was made with especial care. The fieldworkers were instructed to seek out elderly men and women – more often men, since women seemed in general to encourage the social upgrading of the speech of their families – who were themselves of the place and both of whose parents were preferably natives also. They were to be over 60 years of age, with good mouths, teeth and hearing and of the class of agricultural workers who would be familiar with the subject matter of the questionnaire and capable of responding perceptively and authoritatively.

Typically, both informants and field workers were male. As Coates (2004, 10–11) says, 'Dialectology … marginalized women speakers. Traditional dialectologists defined the true vernacular in terms of male informants, and organised their questionnaires around what was seen as the man's world.'

Another methodological issue involves basic ideas about language. The data collection methodology often used in earlier dialect geography studies assumes that individual speakers do not have variation in their speech; for instance, if they use the word 'pop' to talk about carbonated beverages they never use the term 'soda' to refer to the same thing, or if they merge the vowels in 'pin' and 'pen,' they always do this. This assumption has been called '**the axiom of categoricity**' (Chambers 1995: 25–33) as it treats linguistic variables as if they are categorical in the speech of an individual – and from there it is implied that they are categorical in regional dialects. This is dangerously close to the 'ideal speaker-listener' (referred to in chapter 1) that sociolinguistics eschews. As Gordon (2013, 32–3) observes, not taking variation in the speech of an individual speaker into account leads to an interpretation of the results which is misleading; presenting speakers as using variables categorically is 'taken to represent how languages work rather than how linguists work.'

Furthermore, since most of us realize that it is not only where you come from that affects your speech but also your social and cultural background, age, gender, race, occupation, and group loyalty, the traditional bias toward geographic origin alone now appears to be a serious weakness. Then, too, the overriding model of language change and differentiation is an extremely static one, and one that is reinforced, rather than questioned, by the types of data selected for analysis. Speakers from different regions certainly interact with one another; dialect breaks or boundaries are not 'clean'; and change can be said to be 'regular' only if you are prepared to categorize certain kinds of irregularities as exceptions, relics, borrowings, 'minor' variations, and so on. Furthermore, the varieties of a language spoken within large

gatherings of people in towns and cities must influence what happens to other varieties of that language: to attempt to discuss the history of English, French, or Italian while ignoring the influences of London, Paris, or Florence would seem to be something like attempting to produce *Hamlet* without the prince!

Dialect mixture and free variation

All of this is not to say that this kind of individual and social variation has gone unnoticed in linguistics. Linguists have long been aware of variation in the use of language: individuals do speak one way on one occasion and other ways on other occasions, and this kind of variation can be seen to occur within even the most localized groups. Such variation is often ascribed to **dialect mixture**, that is, the existence in one locality of two or more dialects which allow a speaker or speakers to draw now on one dialect and then on the other. An alternative explanation is **free variation**, that is, variation of no social significance. However, no one has ever devised a suitable theory to explain either dialect mixture or free variation, and the latter turns out not to be so free after all because close analyses generally reveal that complex linguistic and social factors appear to explain much of the variation.

Exploration 6.1:   Free Variation?

What vowel do you use in the first vowel in the word 'data' (/e/ or /a/), or the initial sound of the words 'economic' (/i/ or /ɛ/) or 'either' (/ai/ or /i/)? Is there any difference in social meaning between the two pronunciations?

Linguistic atlases

There have been some recent developments in linguistic atlas work which hold promise for future discoveries. They result largely from our growing ability to process and analyze large quantities of linguistic data. One, for example, is Kretzschmar's work on the *Linguistic Atlas of the Middle and South Atlantic States* (LAMSAS). He shows (1996) how it is possible to use quantitative methods to demonstrate the probability of occurrence of specific words or sounds in specific areas. Another quantitative survey (Labov et al. 2005) used a very simple sampling technique to survey the whole of North American English in order to produce the *Atlas of North American English* (ANAE), a study of all the cities on the continent with populations of over fifty thousand. This study showed that 'regional dialects are getting stronger and more diverse as language change is continuing and that the

structural divisions between them are very sharp, with very tight bundling of the isoglosses' (Labov et al. 2005, 348). (See also links to dialect atlas projects in the US and the UK in the chapter 6 materials on our website companion to this text.)

In still another approach to dialects, this one focusing on how a specific dialect emerged, Lane (2000) used a variety of economic, demographic, and social data from 3,797 residents of Thyborøn, Denmark, covering the years 1890–1996, to reveal how the local dialect 'is the result of a constant situation that led to the formation of a new dialect as a result of massive in-migration … a new system created largely out of materials selected from competing systems in contact and from innovations that indexed the new local linguistic community' (Lane 2000, 287). It was clearly another triumph for an aspiration to achieve a local identity. We can see a similar emphasis on using traditional dialect materials to help us account for current language varieties in recent writings on new Englishes (see Gordon et al. 2004, Hickey 2004, and Trudgill 2004). This discussion of dialect geography raises a number of issues which are important to our concerns. One is the kind of variation that we should try to account for in language. Another has to do with sampling the population among which we believe there is variation. Still another is the collection, analysis, and treatment of the data that we consider relevant. And, finally, there are the overriding issues of what implications there are in our findings for theoretical matters concerning the nature of language, variation in language, the language-learning and language-using abilities of human beings, and the processes involved in language change. It is to these issues that we will now turn, and in doing so, focus on social rather than regional variation in language. The major conceptual tool for investigation of such variation will be the linguistic variable.

The Linguistic Variable

The investigation of social dialects has required the development of an array of techniques quite different from those used in dialect geography. Many of these derive from the pioneering work of Labov, who, along with other sociolinguists, has attempted to describe how language varies in any community and to draw conclusions from that variation not only for linguistic theory but also sometimes for the conduct of everyday life, for example, suggestions as to how educators should view linguistic variation (see chapter 13). As we will see, investigators now pay serious attention to such matters as stating hypotheses, sampling, the statistical treatment of data, drawing conclusions, and relating these conclusions to such matters as the inherent nature of language, the processes of language acquisition and language change, and the social functions of variation.

Possibly the greatest contribution has been in the development of the use of the linguistic variable, the basic conceptual tool necessary to do this kind of work (see Wolfram 1991). As we have just indicated, variation has long been of interest to linguists, but the use of the linguistic variable has added a new dimension to linguistic investigations.

Variants

A **linguistic variable** is a linguistic item which has identifiable **variants,** which are the different forms which can be used in an environment. For example, words like *singing* and *fishing* are sometimes pronounced as *singin'* and *fishin'*. The final sound in these words may be called the linguistic variable (ng) with its two variants [ŋ] in *singing* and [n] in *singin'*. Another example of a linguistic variable can be seen in words like *farm* and *far*. These words are sometimes given *r*-less pronunciations; in this case we have the linguistic variable (r) with two variants [r] and Ø (i.e., 'zero,' or 'null'). There are at least two basically different kinds of variation. One is of the kind (ng) with its variants [ŋ] or [n], or (th) with its variants [θ], [t], or [f], as in *with* pronounced as *with*, *wit*, or *wif*. In this first case the concern is with which quite clearly distinct variant is used, with, of course, the possibility of Ø, the zero variant. that is, neither. The other kind of variation is a matter of degree, such as the quantity of nasalization of a vowel, rather than its presence or absence. How can you best quantify nasalization when the phenomenon is actually a continuous one? The same issue occurs with quantifying variation in other vowel variables: quantifying their relative frontness or backness, tenseness or laxness, and rounding or unrounding. Moreover, more than one dimension may be involved, for example, amount of nasalization *and* frontness or backness.

An important principle in the analysis of variants is the **principle of accountability**, which holds that if it is possible to define a variable as a closed set of variants, all of the variants (including non-occurrence if relevant) must be counted. So, for instance, in the study of copula usage, the use of a conjugated form of *be* (i.e., *am, is, are*), invariant *be*, and zero copula would all be included in the analysis. While in general this principle applies to grammatical variables, for pragmatically motivated variables such as discourse markers (e.g., *you know*, *well*) the principle of accountability cannot be applied, as there are no mandatory environments for such particles.

Types of linguistic variables

Linguists who have studied variation in this way have used a number of linguistic variables, many of which have been phonological. The (ng) variable has been widely used; Labov (2006, 259) says it 'has been found to have the greatest generality over the English-speaking world, and has been the subject of the most fruitful study.' The (r) variable mentioned above has also been much used. Other useful variables are the (h) variable in words like *house* and *hospital*, that is, (h): [h] or Ø; the (t) variable in *bet* and *better*, that is, (t): [t] or [ʔ]; the (th) and (dh) variables in *thin* and *they*, that is, (th): [θ] or [t] and (dh): [ð] or [d]; the (l) variable in French in *il*, that is, (l): [l] or Ø; and variables like the final (t) and (d) in words like *test* and *told*, that is, their presence or absence. Vowel variables used have included the vowel (e) in

words like *pen* and *men*; the (o) in *dog*, *caught*, and *coffee*; the (e) in *beg*; the (a) in *back*, *bag*, *bad*, and *half*; and the (u) in *pull* (see discussion in chapter 8 on the Northern Cities Vowel Shift, which addresses variation in vowel sounds).

Studies of variation employing the linguistic variable are not confined solely to phonological matters. Investigators have looked at the (s) of the third-person singular, as in *he talks*, that is, its presence or absence; the occurrence or non-occurrence of *be* (and of its various inflected forms) in sentences such as *He's happy*, *He be happy*, and *He happy*; the occurrence (actually, virtual nonoccurrence) of the negative particle *ne* in French; various aspects of the phenomenon of multiple negation in English, for example, *He don't mean no harm to nobody*; and the beginnings of English relative clauses, as in *She is the girl who(m) I praised*, *She is the girl that I praised*, and *She is the girl I praised*.

To see how individual researchers choose variables, we can look briefly at three landmark studies carried out in three urban areas by prominent sociolinguists in the 1960s and 1970s: New York City (Labov), Norwich (Trudgill), and Detroit (Shuy et al. and Wolfram).

Variation in New York City

In a major part of his work in New York City, Labov (1966) chose five phonological variables: the (th) variable, the initial consonant in words like *thin* and *three*; the (dh) variable, the initial consonant in words like *there* and *then*; the (r) variable, *r*-pronunciation in words like *farm* and *far*; the (a) variable, the pronunciation of the vowel in words like *bad* and *back*; and the (o) variable, the pronunciation of the vowel in words like *dog* and *caught*. We should note that some of these have discrete variants, for example, (r): [r] or Ø, whereas others require the investigator to quantify the variants because the variation is a continuous phenomenon, for example, the (a) variable, where there can be both raising and retraction of the vowel, that is, a pronunciation made higher and further back in the mouth, and, of course, in some environments nasalization too.

Variation in Norwich

Trudgill (1974) also chose certain phonological variables in his study of the speech of Norwich: three consonant variables and thirteen vowel variables. The consonant variables were the (h) in *happy* and *home*, the (ng) in *walking* and *running*, and the (t) in *bet* and *better*. In the first two cases only the presence or absence of *h*-pronunciation and the [ŋ] versus [n] realizations of (ng) were of concern to Trudgill. In the last there were four variants of (t) to consider: an aspirated variant; an unaspirated one; a glottalized one; and a glottal stop. These variants were ordered, with the first two combined and weighted as being least marked as nonstandard, the third as more marked, and the last, the glottal stop, as definitely marked as

nonstandard. The thirteen vowel variables were the vowels used in words such as *bad*, *name*, *path*, *tell*, *here*, *hair*, *ride*, *bird*, *top*, *know*, *boat*, *boot*, and *tune*. Most of these had more than two variants, so weighting, that is, some imposed quantification, was again required to differentiate the least preferred varieties, that is, the most nonstandard, from the most preferred variety, that is, the most standard.

Variation in Detroit

One Detroit study (Shuy et al. 1968) focused on the use of three variables: one phonological variable and two grammatical variables. The phonological variable was the realization of a vowel plus a following nasal consonant as a nasalized vowel. The grammatical variables were multiple negation, which we have already mentioned, and pronominal apposition, for example, *That guy, he don't care*. In another study of Detroit speech, Wolfram (1969) considered certain other linguistic variables. These included the pronunciation of final consonant clusters, that is, combinations of final consonants in words like *test*, *wasp*, and *left*, *th* in words like *tooth* and *nothing*, final stops in words like *good* and *shed*, and *r*-pronouncing in words like *sister* and *pair*. So far as grammatical variables were concerned, Wolfram looked at matters such as *he talk/talks*, *two year/years*, *she nice/she's nice*, *he's ready/he ready/ he be ready*, and multiple negation as in *He ain't got none neither*.

This brief sample indicates some of the range of variables that have been investigated. The important fact to remember is that a linguistic variable is an item in the structure of a language, an item that has alternate realizations, as one speaker realizes it one way and another speaker in a different way, or the same speaker realizes it differently on different occasions (see the above discussion of the axiom of categoricity) . For example, one speaker may say *singing* most of the time whereas another prefers *singin'*, but the first is likely to say *singin'* on occasion just as the second may be found to use the occasional *singing*. What might be interesting is any relationship we find between these habits and either (or both) the social class to which each speaker belongs or the circumstances which bring about one pronunciation rather than the other.

Indicators, markers, and stereotypes

Labov (1972) has also distinguished among what he calls indicators, markers, and stereotypes. An **indicator** is a linguistic variable to which little or no social import is attached. Only a linguistically trained observer is aware of indicators. For example, some speakers in North America distinguish the vowels in *cot* and *caught* and others do not; this is not salient to most non-linguists. On the other hand, a **marker** can be quite noticeable and potent carriers of social information. You do not always have to drop every *g*, that is, always say *singing* as *singin'*. Labov says that 'we observe listeners reacting in a discrete way. Up to a certain point they do not perceive the

speaker "dropping his g's" at all; beyond a certain point, they perceive him as always doing so' (Labov 1972, 226). G-dropping is a marker everywhere English is spoken. People are aware of markers, and the distribution of markers is clearly related to social groupings and to styles of speaking. A **stereotype** is a popular and, therefore, conscious characterization of the speech of a particular group: New York *boid* for *bird* or *Toitytoid Street* for *33rd Street*; a Northumbrian *Wot-cher* (What cheer?) greeting; the British use of *chap*; or a Bostonian's *Pahk the cah in Hahvahd Yahd*. Often such stereotypes are stigmatized everywhere, and in at least one reported case (see Judges 12: 4–6 in the Old Testament) a stereotypical pronunciation of *shibboleth* had fatal consequences. A stereotype need not conform to reality; rather, it offers people a rough and ready categorization with all the attendant problems of such categorizations. Studies of variation tend therefore to focus on describing the distributions of linguistic variables which are markers. (Although see Johnstone 2004 for a discussion of stereotypes in Pittsburgh speech.)

Exploration 6.2:   Stereotypes

Are there stereotypes about the variety you speak? Can you give examples of how these stereotypes might be embraced by speakers of that variety, but also stigmatized in a wider context? To what extent do you think these stereotypes are accurate portrayals of local speech?

Social Variation

Once we have identified the linguistic variable as our basic working tool, the next question is how linguistic variation relates to social variation. That is, can we correlate the use of specific linguistics features – r-lessness, for example – with membership in a particular social group?

In order to address this question, the next task becomes one of collecting data concerning the variants of a linguistic variable in such a way that we can draw certain conclusions about the social distribution of these variants. To draw such conclusions, we must be able to relate the variants in some way to quantifiable factors in society, for example, social-class membership, gender, age, ethnicity, and so on. As we will see, there are numerous difficulties in attempting this task, but considerable progress has been made in overcoming them, particularly as studies have built on those that have gone before in such a way as to strengthen the quality of the work done in this area of sociolinguistics.

Social class membership

One factor which has been prominent in sociolinguistic studies of variation is **social class** membership. If we consider 'social class' to be a useful concept to apply in stratifying society – and few indeed would deny its relevance! – we need a way to determine the social class of particular speakers. This raises various difficulties, as in many societies there are not strict guidelines, and terms such as 'middle class' may have many different meanings for the speakers themselves. Further, we must be cautious in any claims we make about social-class structures in a particular society, particularly if we attempt regional or historical comparisons. The social-class system of England in the 1950s was different from what it is today and, presumably, it will be different again in another half century, and all these class systems were and are different from those existing contemporaneously in New York, Brazil, Japan, and so on.

Sociologists use a number of different scales for classifying people when they attempt to place individuals somewhere within a social system. An occupational scale may divide people into a number of categories as follows: major professionals and executives of large businesses; lesser professionals and executives of medium-sized businesses; semi-professionals; technicians and owners of small businesses; skilled workers; semi-skilled workers; and unskilled workers. An educational scale may employ the following categories: graduate or professional education; college or university degree; attendance at college or university but no degree; high school graduation; some high school education; and less than seven years of formal education. Once again, however, some caution is necessary in making comparison across time: graduating from college or university in the 1950s indicated something quite different from what it does today. Income level and source of income are important factors in any classification system that focuses on how much money people have. Likewise, in considering where people live, investigators must concern themselves with both the type and cost of housing and its location.

In assigning individuals to social classes, investigators may use any or all of the above criteria (and others too) and assign different weights to them. Accordingly, the resulting social-class designation given to any individual may differ from study to study. We can also see how social class itself is a sociological construct; people probably do not classify themselves as members of groups defined by such criteria. Wolfram and Fasold (1974, 44) point out that 'there are other objective approaches [to establishing social groupings] not exclusively dependent on socio-economic ranking. … An investigator may look at such things as church membership, leisure-time activities, or community organizations.' They admit that such alternative approaches are not at all simple to devise but argue that a classification so obtained is probably more directly related to social class than the simple measurement of economic factors. We should note that the concept of **lifestyle** has been introduced into classifying people in sociolinguistics, so obviously patterns of consumption of goods and appearance are important for a number of people in arriving at some

kind of social classification. Coupland (2007, 29–30) calls the current era 'late-modernity.' It is a time in which 'Social life seems increasingly to come packaged as a set of lifestyle options able to be picked up and dropped, though always against a social backdrop of economic possibilities and constraints. … Social class … membership in the West is not the straitjacket that it was. Within limits, some people can make choices in their patterns of consumption and take on the social attributes of different social classes. … the meaning of class is shifted.'

In his early work on linguistic variation in New York City, Labov (1966) used the three criteria of education, occupation, and income to set up ten social classes. His class 0, his lowest class, had grade school education or less, were laborers, and found it difficult to make ends meet. His classes 1 to 5, his working class, had had some high school education, were blue-collar workers, but earned enough to own such things as cars. His classes 6 to 8, his lower middle class, were high school graduates and semi-professional and white-collar workers who could send their children to college. His highest class, 9, his upper middle class, were well educated and professional or business-oriented. In this classification system for people in the United States about 10 percent of the population are said to be lower class, about 40 percent working class, another 40 percent lower middle class, and the remaining 10 percent fall into the upper middle class or an upper class, the latter not included in Labov's study. In his later study (2001b) of variation in Philadelphia, Labov used a socio-economic index based on occupation, education, and house value.

In an early study of linguistic variation in Norwich, England, Trudgill (1974) distinguishes five social classes: middle middle class (MMC), lower middle class (LMC), upper working class (UWC), middle working class (MWC), and lower working class (LWC). Trudgill interviewed ten speakers from each of five electoral wards in Norwich plus ten school-age children from two schools. These sixty informants were then classified on six factors, each of which was scored on a six-point scale (0–5): occupation, education, income, type of housing, locality, and father's occupation. Trudgill himself decided the cut-off points among his classes. In doing so, he shows a certain circularity. His lower working class is defined as those who use certain linguistic features (e.g., *he go*) more than 80 percent of the time. Out of the total possible score of 30 on his combined scales, those scoring 6 or less fall into this category. Members of Trudgill's middle middle class always use *he goes*, and that behavior is typical of those scoring 19 or more. His study is an attempt to relate linguistic behavior to social class, but he uses linguistic behavior to assign membership in social class. What we can be sure of is that there is a difference in linguistic behavior between those at the top and bottom of Trudgill's 30-point scale, but this difference is not one that has been established completely independently because of the underlying circularity.

Shuy's Detroit study (Shuy et al. 1968) attempted to sample the speech of that city using a sample of 702 informants. Eleven field workers collected the data by means of a questionnaire over a period of ten weeks. They assigned each of their informants to a social class using three sets of criteria: amount of education, occupation, and place of residence. Each informant was ranked on a six- or seven-point scale for each set, the rankings were weighted (multiplied by 5 for education, 9 for

occupation, and 6 for residence), and each informant was given a social-class placement. Four social-class designations were used: upper middle class, those with scores of 20–48; lower middle class, those with scores of 49–77; upper working class, those with scores of 78–106; and lower working class, those with scores of 107–134.

There are some serious drawbacks to using social-class designations of this kind. Bainbridge (1994, 4023) says:

> While sociolinguists without number have documented class-related variation in speech, hardly any of them asked themselves what social class was. They treated class as a key independent variable, with variations in speech dependent upon class variations, yet they never considered the meaning of the independent variable. In consequence, they seldom attempted anything like a theory of why class should have an impact, and even more rarely examined their measures of class to see if they were methodologically defensible.

Woolard (1985, 738) expresses a similar view: 'sociolinguists have often borrowed sociological concepts in an ad hoc and unreflecting fashion, not usually considering critically the implicit theoretical frameworks that are imported.' She adds, 'However, to say that our underlying social theories are in need of examination, elaboration, or reconsideration is not to say that the work sociolinguists have done or the concepts we have employed are without merit.'

Milroy and Gordon (2008) discuss two problematic issues inherent in the study of social class. First, as a concept it combines economic aspects with status ones; this creates particular difficulty when we try to make comparison across communities, as a university professor may have a very different type of status (as well as economic standing) in one community when compared to another. Another issue has to do with mobility between social classes; again we see variation in this across societies, with mobility being greater in, for example, the United States than in the United Kingdom. In short, any categorization of speakers into social class categories must be done with careful attention to the community norms and understandings of economic and status factors. (Go to the online companion for the text for a link to a BBC study about social class in the UK which specifies seven social class categories.)

Exploration 6.3:   Social Class

How would you try to place individuals in the community in which you live into some kind of social-class system? What factors would you consider to be relevant? How would you weigh each of these? What class designations would seem to be appropriate? Where would you place yourself? You might also compare the scale you have devised for your community with similar scales constructed by others to find out how much agreement exists.

Another way of looking at speakers is to try to specify what kinds of groups they belong to and then relate the observed uses of language to membership in these groups. The obvious disadvantage of such an approach is the lack of generalizability of the results: we might be able to say a lot about the linguistic behavior of particular speakers *vis-à-vis* their membership in these groups, but we would not be able to say anything at all about anyone else's linguistic behavior. We can contrast this result with the statements we can make from using the aforementioned social-class designations: they say something about the linguistic usage of the 'middle middle class' without assuring us that there is really such an entity as that class; nor do they guarantee that we can ever find a 'typical' member.

One of the major problems in talking about social class is that social space is multi-dimensional whereas systems of social classification are almost always one-dimensional. As we have seen, at any particular moment, individuals locate themselves in social space according to the factors that are relevant to them at that moment. While they may indeed have certain feelings about being a member of the lower middle class, at any moment it might be more important to be female, or to be a member of a particular church or ethnic group, or to be an in-patient in a hospital, or to be a sister-in-law. That is, creating an identity, role-playing, networking, and so on, may be far more important than a certain social-class membership. This is the reason why some investigators find such concepts as social network and communities of practice attractive. Sometimes, too, experience tells the investigator that social class is not a factor in a particular situation and that something else is more important. For example, Rickford's work (1986) on language variation in a non-American, East Indian sugar-estate community in Cane Walk, Guyana, showed him that using a social-class-based model of the community would be inappropriate. What was needed was a conflict model, one that recognized schisms, struggles, and clashes on certain issues. It was a somewhat similar perspective that Mendoza-Denton (2008) brought to her work among rival Latina groups in a California school where the main issue was Norteña–Sureña rivalry.

One of the problems in sociolinguistics, then, is the tension between the desire to accurately portray particular speakers and to make generalizations about groups of speakers. To the extent that the groups are real, that is, that the members actually feel that they do belong to a group, a description of a social dialect has validity; to the extent that they are not, it is just an artifact. In the extremely complex societies in which most of us live, there must always be some question as to the reality of any kind of social grouping: each of us experiences society differently, multiple-group membership is normal, and both change and stability seem to be natural conditions of our existence. We must therefore exercise a certain caution about interpreting any claims made about 'lower working-class speech,' 'upper middle-class speech,' or the speech of any other social group designated with a class label – or any label for that matter.

Distinguishing among social classes in complex modern urban societies is probably becoming more and more difficult. The very usefulness of social class as a

concept that should be employed in trying to explain the distribution of particular kinds of behavior, linguistic or otherwise, may need rethinking.

Social networks

It was for reasons not unlike these that Milroy (1987) preferred to explore social network relationships and the possible connection of these to linguistic variation, rather than to use the concept of social class (see chapter 3 for an introductory discussion of social networks). In her work, Milroy found that it was the network of relationships that an individual belonged to that exerted the most powerful and interesting influences on that individual's linguistic behavior. When the group of speakers being investigated shows little variation in social class, however that is defined, a study of the network of social relationships within the group may allow you to discover how particular linguistic usages can be related to the frequency and density of certain kinds of contacts among speakers. Network relationships, however, tend to be unique in a way that social-class categories are not. That is, no two networks are alike, and network structures vary from place to place and group to group, for example, in Belfast and Boston, or among Jamaican immigrants to London and Old Etonians. But whom a person associates with regularly may be more 'real' than any feeling he or she has of belonging to this or that social class. We will have more to say in chapter 7 about this use of network structure in the study of linguistic variation.

Data Collection and Analysis

Once an investigator has made some decision concerning which social variables must be taken into account and has formed a hypothesis about a possible relationship between social and linguistic variation, the next task becomes one of collecting data that will either confirm or refute that hypothesis. In sociolinguistics, this task has two basic dimensions: devising some kind of plan for collecting relevant data, and then collecting such data from a representative sample of speakers. As we will see, neither task is an easy one.

The observer's paradox

An immediate problem is one that we have previously referred to as the **observer's paradox**. How can you obtain objective data from the real world without injecting your own self into the data and thereby confounding the results before you even begin? How can you be sure that the data you have collected are uncontaminated by the process of investigation itself? This is a basic scientific quandary, particularly observable in the social sciences where, in almost every possible situation, there is

one variable that cannot be controlled in every possible way, namely, the observer/ recorder/analyst/investigator/theorist him- or herself. If language varies according to the social context, the presence of an observer will have some effect on that variation. How can we minimize this effect? Even data recorded by remote means, for example, by hidden cameras and sound recorders, may not be entirely 'clean' and will require us to address additional ethical issues which severely limit what we can do and which we would be extremely unwise to disregard. We know, too, that observations vary from observer to observer and that we must confront the issue of the reliability of any observations that we make. Sociolinguists are aware that there are several serious issues here, and, as we will see, they have attempted to deal with them.

The sociolinguistic interview

Unlike the methodology used in dialect geography studies, which often involved explicitly asking speakers to provide linguistic information, the methodology in sociolinguistics is geared toward having the research participants (the term preferred over 'informants' or 'subjects' in sociolinguistics today) provide speech in context. This approach addresses the issues of both non-categorical use and stylistic variation. That is, the interviewer manipulates the context to try to have interviewees focus more or less on how they are speaking. The traditional sociolinguistic interview involves a casual interview, which ideally resembles a conversation more than a formal question and answer session. In addition to trying to make the interviewee feel comfortable enough to talk in a casual speech style, Labov also introduced the 'danger of death' question, in which interviewees were asked to talk about situations in which they had felt themselves to be in serious danger. The idea behind this is that the interviewees would become emotionally involved in the narrative and forget about how they are talking in their involvement with what they are saying.

To get more formal styles of speech, investigators also ask research participants to do various reading tasks: a story passage, lists of words, and **minimal pairs**. Each of these tasks requires an increased level of attention to speech. The texts are designed to contain words which illustrate important distinctions in the regional or social dialect being studied; for instance, if it is known that some speakers in the regional or social group of this speaker pronounce 'cot' and 'caught' with the same vowel, these words, or other words with these vowels, will be present in the reading materials, and be presented as a minimal pair in the final task. Speakers are obviously most likely to pronounce these words differently if they are reading them as a pair. This methodology assumes that if speakers are going to adjust their speaking style, they will use what they consider to be increasingly formal and correct speech in these elicitations.

While many researchers have followed this approach to sociolinguistic fieldwork, sociolinguists continue to rethink and develop data collection methods. For example,

the idea that the conversation in a sociolinguistic interview can be described as 'natural' has been challenged, and many linguists recognize 'that there is no one single "genuine" vernacular for any one speaker, since speakers always shape their speech in some way to fit the situation or suit their purposes' (Schilling 2013, 104). Mendoza-Denton (2008, 222–5) also questions the naturalness of such interview-derived data and the usefulness of the danger of death question. She says that in her work using the latter would have been an 'outright *faux pas* … highly suspicious to gang members … very personal, and only to be told to trusted friends.' However, she does admit that 'the sociolinguistic interview paradigm … has yielded replicable results that allow us to contextualize variation in a broader context.' Labov's own recent work (2001a) still distinguishes between casual and careful speech but provides for a more nuanced assessment of how the research participant views the speech situation.

Sampling

Another critical aspect of sociolinguistic research is sampling: finding a representative group of speakers. The conclusions we draw about the behavior of any group are only as good as the sample on which we base our conclusions. If we choose the sample badly, we cannot generalize beyond the actual group that comprised the sample. If we intend to make claims about the characteristics of a population, we must either assess every member of that population for those characteristics or sample the whole population in some way. Sampling a population so as to generalize concerning its characteristics requires considerable skill. A genuine sample drawn from the population must be thoroughly representative and completely unbiased. All parts of the population must be adequately represented, and no part should be overrepresented or underrepresented, thereby creating bias of some kind. The best sample of all is a **random sample**. In a random sample everyone in the population to be sampled has an equal chance of being selected. In contrast, in a **judgment sample** (also known as a **quota sample**) the investigator chooses the subjects according to a set of criteria, for example, age, gender, social class, education, and so on. The goal is to have a certain quota of research participants in each category; for example, if the study aims to look at age and social class, the goal is to include X number of people in each age group from each social class. Sometimes, too, it is the investigator who judges each of these categories, for example, to which social class a subject belongs. A judgment sample, although it does not allow for the same kind of generalization of findings as a random sample, is clearly more practical for a sociolinguist and it is the kind of sample preferred in most sociolinguistic studies (see Chambers 2003, 44–5 and Milroy and Gordon 2008, 30 ff).

In sampling the speech of the Lower East Side in New York City, Labov did not use a completely random sample because such a sample would have produced subjects who were not native to the area, for example, immigrants from abroad and

elsewhere in the United States. He used the sampling data from a previous survey that had been made by Mobilization for Youth, a random sample which used a thousand informants. Labov's own sample size was eighty-nine. He used a **stratified sample**, that is, one chosen for specific characteristics, from that survey. He also wanted to be sure that he had representatives of certain groups which he believed to exist on the Lower East Side. When he could not, for various reasons, interview some of the subjects chosen in the sample, he tried to find out by telephoning the missing subjects if his actual sample had been made unrepresentative by their absence. He was able to contact about half of his missing subjects in this way and, on the basis of these brief telephone conversations, he decided that his actual sample was unbiased and was typical of the total population he was interested in surveying.

The Detroit study (Shuy et al. 1968) initially collected data from 702 informants in the city. However, the data used for the actual analysis came from only thirty-six informants chosen from this much larger number. In selecting these thirty-six, the investigators wanted to be sure that each informant used had been a resident of Detroit for at least ten years, was 'representative,' had given a successful interview, and had provided an adequate amount of taped material for analysis. In other words, to any initial biases that might have been created in choosing the first set of 702 informants was added the possibility of still further bias by choosing non-randomly from the data that had become available. This is not to suggest that any such biases vitiate the results: they do not appear to do so. Rather, it is to point out that the kinds of concerns sociolinguists have about data and sources of data have not necessarily been the same as those of statisticians.

Wolfram (1969) chose forty-eight Black informants from those interviewed in the Detroit study. These informants were evenly divided into four social classes used in that study. Each group of twelve was further divided into three age groups: four informants in the 10–12 age group, four in the 14–17 age group, and four in the 30–55 age group. Wolfram also selected twelve White informants from the highest social class in the Detroit project, again by age and sex. Wolfram's study therefore used a total of sixty informants: twenty-four (twelve White and twelve Black) from the upper middle class and thirty-six who were Black and were members of the working classes. Such a sample is very obviously highly stratified in nature.

It is actually possible to use a very small sample from a very large area and get good results. For their *Atlas of North American English* (ANAE) Labov and his co-workers sampled all North American cities with populations over 50,000. Labov (2006, 396) reports that they did this through a telephone survey: 'Names were selected from telephone directories, selecting by preference clusters of family names representing the majority ethnic groups in the area. The first two persons who answered the telephone and said that they had grown up in the city from the age of four or earlier, were accepted as representing that city (four or six persons for the largest cities). A total of 762 subjects were interviewed.' The investigators were very pleased with the results of this sampling procedure for the ANAE.

Apparent time and real time

Investigations may also have a 'time' dimension to them because one purpose of sociolinguistic studies is trying to understand language change. They may be **apparent-time** studies in which the subjects are grouped by age, for example, people in their 20s, 40s, 60s, and so on. Any differences found in their behavior may then be associated with changes that are occurring in the language. **Real-time** studies elicit the same kind of data after an interval of say ten, twenty, or thirty years. If the same informants are involved, this would be in a **panel study**; if different people are used it would be in a **trend study**. Obviously, real-time studies are difficult to do. The study of the Queen's English is one such study (Harrington et al. 2000, mentioned in chapter 2), but she was the sole panel member. The study that replicated Labov's work on Martha's Vineyard (Pope et al. 2007) was a real-time trend study. As we will see in the following pages, most studies of change in progress are apparent-time studies for reasons which should now be obvious.

Exploration 6.4:   Research Design

What are the advantages/disadvantages of: random versus quota sampling; real versus apparent time studies; sociolinguistic interviews versus recordings of naturally occurring data? Think about what kinds of data are collected using these different approaches, and also about what is practical in terms of carrying out research. How are the choices researchers make linked to their research questions?

Correlations: dependent and independent variables

Studies employing the linguistic variable are essentially correlational in nature: that is, they attempt to show how the variants of a linguistic variable are related to social variation in much the same way that we can show how children's ages, heights, and weights are related to one another. However, a word of caution is necessary: correlation is not the same as causation. It is quite possible for two characteristics in a population to covary without one being the cause of the other. If A and B appear to be related, it may be because either A causes B or B causes A. However, it is also possible that some third factor C causes both A and B. The relationship could even be a chance one.

To avoid the problems just mentioned, we must distinguish between **dependent variables** and **independent variables**. The linguistic variable is a dependent variable, the one we are interested in. We want to see what happens to language when

we look at it in relation to some factor we can manipulate, the independent variable, for example, social class, age, gender, ethnicity, and so on: as one of these changes, what happens to language? As Chambers (2003, 26) expresses it, 'Socially significant linguistic variation requires correlation: the dependent (linguistic) variable must change when some independent variable changes. It also requires that the change be orderly: the dependent variable must stratify the subjects in ways that are socially or stylistically coherent.'

Quantitative sociolinguistics

This kind of sociolinguistic investigation is often called **quantitative sociolinguistics** (or **variationist** sociolinguistics) and it is, as we have indicated previously, for some sociolinguists the 'heart of sociolinguistics' (Chambers 2003, xix). Quantitative studies must therefore be statistically sound if they are to be useful. Investigators must be prepared to employ proper statistical procedures not only in their sampling but also in the treatment of the data they collect and in testing the various hypotheses they formulate. They must be sure that what they are doing is both valid and reliable. **Validity** implies that, as said by Lepper (2000, 173): 'the researcher must show that what is being described is accurately "named" – that is, that the research process has accurately represented a phenomenon which is recognizable to the scientific community being addressed.' **Reliability** is how objective and consistent the measurements of the actual linguistic data are. Data collection methodology is part of this issue; if only one person collected the data, how consistent was that person in the actual collection? If two or more were involved, how consistently and uniformly did they employ whatever criteria they were using? Bailey and Tillery (2004, 27–8) have identified a cluster of such issues, for example, the effects of different interviewers, elicitation strategies, sampling procedures, and analytical strategies, and pointed out that these can produce significant effects on the data that are collected and, consequently, on any results that are reported. Therefore, there may still be room for improving the reliability of our results.

Serious empirical studies also require experimental hypotheses to be stated *before* the data are collected, and suitable tests to be chosen to decide whether these hypotheses are confirmed or not and with what degree of confidence. (For more discussion of statistical analyses in sociolinguistics, see Bayley 2013 and Tagliamonte 2006.)

Petyt (1980, 188–90) points out how the kinds of figures that sociolinguists use in their tables may be misleading in a very serious way. Sociolinguists stratify society into sub-groups, the members of which are measured in certain ways, and then these measurements are pooled. Individual variation is eliminated. Hudson (1996, 181) offers a similar criticism, declaring that such pooling

> loses too much information which may be important. Information about the use of individual variants is lost when they are merged into variable scores, and information

about the speech of individuals is also lost if these are included in group averages. At each stage the method imposes a structure on the data which may be more rigid than was inherent in the data, and to that extent distorts the results – discrete boundaries are imposed on non-discrete phonetic parameters, artificial orderings are used for variants which are related in more than one way, and speakers are assigned to discrete groups when they are actually related to each other in more complex ways.

Petyt (1980, 189) provides the data given in figure 6.3. These data come from an investigation of *h*-dropping in West Yorkshire, and the figure shows the means for five sub-groups, that is, social classes. As can be seen, these groups appear to vary quite a bit. However, Petyt points out that, if the range of variation within each sub-group is also acknowledged to be of consequence, there is a considerable overlap among the performances of individuals, so that 'it is not the case that this continuum can be divided in such a way that the members of each social class fall within a certain range, and members of other classes fall outside this.' He indicates the range of individual scores in figure 6.4, and adds that for Classes II and V, there was one individual in each group which provided the lowest and highest figure, respectively. These outliers could be eliminated and the groups would then be more uniform, but their presence shows that the groups are not discrete groups which are unified in their linguistic behavior.

It is quite obvious that if we look only at means in such a case we are tempted to say one thing, whereas if we consider the distribution of responses within each class we may draw some other conclusion. The overriding issue is that there are approved procedures to help investigators to decide how far they can be confident that any differences that they observe to exist among the various classes, that is, among the various means, are due to something other than errors in measurement or peculiarities of distribution. Such procedures require an investigator not only to calculate the means for each class, but also to assess the amount of variation in the responses
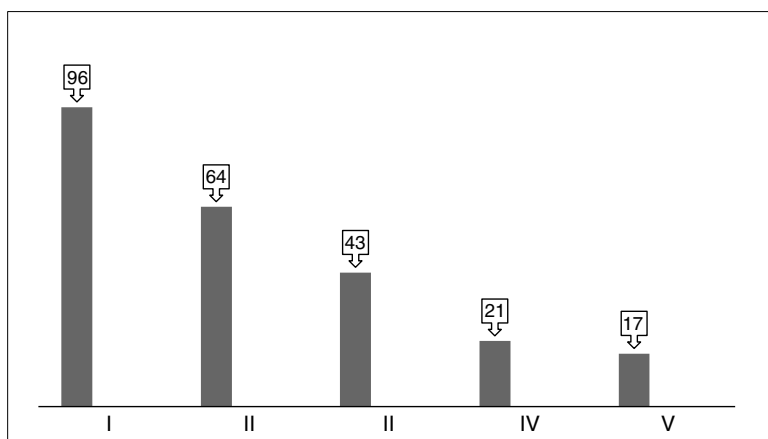

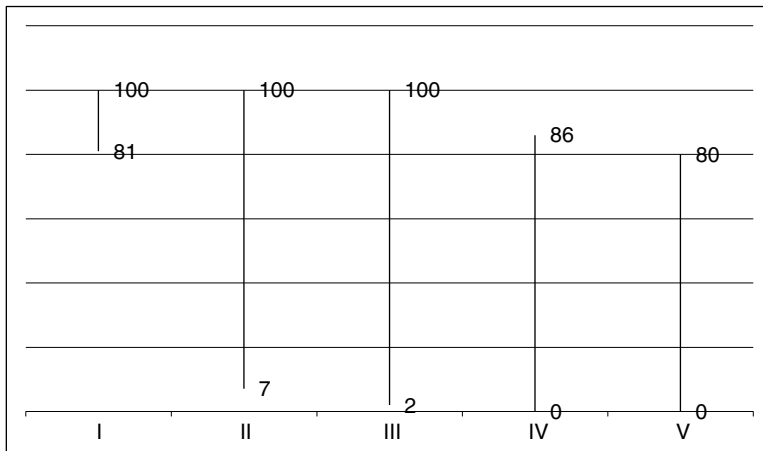**Figure 6.3** H-dropping means for five social groups

Figure 6.4   H-dropping: within-group ranges for five social groups

within each class, and then to test pairs of differences of means among the classes using a procedure which will indicate what likelihood there is that any difference found occurs by chance, for example, one chance in twenty.

Most social scientists employing statistical procedures regard this last level of significance as a suitable test of a hypothesis. In other words, unless their statistical procedures indicate that the same results would occur by chance in less than one case in twenty, they will *not* say that two groups differ in some respect or on a particular characteristic; that is, they insist that their claims be significant at what they call the 0.05 level of significance. We are also much more likely to find two means to be significantly different if they are obtained from averaging a large number of observations than from a small number.

Whenever you look at results reported by sociolinguists, you should keep in mind the above-mentioned issues concerning the formulation of hypotheses and the collection, analysis, and interpretation of data. In examining individual sociolinguistic investigations, therefore, you must ask what exactly are the hypotheses; how reliable are the methods used for collecting the data; what is the actual significance of results that are reported on a simple graph or histogram; and what do the findings tell us about the initial hypotheses.

Milroy and Gordon (2008, 168) provide another perspective on the use of statistics in the study of language, asking: 'should we equate failure to achieve statistical significance with sociolinguistic irrelevance?' Their answer is that 'statistical tests, like all quantitative procedures are tools to provide insight into patterning in variation. They must be used critically.' Labov himself (1969, 731) has stated that statistical tests are not always necessary: 'We are not dealing here with effects which are so erratic or marginal that statistical tests are required to determine whether or not they might have been produced by chance.' Dealing with a critic of Labov's work, Milroy (1992, 78) says:

It is not surprising that an anti-quantitative linguist should advocate confirmatory statistical testing, but it is very important to understand the proposition put forward here is simply wrong. If Labov's interpretations were suspect (and of course they are not), this would not arise from the fact that he failed to test for significance. There was no reason for him to do so because the claims he wished to make were quite simple … and because in his analysis the same patterns were repeated for every variable studied.

According to Milroy, since this kind of sociolinguistic inquiry is 'exploratory' in nature, it can be likewise 'exploratory' in its quantitative approach. Labov's recent work (2001b) is still exploratory in nature but it is also extremely sophisticated in its sampling, data collection, and hypothesis-testing.

Chapter Summary

In this chapter we present an overview of the development of research on regional dialects, including methodologies used to create dialect maps and study the patterns in local vernaculars. We also introduce the concept of the linguistic variable, which is central to linguistics and variationist sociolinguistics in particular. We review some early important work on regional varieties of English by well-known socio-linguists who were responsible for the growth of the field. Further, we look at social dialects and how they are studied, focusing in particular on social class, and what methodologies have traditionally been used to study this variation.

Exercises

1. As we have said, the (ng) variable, realized as [n] or [ŋ], is generally a noticeable phonological variable throughout the English-speaking world. This task requires you to do some 'field work.' Devise a way of collecting instances of the use of (ng) in naturally occurring discourse. You may want to listen to song lyrics, recorded interviews, talk shows, news reports, and so on. The key is to access both unmonitored speech, that is, talk that is focused on 'content' rather than on 'form,' and more conscious varieties, in which speakers are clearly trying to speak Standard English. After you have collected some data and analyzed what you have, try to figure out how you might improve your results if you were to repeat the task. (You could then repeat it to see what progress you made.) You can be sure that none of the research findings reported in this chapter and in the following two came from first attempts at data collection, but were preceded by such pilot studies!

2. In the following text, identify all of the contexts for the linguistic variable of the copula (that is, the verb *to be*). What are the variants which appear here? (Hint: be sure to include the zero copula variant). Can you describe the contexts in which they occur? (You may wish to consult the description of AAVE from chapter 2, as some usages are from that social dialect.)

Today movie prices are entirely too high. It doesn't make no sense to pay that much, because the picture the people be showing is not worth it. If you going to pay that much for a movie you should at least have a cut on the prices of the food. Not only the food is high, but you cannot sit in a nice clean place. But still you paying that very high price to get inside the place. Another reason you got against paying such a high price is that the people at the movies be throwing popcorn all in your head. You not paying that much money to come to a movie and get food stains all on your clothes and hair.

Further Reading

Chambers, Jack K. and Natalie Schilling (eds.) (2002). *The Handbook of Language Variation and Change*. Oxford: Blackwell.
>   This collection of papers on language variation and change presents articles by leading sociolinguistics which address issues of theory and method in sociolinguistic research.

Gordon, Matthew J. (2013). *Labov: A Guide for the Perplexed*. New York: Bloomsbury.
>   Written for a general audience, this is an especially good and accessible discussion of Labov's influential work and how it has shaped sociolinguistics.

Milroy, Lesley and Matthew Gordon (2008). *Sociolinguistics: Method and Interpretation*. 2nd edn. Oxford: Blackwell.
>   A comprehensive introduction to research in the field of sociolinguistics, with especial foci on phonological variation and style-shifting and code-switching.

Schilling, Natalie (2013). *Sociolinguistic Fieldwork*. Cambridge: Cambridge University Press.
>   This is a thorough treatment on research methodology in sociolinguistics, with special attention given to research on speech style, addressing practical, ethical, and theoretical issues.

Tagliamonte, Sali (2012). *Variationist Sociolinguistics: Change, Observation and Interpretation*. Oxford: Wiley-Blackwell.
>   A research guide with a primary focus on quantitative research methods with both a treatment of social factors and detailed sections on the analysis of phonological and grammatical variation.

For further resources for this chapter visit the companion website at

🖥 **www.wiley.com/go/wardhaugh/sociolinguistics**

References

Bailey, G. and J. Tillery (2004). Some Sources of Divergent Data in Sociolinguistics. In C. Fought (ed.) (2004), *Sociolinguistic Variation: Critical Reflections*. New York: Oxford University Press.

Bainbridge, W. S. (1994). Sociology of Language. In R. E. Asher and J. M. Simpson (eds.), *The Encyclopedia of Language and Linguistics*. Oxford: Pergamon.

Bayley, R. (2013) The Quantitative Paradigm. In Jack K. Chambers and Natalie Schilling (eds.), *The Handbook of Language Variation and Change*. 2nd edn. Oxford: Wiley-Blackwell, 85–107.

Chambers, J. K. (1995). *Sociolinguistic Theory: Linguistic Variation and its Social Significance*. Oxford: Blackwell.

Chambers, J. K. (2003). *Sociolinguistic Theory: Linguistic Variation and its Social Significance*. 2nd edn. Oxford: Blackwell.

Chambers, J. K. and P. Trudgill (1998). *Dialectology*. 2nd edn. Cambridge: Cambridge University Press.

Coates, J. (2004). *Women, Men and Language*. 3rd edn. London: Longman.

Coupland, N. (2007). *Style: Language Variation and Identity*. Cambridge: Cambridge University Press.

Gordon, E., L. Campbell, J. Hay et al. (2004). *New Zealand English: Its Origin and Evolution*. Cambridge: Cambridge University Press.

Gordon, M. J. (2013) *Labov: A Guide for the Perplexed*. New York: Bloomsbury.

Harrington, J., S. Palethorpe, and C. I. Watson (2000). Does the Queen Speak the Queen's English? *Nature* 408: 927–8.

Hickey, R. (2004). *Legacies of Colonial English*. Cambridge: Cambridge University Press.

Hudson, R. A. (1996). *Sociolinguistics*. 2nd edn. Cambridge: Cambridge University Press.

Johnstone, B. (2004). Place, Globalization, and Linguistic Variation. In C. Fought (ed.), *Sociolinguistic Variation: Critical Reflections*. New York: Oxford University Press.

Kretzschmar, W. A., Jr (1996). Quantitative Aerial Analysis of Dialect Features. *Language Variation and Change* 8: 13–39.

Kurath, H. (1939). *Handbook of the Linguistic Geography of New England*. Providence, RI: Brown University Press.

Labov, W. (1966). *The Social Stratification of English in New York City*. Washington, DC: Center for Applied Linguistics.

Labov, W. (1969). Contraction, Deletion, and Inherent Variability of the English Copula. *Language* 45: 715–62.

Labov, W. (1972). *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.

Labov, W. (2001a) The Anatomy of Style Shifting. In P. Eckert and J. R. Rickrods (eds.), *Style and Sociolinguistic Variation*. Cambridge: Cambridge University Press, 85–108.

Labov, W. (2001b). *Principles of Linguistic Change, II: Social Factors*. Oxford: Blackwell.

Labov, W. (2006). *The Social Stratification of English in New York City*. 2nd edn. Cambridge: Cambridge University Press.

Labov, W. (2007). Transmission and Diffusion. *Language* 83(2): 344–87.

Labov, W., C. Boberg, and S. Ash (2005). *The Atlas of North American English*. Berlin: Mouton de Gruyter.

Lane, L. A. (2000). Trajectories of Linguistic Variation: Emergence of a Dialect. *Language Variation and Change* 12: 267–94.

Le Page, R. B. (1997). The Evolution of a Sociolinguistic Theory of Language. In F. Coulmas (ed.), *The Handbook of Sociolinguistics*. Oxford: Blackwell.

Lepper, G. (2000). *Categories in Text and Talk: A Practical Introduction to Categorization Analysis*. London: Sage.

Marshall, J. (2003). The Changing Sociolinguistic Status of the Glottal Stop in Northeast Scottish English. *English World-Wide* 24(1): 89–108.

Marshall, J. (2004). *Language Change and Sociolinguistics: Rethinking Social Networks*. Basingstoke: Palgrave Macmillan.

Mendoza-Denton, N. (2008). *Homegirls*. Oxford: Blackwell.

Milroy, J. (1992). *Language Variation and Change*. Oxford: Blackwell.

Milroy, Lesley (1987). *Language and Social Networks*. 2nd edn. Oxford: Blackwell.

Milroy, Lesley and Matthew Gordon (2008). *Sociolinguistics: Method and Interpretation*, vol. 13. 2nd edn. Chichester: John Wiley & Sons, Ltd.

Orton, H., S. Sanderson, and J. Widdowson (eds.) (1978). *The Linguistic Atlas of England*. London: Croom Helm.

Petyt, K. M. (1980). *The Study of Dialect: An Introduction to Dialectology*. London: André Deutsch.

Pope, Jennifer, Miriam Meyerhoff, and D. Robert Ladd (2007). Forty Years of Language Change on Martha's Vineyard. *Language* 83(3): 615–27.

Rickford, J. R. (1986). The Need for New Approaches to Social Class Analysis in Sociolinguistics. *Language & Communication* 6(3): 215–21.

Schilling, Natalie (2013). *Sociolinguistic Fieldwork*. Cambridge: Cambridge University Press.

Shuy, R. W., W. Wolfram, and W. K. Riley (1968). *Field Techniques in an Urban Language Study*. Washington, DC: Center for Applied Linguistics.

Tagliamonte, Sali (2006). *Analysing Sociolinguistic Variation*. New York: Cambridge University Press.

Trudgill, P. (1974). *The Social Differentiation of English in Norwich*. Cambridge: Cambridge University Press.

Trudgill, P. (2004). *New-Dialect Formation: The Inevitability of Colonial Englishes*. Oxford: Oxford University Press.

Wakelin, M. F. (1977). *English Dialects: An Introduction*. Rev. edn. London: Athlone Press.

Wolfram, W. (1969). *A Sociolinguistic Description of Detroit Negro Speech*. Washington, DC: Center for Applied Linguistics.

Wolfram, W. (1991). The Linguistic Variable: Fact and Fantasy. *American Speech* 66(1): 22–32.

Wolfram, W. (2004). The Sociolinguistic Construction of Remnant Dialects. In C. Fought (ed.), *Sociolinguistic Variation: Critical Reflections*. New York: Oxford University Press.

Wolfram, W. and R. W. Fasold (1974). *The Study of Social Dialects in American English*. Englewood Cliffs, NJ: Prentice-Hall.

Woolard, K. A. (1985). Language Variation and Cultural Hegemony: Toward an Integration of Linguistic and Sociological Theory. *American Ethnologist* 12: 738–48.