

# 1

## *Correlations*

It is precisely because language is as strictly socialized a type of behavior as anything else in culture and yet betrays in its outlines such regularities as only the natural scientist is in the habit of formulating, that linguistics is of strategic importance for the methodology of social science. Behind the apparent lawlessness of social phenomena there is a regularity of configuration and tendency which is just as real as the regularity of physical processes in a mechanical world . . . Language is primarily a cultural or social product and must be understood as such . . . It is peculiarly important that linguists, who are often accused, and accused justly, of failure to look beyond the pretty patterns of their subject matter, should become aware of what their science may mean for the interpretation of human conduct in general.

*Edward Sapir (1929: 76–7)*

This book is about language variation and its social significance. By now, the research literature on this topic, from the first breakthroughs almost 40 years ago to the most recent refinements, amounts to a formidable accumulation. It includes, by any reasonable yardstick, some of the most incisive discoveries in the long history of humanity's inquiries into the structure and function of language. My purpose is to make a critical synthesis of as much of that research, great and small, as I can handle within the covers of one book.

Looked at that way, my topic perhaps looks grand. But there is a sense in which it is narrow. The social significance of language variation is only one aspect of the discipline of sociolinguistics, broadly conceived. I will be dealing only with what might be called urban dialectology, that is, with accent or dialect as an emblem of an individual's class, sex, age, ethnicity, ambition, or some other social attribute. When we consider the enormous

number of uses that language serves in our daily interactions with other people, its social significance does really not cover much of the territory. In §1.1 below, I sketch the various social uses of language in order to put into a larger perspective the area to be covered in detail in this book.

The rest of this chapter is also devoted to providing perspectives on the subject matter of the chapters that follow. In §1.2, I explore the main theoretical construct of sociolinguistics, the linguistic variable, and look at its historical development, methodological premises, and theoretical basis. In §1.3 I compare and contrast categorical theories, especially Chomskyan linguistics, with sociolinguistics, a variationist theory, emphasizing the essential difference between them.

## 1.1 The Domain of Sociolinguistics

Sociolinguistics, as the study of the social uses of language, encompasses a multitude of possible inquiries. Ordinarily, we simply take for granted the numerous ways we use language in our social interactions because they are so deeply embedded in our daily affairs. It is sometimes hard for people to understand that a brief telephone conversation could possibly be of interest as an object of serious linguistic study. It is also hard for them to understand how much we reveal about ourselves – our backgrounds, our predilections, our characters – in the simplest verbal exchange.

What we need is a degree of objectivity – the willingness to step back and take a fresh look at our mundane activities in order to see them as the fascinating and exotic and often very complex events that they really are. We must reflect upon, for instance, the multitude of inferences individuals make when they are engaged in a conversation. The best kind of conversational exchange for reflecting upon is one in which the information is almost exclusively linguistic, as when you overhear a conversation between strangers sitting behind you in a bus or when you receive a telephone call from a total stranger. On those occasions, you begin the exchange with the minimum of knowledge and presupposition. And yet, after hearing only a few sentences, you find yourself in possession of a great deal of information of various kinds about people whom you have never seen.

The kinds of inferences you tacitly make fit into five general categories. In the following sections I call them personal, stylistic, social, sociocultural, and sociological.

### 1.1.1 Personal characteristics

One level of information is *personal*. Is the voice high-pitched or low? Nasal or open? Does the pitch move up and down the scale or is it relatively monotonal? Does the speaker lisp?

Like all the other linguistic observations we make, even those at much more sophisticated levels, these take place spontaneously, with very little consciousness on our part. And they are very often accompanied by spontaneous judgements, partly culture-driven and partly experience-driven. One obvious one is that monotonal speech is monotonous. Indeed, those two words – monotonal and monotonous – are etymologically almost identical as adjectives derived (by different Latin suffixes *-al* and *-ous*) from a complex noun meaning “one tone.”

Also at the personal level are inferences about the *speaking ability* of the individuals you are listening to. Is their speech fluent or hesitant? Is it articulate or vague? These are among the simplest, most superficial observations we make but, even at this level, the observations interact to give strong (though not necessarily accurate) impressions of character. A speaker who is fluent but vague will seem to us to be evasive, perhaps deceitful, and one who is articulate but hesitant will seem pensive and thoughtful. And there are of course many other possible judgements at this level. Is the person’s vocabulary current and slang-inflected or ornate and careful?

Observations like these at the personal linguistic level have attracted relatively little serious linguistic study. Traditionally, they were considered too idiosyncratic or individualistic for framing hypotheses about language in general. With the insurgence of studies of the social use of language, including sociolinguistics (as discussed in §1.3 below), research into personal characteristics has increased.

Still, it is probably true that most personal linguistic characteristics offer little of interest to sociolinguists. If some aspect of a person’s voice quality comes to be thought of as pathological, as are some kinds of lisp or stuttering, that person might be referred to a speech therapist, and speech therapists naturally classify the kinds of conditions referred to them in order to develop treatments for them, but their studies are outside the domain of sociolinguistics. By the same token, if some aspect of the person’s speaking ability is deemed an impediment for cosmetic or occupational reasons, that person might seek the help of an elocutionist in hopes of learning how to speak more “attractively” (whatever that might mean) or more convention-

ally. The elocutionists' manual of speaking aids is irrelevant to sociolinguistics, except perhaps in the way that a manual of etiquette might be of interest in sociology, as an indicator of the social values attached to particular mannerisms at a particular time.

Observations about personal speech characteristics could perhaps be better integrated into sociolinguistic research than they are. Sapir (1927) made an attempt at considering speech as a "personality trait" but his fascinating study has not inspired productive research by others. One avenue that would surely be interesting and possibly productive would be studying how (if at all) personal speech characteristics differ from society to society or, conversely, how they remain constant across social and cultural boundaries. It would also be of considerable sociolinguistic interest to discover how consistently these varied personal characteristics are used by listeners to form judgements about the speakers. For the time being, however, considerations like these are at the fringe of sociolinguistic research.

### 1.1.2 *Linguistic styles*

Another level of observation is *stylistic*. Here again listeners are capable of considerable discrimination, spontaneously and almost instantaneously, concerning the degree of familiarity between the participants in a conversation, their relative ages and ranks, the function of their conversation, and many other aspects. The main determinant is the speech styles they are using. The range of possibilities encompasses, on the one hand, the casualness of utterly familiar, long-time friends who share a wealth of common experience and, on the other hand, the formality of unequal participants who have no common ground but are forced to interact for some reason or other – perhaps one is hiring the other to mow the lawn, or instructing the other to serve the tea – with numerous possibilities in between.

Unlike the personal traits discussed above, speech styles fall squarely into the domain of sociolinguistics. Stylistic differences have a simple social correlate: *formality tends to increase in direct proportion to the number of social differences between the participants*. The most relevant social factors are the topic of the next heading (§1.1.3), but for now it is enough to know that age is one of them, and to think of the effect that age differences often impose upon a discussion. Imagine a conversation between two women from the same neighborhood who unexpectedly meet in the waiting room of a dentist's office; imagine first that both women were, say, 30, and then imagine the difference in the conversation if one was 30 and the other 70.

The sociolinguistic relevance comes about because our ability to judge the formality of a conversation is largely determined by linguistic cues. Casual conversations tend to be more rapid, with more syntactic ellipses and contractions, and more phonological assimilations and coalescences. Highly formal conversations can also be very rapid if a participant is very nervous, but in that instance the syntax is usually stilted and somewhat breathless and the phonology articulated unnaturally. In English, one stereotype of hyper-formality is the pronunciation of the indefinite article “a,” which is ordinarily pronounced [ə], as hyper-correct [eɪ]. There is also a middle ground between casual style and formal style, typically found in linguistic interactions between peers, that is, people who share many social characteristics, called careful style.

Clearly, if the relative formality of a conversation can cause speakers to adjust their phonology and other aspects of dialect and accent, then style is an independent variable that affects the dependent speech variables. The importance of style was recognized in a study that proved to be the most important precursor of modern sociolinguistics, when Fischer (1958: 49) noted that the choice of the suffix [ɪn] for [ɪŋ] in participles like *walkin'*, *talkin'* and *thinkin'* in the speech of Boston schoolchildren “changed from an almost exclusive use of *-ing* in the [formal] situation to a predominance of *-in* in the informal interviews.” (Fischer’s study is discussed further in §2.9.4.1 and §3.2.1.)

Style was firmly established as an independent variable in sociolinguistics, as were so many other factors, when Labov made it an integral part of his interview protocols in his ground-breaking survey of New York City (1966a: 90–135 and passim; see §1.2.2.5 below). Labov asked his subjects to talk about topics such as street games and life-threatening experiences. He also asked them to read passages of connected prose and lists of words into the tape recorder. These tasks elicit a range of styles from the speakers. The essential difference between speech styles is the amount of self-monitoring people do when they are speaking. When people are asked to read lists of words, they obviously concentrate on their pronunciation almost completely, especially when the reading is being recorded by someone who is admittedly studying the way they speak. The care and attention is even greater than usual if the words are arranged as minimal pairs – “cot” and “caught,” or “poor” and “pour” or (from Labov’s list) “God” and “guard” (§1.2.2.6 below).

The reading of connected prose is also highly monitored – so much so that most people are well aware of sounding different when they read – but the requirement of maintaining coherence when reading a passage aloud

deflects some attention away from speech and on to the content of the passage.

In a free discussion, the content becomes even more important. Though self-monitoring is normal as an interviewee frames answers to the interviewer's questions, it must obviously be less than when reading a passage because the content of the answer must be foremost.

The unmonitored style – casual speech – is the one that sociolinguists want most to study, and it is the one that cannot be elicited by any foolproof devices. After the interviews have been going on for several minutes, the subjects normally become accustomed to the recording apparatus and more relaxed with the interviewer. When they are asked to tell the interviewer about near-fatal car accidents or fires in the toaster or other events that involved them, they are likely to get caught up in the recollected urgency of the situation and forget their self-consciousness. As interviewers, we can work at developing good rapport in the course of the interview, and at finding some topic that will touch a nerve. Apart from that, the best prospect of eliciting casual speech comes about when some intimate third person interrupts the interview, by telephone or in person, while the recording is taking place. (The elicitation of a range of styles is further exemplified in the summary of Labov's New York interview protocol in the next section.)

Elicitation of a range of styles is routinely included in sociolinguistic interviews. In the discussion of results throughout this book, style is often included as an independent variable. I refer to the styles in the conventional way by using self-explanatory terms (and their abbreviations): word list style (WL) is elicited by the reading of a list of words; the more self-conscious variant elicited by arranging the words based on their phonological similarities is called, simply, minimal pairs (MP); reading passage style (RP) is elicited by recording a prepared text; interview style (IS) is the free discussion of topics with perhaps some direction by the interviewer; and casual style (CS) is the unmonitored natural vernacular.

Throughout the book, style is an important independent variable but it is never the focal point. (For fuller discussions of sociolinguistic style, see Bell 1984, Schilling-Estes 2002.) The focal point in this book will be social variables of the type to which we now turn.

### *1.1.3 Social characteristics*

Whenever we speak we reveal not only some personality traits and a certain sensitivity to the contextual style, but also a whole configuration of

characteristics that we by and large share with everyone who resembles us socially. Usually without any conscious effort on our part, we embody in our speech, as in our dress, manners, and material possessions, the hallmarks of our social background. Our speech, from this perspective, is emblematic in the same sense as is the car we drive or the way we dress for work but, obviously, our speech is much less manipulable, much harder to control consciously, and for that reason much more revealing.

The *social class* to which we belong imposes certain norms of behavior on us and reinforces them by the strength of the example of the people with whom we associate most closely. The sub-elements of social class include education, occupation, income and type of housing, all of which play a role in determining the people with whom we will have daily contacts and more permanent relationships. They tend to be similar to those of our parents, so that the class trappings that most adults surround themselves with are to some degree an updated replication of those they grew up with. In all of this, of course, there is some latitude and, in relatively free societies, some mobility. The effects of social class on speech are the subject of chapter 2.

The other major social factors that exert a tacit and largely irrepressible effect on our behavior, including the way we speak, are *sex* and *age*. Their effects on our speech are the subjects of chapter 3 and chapter 4, respectively.

In modern industrial societies, these three social characteristics – class, sex, and age – are the primary determinants of social roles. They are, of course, enormously complex, subsuming a host of social factors. The chapters on the primary characteristics break them down into their molecular elements in so far as those elements have a demonstrable effect on the way people speak.

For social class, the essential distinction separates non-manual and manual workers (§2.1). The effect of occupational mobility blurs the class lines not only socially but also linguistically (§2.4). In close-knit social clusters of the kind often (but not exclusively) found in manual workers' communities, the degree to which individuals are integrated into their local networks may affect their uses of regional markers (§2.6–§2.8). Even with class distinctions and network pressures impinging upon the individual, linguistic behavior is by no means rigidly defined but can vary within certain limits (§2.9).

For sex, the essential distinction separates sex roles, which are biological, and gender roles, which are sociological (§3.1). In various communities, men and women divide the social labor in different ways, with

demonstrable consequences in their use of the local dialect (§3.3, §3.5). In many other communities, including some of the best-studied ones, the roles of men and women are not fundamentally different with respect to language use, and in those instances there are still found typical differences in female–male use of sociolinguistic markers. These differences appear to find their explanation in inherent differences in language skills between the sexes (§3.4).

For age, the relative ease with which listeners can estimate speakers' ages within a few years appears to be a function of non-idiosyncratic voice quality changes that are a function of normal aging (§4.1). Of much greater social significance is the acquisition of sociolectal features beginning in infancy, as a normal development along with basic linguistic competence (§4.2). Individuals in complex modern societies pass through at least three stages that have significant linguistic consequences, from the formation of a local peer group in childhood (§4.3), through the dense networks of adolescence (§4.4), to the settling into a style of life in early adulthood (§4.5). Linguistic differences between groups that differ from one another only in age can signal either a regular, maturational change (§4.6.1) or, more likely, a linguistic change in progress in the community (§4.6.2).

The correlations between these independent variables and speech is complex, sometimes dauntingly so, and endlessly interesting. These correlations are the ones with which this book is centrally concerned.

Our speech also – inevitably – reifies some less pressing but nevertheless important realities, which, for want of a ready set of terms for the broad categories, I will call sociocultural and sociological factors.

#### *1.1.4 Sociocultural factors*

To some extent, the very topics we talk about are culturally determined, and to a greater extent the way in which we talk about them is too.

This is obviously true of phatic communion, the repertoire of speech events we use when we greet people, pass the time in brief encounters at bus stops or waiting rooms, and take our leave of them. Consider, as an example, the conversation-ender, “Have a nice day.” It has a peculiarly American ring to it because it came into use in the United States and started spreading outward from there in the 1970s. Outside of the United States, its use is occasionally unself-conscious, a sign that the users have either lost track of its foreignness altogether or have adopted it as a stylish Americanism. But often its use outside of the United States is imbued



with a sense of parody, and its users are well aware not only of it as foreign but also as a stereotype of jaunty, hail-fellow American tourists.

Culture-laden interactions abound, and usually their verbal component is less important than the non-verbal. For instance, for Canadians and Americans in Budapest, it takes some time to adjust to the fact that shopkeepers will sell them their wares in silence and then end the transaction by placing their change on the counter, not in their hand. It no doubt takes Eastern Europeans at least as long to come to grips with the fact that shopkeepers in Toronto and Chicago make small talk and maintain eye contact with their customers.

Similarly, almost all people who are not native New Yorkers are likely to feel they are being insulted and shouted at by taxi and bus drivers or waiters when they visit that city. Partly this is the result of a cultural difference, what Tannen identified as the New York conversational style, which favors “cooperative overlap – i.e. speakers who like a lot of talk going on in casual conversation, much of which occurs at the same time that others are speaking” (Tannen 1987: 581). To outsiders, it seems as if they are being constantly interrupted and shouted down. The rules governing speech acts, though not encoded anywhere (except by ethnographers), are understood thoroughly by natives and often misunderstood by outsiders.

Less obvious sociocultural influences are found in the use of conversational implicatures. In some middle-class settings, it is permissible to request an action by asking a question that is, taken literally, highly indirect. “Do you find it cold in here?” literally invites assent or denial, but it might in a polite gathering in most cultures lead to a discussion of the weight of one’s cardigan or the unseasonal June weather. In certain English-speaking social circles a perfectly polite response to the question would be for the host to rise and stoke the fire or close the window.

Numerous discourse rules differ subtly from culture to culture, such as the conventions for maintaining the conversational topic, ways of assuming a turn as speaker, the intimacy of disclosures, and the amount of overlapping or interrupting. (For an overview of discourse analysis, see Macaulay 2002.)

### *1.1.5 Sociological factors*

Language also functions sociologically in various ways as an accoutrement of the social structure. This is especially clear in the conventionalized use

of address forms. Linguistically it is irrelevant whether someone addresses someone else as “Mr Jones” or as “Sam,” or whether someone chooses the pronoun *to* or *shomā* “you” in a Persian-speaking community (or Spanish *tu* or *Usted*, Italian *tu* or *Lei*, Ukrainian *ty* or *vy*, Greek *esi* or *esis*). Sociologically, it can make an enormous difference, if the distinction obligatorily marks the social ranks of the participants.

Also largely sociological is the importance of particular languages as “codes” in multilingual societies. Coexistent languages are never sociologically equal, though of course they are linguistically equal. In Haiti, it is important to know when to use French and when to use creole in order to gain and maintain status in the diglossic society. In the bilingual belt of Canada, it is important to know when to use French and when to use English and, more subtly, when to mix the codes (Poplack 1988). In all bilingual regions, there are times when it is appropriate to code-switch and times when it is incongruous. Native speakers understand the unwritten rules for diglossia and for code-switching because they are integral elements of the value system of their societies.

Just as the most idiosyncratic personal factors in our speech are outside the domain of sociolinguistics, so at the sociological end of the continuum the topics tail off into linguistically extrinsic matters. Purely ideological issues that impinge upon language planning in multinational administrations belong properly to political science, and debates about linguistically equivalent but sociologically distinct spelling reforms may touch educators, politicians, and sociologists. In these areas, the linguist as linguist will have little or nothing to offer, whatever the linguist as teacher, politician, or citizen may think.

### *1.1.6 Sociolinguistics and the sociology of language*

The preceding outline of the various ways in which social structure and linguistic structure come together is intended to cover the field in the broadest possible terms. Studies of the various aspects – personal, stylistic, social, sociocultural, and sociological – are subsumed by either sociolinguistics or the sociology of language, depending upon the purposes of the research.

The domain of sociolinguistics deals centrally with stylistic and social aspects and also takes in some personal and some sociocultural aspects. The subject matter of this book will concentrate, as I have said, on the core area: linguistic variation and its social significance. Each of the three chap-

ters following this one concentrates on one of the three major social correlates of linguistic variation – class, sex, and age – along with the web of related social factors that each of them entails. The final chapter steps back and considers the purposes of linguistic variation in general, and the reasons for its perpetuation.

As necessary background for the synthesis of the major results of sociolinguistic research that follow, the rest of this chapter reviews the essentials of data analysis and data presentation that underlie those results. Methods of data gathering will not be reviewed here. For that, see Labov (1975, 1984b), Chambers and Trudgill (1998: chapters 4–6), L. Milroy (1987), and Chambers et al. (2002: Part 1).

Sociolinguists are sometimes accused of giving more space to their methods than to their results. That will certainly not be true of this book. Still, it seems prudent to allot some space to analytical matters because they contextualize the results as nothing else can.

The methodological matters that are dealt with in the following sections arise incidentally and (I hope) unavoidably as adjuncts of the main theme. In the rest of this chapter, I provide a theoretical perspective on linguistic variation. I will show that the admission of the variable as a structural unit in linguistic analysis – the topic of §1.2, immediately below – represents a breakthrough of considerable magnitude in linguistic theory – the topic of §1.3.

These theoretical considerations may be less well understood even by experienced sociolinguists than they perhaps should be. They are, at any rate, almost never discussed by them. One reason for the relative paucity of discussion may simply be that the social-science content of sociolinguistics has overshadowed its theoretical implications. Where research in more formal branches of linguistics has little content apart from its theoretical implications, sociolinguists have had to face empirical matters of enormous complexity as well. Those empirical matters dominate this book as they do the discipline, but for the rest of this chapter they are relegated to the background.

## 1.2 The Variable as a Structural Unit

In 1885, Schuchardt (1972: 48) noted that “the pronunciation of the individual is never free from variations.” A few decades later, Sapir (1921: 147) wrote, “Everyone knows that language is variable.” Though linguistic

variation may be obvious, no linguists analyzed it systematically until the inception of sociolinguistics in the 1960s.

Even Sapir, in practice, was unprepared to incorporate insights about variability into his linguistic analysis. While he was more willing than his contemporaries – or, for that matter, most of his predecessors and successors – to recognize variation, he did it mainly by acknowledging and discussing the loose ends of a particular analysis. It was his awareness of these loose ends that led to his epigrammatic conclusion: “Unfortunately or luckily, no language is tyrannically consistent. All grammars leak” (1921: 38).

Sapir’s persistent allusions to the fact of variation in language, though never more than an addendum in his work, were not considered commendable by some of his immediate successors. Thus Joos (1957: 25) summarized Sapir’s contribution by saying, “We welcome the insights of his genius, which allowed no scrap of evidence to escape at least subconscious weighing; where it is possible to check up, we normally find him right; thus we may seem captious when we point out that he also said many things which are essentially uncheckable (‘invulnerable’) and not science.” Sapir’s contemporaries and his immediate successors (Joos among them) viewed science as behaviorist and anti-mentalist, and that view perhaps forestalled any significant influence from “the insights of [Sapir’s] genius.” Though Sapir was respected and emulated as a practicing linguist, he inspired no movement devoted to accounting for variability within linguistic theory.

Instead, the generation that came after Sapir adopted a very strong form of what I will call the axiom of categoricity, the simplifying assumption that data for linguistic analysis must be regularized to eliminate real-world variability. (The axiom of categoricity is discussed in greater detail in §1.3.2 below.)

Among many statements of the axiom, one by Joos stands out as the most pellucid. “We must make our ‘linguistics’ a kind of mathematics within which inconsistency is by definition impossible,” Joos said, and he offered this elaboration:

Ordinary mathematical techniques fall mostly into two classes, the continuous (e.g. the infinitesimal calculus) and the discrete or discontinuous (e.g. finite group theory). Now it will turn out that the mathematics called “linguistics” belongs to the second class. It does not even make any compromise with continuity as statistics does, or infinite-group theory. Linguistics is a quantum mechanics in the most extreme sense. All continuities, all possibilities of infinitesimal gradation, are shoved outside of linguistics in one direction or the other. (Joos 1950: 701–2)

As we shall see in the next section, Chomsky influentially perpetuated the axiom of categoricity for the next generation.

### 1.2.1 *Coexistent systems and free variation*

Of course the fact that variation exists in language remains as obvious as Schuchardt and Sapir said it is. People use the lexical item *car* on one occasion and *automobile* on another, pronounce the participle *walking* sometimes and *walkin'* other times, and ask the technician to *back up the disk* on their computer one day and to *back the disk up* the next. These mundane examples – lexical, phonological, and syntactic as illustrated, but pronunciation and morphological examples also abound – represent dozens of others like them. They are perhaps easy to ignore because linguistically they make no difference: the sentences mean the same thing no matter which variant is used.

When variants attracted the attention of linguists at all, they were generally regarded either as belonging to different coexistent linguistic systems or as unpredictably free substitutes. The notion of coexistent systems, explicated most influentially by Fries and Pike (1949), held that speakers maintained separate phonologies (and also, by inference, separate grammars) that gave them access to more than one code, thus allowing them or perhaps causing them to switch from one to the other. The notion of free variation, widely subscribed to but, as far as I can discover, never discussed or examined critically, held that the variants were merely random fluctuations.

The idea of coexistent systems carries certain implications that must have made it seem dubious from the beginning. It implies, for instance, that speakers should maintain one phonology consistently until the circumstances arise that trigger the second system. Mixing elements from the two systems, that is, alternating between the two, should not, in principle, take place as long as the conditions surrounding the speech event remain unchanged. Access to the second system, because it is in some sense separate from the first, should not be momentary or sporadic. Variants, as constituents of separate systems, should not co-occur.

Yet all these things are known to happen in casual conversations. Speakers do mix variants of the same variable in the same discourse, indeed in the same sentence. Subsequent chapters will describe numerous variables in which large groups of people – class cohorts, or age or sex cohorts – maintain similar proportions of variant mixing in similar circumstances.

Rather than the coexistence of linguistic systems, their behavior appears to provide evidence for a single system in which variants coexist (also see Weinreich et al. 1968: 159–65).

The idea of free variation carries a strong implication as well. If the variants are truly free, that is, if the occurrence of one variant or another is arbitrary, then it must follow *a fortiori* that the variants cannot be predicted by any factor. Yet the most casual observations of speech show that its variants are associated with social factors. Discussions of free variation routinely included observations that “free” variants like *automobile* were used in advertisements because the word had a dignity lacking in its counterpart *car*. In other words, the variants were predictable, at least probabilistically, and not free in any meaningful sense of the word.

The impetus for one of the first systematic studies of language variation came about when a sociologist, John L. Fischer, became curious about “certain inconsistencies” in the speech of the New England schoolchildren he was studying in a child-rearing project. He especially noticed the children’s alternations between *-ing* and *-in’* in participles, sometimes saying *running* and other times *runnin’*. When he asked some linguists how they accounted for the alternations, he was told that it was merely “free variation.” Fortunately, Fischer was not satisfied. He later wrote:

“Free variation” is of course a label, not an explanation. It does not tell us where the variants came from nor why the speakers use them in different proportions, but is rather a way of excluding such questions from the scope of immediate inquiry. Historically, I presume that one could investigate the spread of one of these variants into the territory of another through contact and migration, and this would constitute one useful sort of explanation. However, another sort of explanation is possible in terms of current factors which lead a given child in given circumstances to produce one of the variants rather than another. (Fischer 1958: 47–8)

Fischer’s simple statement expresses the most fundamental motivation for sociolinguistics, and the basis for sociolinguistic analysis. He set out to discover “the determinants of the selection of the variants,” and he found them in correlations with social class, sex, and other independent variables (see §2.9.4.1, §3.2.1.2). His research provided a convincing demonstration of the vacuity of “free” variation, and with it an early model for sociolinguistics.

### 1.2.2 *The sociolinguistic enterprise*

Soon after Fischer's study, questions about the frequency and the source of linguistic variants became important to a few linguists. The next crucial step was to gather evidence of variability on a large scale and demonstrate that its occurrence in the speech community was systematic, patterned, and orderly.

The very act of gathering and analyzing variable data would in itself constitute proof that such data was manageable. One of the premises in support of the axiom of categoricity, as Joos suggests above, was that abstracting linguistic data from the vagaries of the real world was necessary in order to make it coherent and manageable. Demonstrating the falseness of this premise would in itself show that categoricity is not an axiom one must accept in order to undertake linguistic analysis, but merely a postulate that one may or may not wish to make.

#### 1.2.2.1 *Precursors*

Several linguists were aware that categoricity was a postulate rather than an axiom, long before the analysis of variability was successfully demonstrated. (See, for example, the citations from Bolinger, Gleason, and Chomsky in the next section.) Nevertheless, the fact that the axiom had been adopted in virtually all language research prior to sociolinguistics, all the way back to classical Greece (§5.5) and including traditional dialectology and anthropological linguistics, must have given it an aura of necessity.

The term *sociolinguistics*, or at least the near-homograph *socio-linguistics*, was coined before any research of the kind that came to be called sociolinguistic was attempted. Haver C. Currie (1952), a poet and philosopher, noted that linguists' definitions of language conventionally included a clause about its social function but actual research usually ignored its social functions. Currie lauded Mencken's popular book, *The American Language* (1919), noting that Mencken "pleases Americans in general by pointing out that they have a national language of their own with respectable regional variants," and that "the American middle class has for thirty years rewarded him by buying every copy of his book that his publishers would print" (1952: 46). In Mencken's popularity Currie saw the impetus for a chauvinist extension of linguistics, which he called "a field for quite conscious study here called *socio-linguistics*," for "the consideration of the social significance of English as spoken in the United States" (1952: 47).

Currie's proposal was not pursued immediately but the term he coined later came into general use, even against opposition. Twenty years after Currie, Labov (1972a: xiii) said, "I have resisted the term *sociolinguistics* for many years, since it implies that there can be a successful linguistic theory or practice which is not social." Indeed there can be if Chomskyan linguistics is successful, but it was not that consideration so much as the universal attachment of the term to Labov's work by others that broke his resistance. A few years later, Chambers and Trudgill (1980: 205) complained that the term *sociolinguistics* had become "perhaps too general to be meaningful" and proposed some alternatives, to no avail (and the discussion disappeared in later editions of their book). By now, there is no hope of detaching the term from the discipline, and, it must be admitted, no good reason.

By the time sociolinguistics came into being a little more than a decade after Currie, virtually no one remembered that Currie had coined the term. It was Labov's exemplary work in the Lower East Side of New York City that inspired other linguists to head into the streets with notepads in their hands and tape recorders over their shoulders.

Labov's New York survey was not the first sociolinguistic study. That honor may also belong to Labov, for his work in Martha's Vineyard in 1962 (Labov 1963; also §2.7.1 below), though a slightly looser definition would give it to Fischer (1958), whose research prefigured most of the essential components of sociolinguistics, as we saw in the preceding section. An even looser definition would credit Louis Gauchat (1905), whose dialect study of the French vernacular in the town of Charmey, Switzerland, broke the conventions of traditional dialectology by correlating linguistic variability with the sex and age of the individual informants, prefiguring at least sociolinguistic network studies (§2.6 below, including discussion of Gauchat's work in §2.6.3). But Labov's New York survey was so enormously influential that it is indisputably the fountainhead.

#### 1.2.2.2 *Labov's New York survey*

In this section, I present an actual case study in order to demonstrate how the analysis of variable data proceeds and what its correlations with independent variables reveal. There are by now literally dozens of analyses in the literature that I could choose as prototypes. The one I will use is one of the very first analyses published by Labov.

The choice of this particular analysis, besides serving an expository purpose at this point, is also meaningful for other reasons. The article in which it appeared was called "The linguistic variable as a structural unit"



(Labov 1966b), the source of my section title. This particular article was obscurely published and seems to have been all but forgotten. Most of Labov's articles have been anthologized – Labov 1972a and 1972b are self-anthologies – but this one never has been. For me, it was his single most influential article. When I came upon it about 10 years after it was published it deflected my attention decisively away from theoretical syntax and the other kinds of research I was involved in and on to urban dialectology.<sup>1</sup>

In it, Labov explains that he undertook a study of the social stratification of English in New York City well aware that his predecessors had come to the conclusion that New York was the site of massive free variation. One predecessor, Hubbell, said,

The pronunciation of a very large number of New Yorkers exhibits a pattern . . . that might most accurately be described as the complete absence of any pattern. Such speakers sometimes pronounce /r/ before a consonant or a pause and sometimes omit it, in a thoroughly haphazard pattern. . . . The speaker hears both types of pronunciation about him all the time, both seem almost equally natural to him, and it is a matter of pure chance which one comes first to his lips. (Hubbell 1950: 48)

When Hubbell undertook his New York study, dialect research in urban areas had been almost completely ignored because of the widespread apprehension that accents in densely populated places with mobile populations would be capricious, vacillating, and disparate. Hubbell's conclusions about "the complete absence of any pattern," "thoroughly haphazard," "a matter of pure chance," accorded perfectly with the expectations of the day. With hindsight, of course, it seems a bizarre conclusion. The essential premise of all linguistic research is that language is systematic. If it is not, there is nothing more to be said. Linguistics cannot exist. Yet Hubbell's conclusion sounded no alarms.

### 1.2.2.3 *Linguistic variables*

Against this background, Labov set out to discover not only when individual speakers pronounced /r/ but also what factors might govern its occurrence. Besides /r/, he isolated four other features known to vary in the city. In order to symbolize the variables graphically, he used parentheses: (r) means "variable r." The parentheses are equivalent to slashes for phoneme /r/ and square brackets for phone [ɹ] but they indicate a level of analysis that is neither phonemic nor phonetic.

In order for something to be a linguistic variable, it must occur in

variant forms. Variable (r) has two variants: [ɹ] and Ø. Another of Labov's variables, (eh), the vowel in words such as *bad*, *ask*, *dance*, and *laugh*, has six variants: [ɪ<sup>e</sup>], [ɛ<sup>e</sup>], [æ], [æɪ], [ɑ:] and [ɑ:].<sup>2</sup> This variable plots the raising and tensing of standard /æ/ from the low front position. The essential pattern of variation is somewhat obscured by the inclusion of the two back variants, which are rare and, in the case of [ɑ:], aberrant in New York City speech.

Like the phonemes /r/ and /æ/, the variables (r) and (eh) represent abstract linguistic entities. Just as phonemes are actualized as one or more allophones, so variables are actualized as two or more variants. Unlike the allophones of phonemes, the variants of variables such as (r) and (eh) are not always predictable by phonological, morphological, or any other kind of linguistic conditioning. Allophones *must* occur when their linguistic conditioning factors are present; for instance, in English, voiceless stops must occur as aspirates in syllable-initial position. Variants may have a tendency to occur when certain linguistic factors are present in the environment; for instance, the higher variants of (eh), [ɪ<sup>e</sup>] and [ɛ<sup>e</sup>], may occur more frequently before alveolar nasals, as in *Ann*, *candy*, and *dandruff*, than before bilabial nasals, as in *Sam*, *ramble*, and *ambulance*, but they do not occur invariably before alveolar nasals and they sometimes do occur before bilabial nasals.

#### 1.2.2.4 Independent variables

The occurrence of one or another of the variants may thus correlate with some linguistic factor in the environment, but as a probability rather than a necessity. Non-linguistic factors may also be found to correlate. One of these is style, as already noted in §1.1.2 above. A particular variant may tend to occur more often in casual interchanges than in formal ones, or vice versa. Various social factors may also exert an influence. Certain variants may be associated with working-class speech more than middle-class, with adolescents more than their parents, with men more than women, or some other social distinction.

Correlating linguistic variation as the dependent variable with independent variables such as linguistic environment, style, or social categories is the primary empirical task of sociolinguistics. From recordings of the speech of a sample of individuals, transcriptions of the variants by which variables get realized can be counted and quantified. If the variable has two variants, as (r) does, the quantification can be made simply in terms of percentages. If it has more variants, as (eh) does, the variants must be counted proportionally.

By a convention established by Trudgill (1974), the basic variant in the set of variants is the one that occurs in standard speech: thus, [ɹ] rather than Ø for (r), and [æ] rather than [ɪ<sup>ɛ</sup>] or [ɛ<sup>ɛ</sup>] for (eh). This means, in the case of a percentage calculation as for (r), that the individual's score will be indicated by the percentage of Ø variants among the total of all variants (all [ɹ] + Ø). The number, whether a percentage or a proportion, is called the weighted index score, or, quite commonly, the index. (For details about the calculation of weighted index scores, see Chambers and Trudgill 1998: 49–53; for other methods of quantifying linguistic variables, see, for example, Cedergren and Sankoff 1974, D. Sankoff 1985, Horvath and Sankoff 1987, Bayley 2002.)

Beyond this primary empirical task lies the problem of interpretation. What is the relative significance of, say, age and style if both are involved in the variation? How do favoring linguistic factors and disfavoring social factors interact? What is the social significance of the linguistic variation?

#### 1.2.2.5 *Speech in the community*

Starting with a stratified random sample of New York's Lower East Side, Labov selected his subjects from a sub-set of the adult native speakers who had not moved for two years. He sought representatives of all ethnic groups, age levels, and social classes. Various problems in getting access to individuals affected the sample's randomness in the end but left him with a large, well-distributed judgement sample, the kind of sample that has proved most judicious in sociolinguistic research. (On sampling, see §2.1 below.)

Labov and his associate interviewer Michael Kac then set about interviewing 157 adults and, as opportunities arose, added interviews with 58 of their children. The interview protocol, as has become standard procedure, involved discussions of various topics broken up by set tasks such as the reading of a minimal pair list (MP style), an isolated word list (WL), and a reading passage (RP). The purpose, as discussed above (§1.1.2), was to elicit a variety of styles from the subjects, based on the amount of self-monitoring they were doing. The bulk of the recorded interviews consisted of relatively careful but free-flowing speech as the subjects responded to various topics introduced by the interviewer. This register is known as interview style (IS).

Of course the style most desired for sociolinguistic purposes is completely unmonitored, and that is the very style that is hardest to elicit in the presence of an interviewer and recording equipment. This fundamen-

tal problem of sociolinguistic research was dubbed by Labov the “observer’s paradox” (1970: 32) and stated thus: “the aim of linguistic research in the community must be to find out how people talk when they are not being systematically observed; yet we can only obtain this data by systematic observation.” The observer’s paradox is the sociolinguistic counterpart of the general problem in social sciences known as the Hawthorne effect (as pointed out by Murray 1985: 327): “the behavior of any given experiment’s subjects is changed *just because* the subjects perceive themselves as participants in the experiment.”

Of the topics used by Labov (1966a: 143–9), the most successful in making the subjects forget the unnaturalness of the situation were the recollection of street games and of life-threatening situations. Most reliable in eliciting truly casual speech were fortuitous interruptions by family members and friends while the tape recorder was turned on.

#### 1.2.2.6 *One subject, Susan Salto*

When the interviews were transcribed and analyzed, Labov correlated the linguistic variables with the social and stylistic variables. One individual, Susan Salto, a 37-year-old third-generation New Yorker, showed striking variation in her use of  $\emptyset$  for (r) in the two polar styles. When she was reading the list of minimal pairs such as *god* and *guard*, which in New York speech are realized as [gɔ:d] “god” and either [gɔ:d], [gɔ:r’d], or [gɔ:r’d] “guard,” she invariably distinguished all pairs that could possibly differ by r-fulness of one of the variants. For the pair just cited, she read [gɔ:d] for “god” and [gɔ:r’d] for “guard.” In other words, in the most self-conscious style, she scored zero because she had no instances of the variant  $\emptyset$  in the set of MP tokens. By contrast, in her most casual speech, as in her breathless account of a life-threatening incident (for instance) and fortuitously when she chatted with a friend on the telephone part-way through the interview, she used the  $\emptyset$  variant 98 percent of the time. Her style-shifting thus covers almost the whole gamut – zero to 98 percent.

In the absence of further information, nothing about Susan Salto’s responses with respect to the (r) variable is incompatible with the hypothesis of free variation. The differences in her use of [ɹ] or  $\emptyset$  could be merely random. Obviously the plausibility of randomness seems somewhat suspicious because the percentages are so extreme, and because they are not evenly distributed throughout her speech but each is associated with a different linguistic context.

When Labov took a further step and examined her responses in the

*Table 1.1* Susan Salto's (r) index in five styles, from most formal to least formal (from Labov 1966b)

<i>Style</i>	<i>(r) index</i>
MP	00
WL	39
RP	42
IS	74
CS	98

other contextual styles, he got the results shown in table 1.1. Now the plausibility of randomness as an explanation of her behavior becomes totally unrealistic. What we notice in table 1.1 is a systematic gradation in the use of variants from style to style. In the styles between the two polar ones, her use of the (r) variants alters in a regular way. For the word list it is 39, a considerable leap from zero in the minimal pairs, and for the reading passage (RP) it is similar, 42; in the conversational styles, there is another considerable leap to 74 in interview style (IS) and then nearly categorical use in the casual style (CS). These results are consistent with a hypothesis that Susan Salto uses more non-standard variants as her speech style becomes more casual. The corollary is that she uses more standard variants as she monitors her speech more attentively.

In terms of phonological structure, we might attempt to analyze Susan Salto's responses by saying that she has a phoneme /r/ with allophones [ɾ] and Ø, but there is no way of capturing the probabilistic occurrence of one or the other allophone. Or we might posit a generative process by which underlying /r/ gets realized as Ø sometimes. Since the variants are not linguistically conditioned, there is no way to make these generalizations explicit unless we incorporate non-linguistic conditions into the structural statements and probabilities into the application. But these are impossible in a phonological theory that includes the axiom of categoricity. The linguistic description of Susan Salto's behavior belongs to a different level.

#### *1.2.2.7 All subjects in three social classes*

From the evidence presented so far, it remains a possibility that Susan Salto's behavior is idiosyncratic. If so, the stylistic correlation is merely personal – an interesting fact about Susan Salto but not generally reveal-

Table 1.2 (r) indices for three social classes in five styles in New York City (from Labov 1966b)

Style	Class		
	LC	WC	MC
MP	50.5	45	30
WL	76.5	65	44.5
RP	85.5	79	71
IS	89.5	87.5	75
CS	97.5	96	87.5

ing. Labov demonstrated that this is not so by agglomerating the data from the entire sample. Table 1.2, which categorizes the subjects according to their social classes, shows that Susan Salto is not the only one whose more formal speech includes more instances of the standard variant [ɹ]. Indeed, according to the table, her behavior conforms to everyone else in her speech community. For all three social classes, whether lower-class (LC), working-class (WC), or middle-class (MC), the (r) indices increase as the formality decreases, that is, as one reads down the columns in table 1.2.

Equally important, table 1.2 shows that the speech community is stratified linguistically with respect to social class. That is, all social classes use more instances of the non-standard variant  $\emptyset$  in more casual styles, but the proportions differ depending upon the class. The lower class (LC) has higher percentages in all styles than either the working class (WC) or the middle class (MC). Similarly, there is a gap between the percentages of the WC and the MC. Each of the three social groups uses the (r) variants in proportionately different ways, although all of them use them in stylistically similar ways. The  $\emptyset$  variant is more characteristic of the LC than of the other two classes because it is more frequent in their speech in all styles.

Results such as these, now replicated hundreds of times in dozens of disparate social circumstances, demonstrate beyond any doubt that linguistic variation is not free but is patterned. The patterns have social significance, as revealed by the systematic frequencies of linguistic variants, the dependent variable, with independent variables such as style and social class. Above all, the patterns show that data can be gathered, analyzed, and interpreted without requiring regularization. The correlations in this straightforward example show that the result is not chaotic – indeed far from it. It

is meaningful, and revealing of the way in which language encodes social relations.

It represents, moreover, a theoretical breakthrough of considerable magnitude, an aspect that has seldom been recognized. The sociolinguistic enterprise, inaugurated so auspiciously in Labov's New York survey, is one of the most significant departures from the conventions of language study in history (as discussed in §1.3).

### 1.2.3 *Figures and tables*

Before going on to an elaboration of the theoretical significance of the sociolinguistic enterprise in the next section, it is worthwhile pausing a moment here and considering the way we have presented the data in table 1.2.

Because sociolinguistic correlations are usually quantitative, like those in table 1.2, it often takes readers a few minutes of study to discover the meaning of the data. If the correlations were qualitative rather than quantitative, this would not be so. For instance, if one social group always used  $\emptyset$  where another social group always used [ɹ], then it would be obvious at a glance how they differed from one another linguistically. But actual sociolinguistic variation is much more subtle. Social groups typically differ by the proportions of particular variants they use in particular circumstances. Hence it is necessary to look closely at a matrix of numbers like table 1.2 in order to discover its trends: in this instance, that the numbers in the rows decrease from left to right, that is, from lower to higher social groups, and that the numbers also increase from top to bottom in the columns, that is, from more formal to less formal styles, and, further, that these trends are consistent in all rows and columns.

As with all quantitative disciplines, sociolinguistics also uses graphic representations of data. Table 1.2 can be graphically represented as figure 1.1. The chapters that follow use both tables and figures, occasionally together but usually one or the other, depending upon what is being discussed. All in all, I use figures more often than tables, especially for complex data. The reason for this is that figures usually require somewhat less effort for the reader to comprehend than tables. This is undoubtedly so once readers are used to their formal conventions.

Although figures and tables may differ in comprehensibility, it is important to understand that they must have exactly the same content. Figure 1.1 contains exactly the same information as table 1.2, and not an iota more

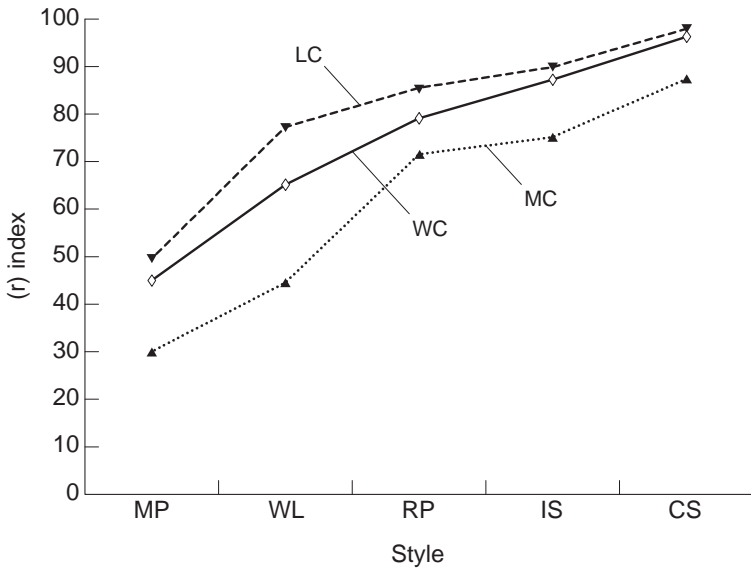


Figure 1.1 Social and stylistic stratification of (r) for three social classes in five styles in New York City (based on Labov 1966b)

or less. Every cell of the matrix in table 1.2 can be located in the graph of figure 1.1, and so can every label on its rows and columns. They are, we say, notationally equivalent. By that, we mean that their form (or notation) is different, but their content (or meaning) is exactly the same. For every table it must be possible to construct an equivalent graph, and for every graph it must be possible to construct an equivalent table.

Figure 1.1 is a typical sociolinguistic figure as far as its form goes, though (as we will see) it is a bit unusual in its content in one respect. It is a line graph, with the (r) index scores – the dependent variable – represented along the vertical axis (also called the Y-axis or the value axis) from zero at the base, and the five styles along the horizontal axis (or X-axis, or series axis). Because there are two independent variables in the data and it is only possible to represent one (at a time) on the horizontal axis, the second one, social class, is represented graphically by the symbols ▲, ◆, and ▼, and identified in the legend. This much is standard graphing procedure for any discipline.

If we look at figure 1.1 for evidence of the stratification we already



observed in table 1.2, it is immediately evident in the way the lines move across the graph separately. The fact that there are gaps between the lines means that each social group occupies its own stratum. Also more obvious in the graph is the near convergence of the two lower social groups in the least formal styles, IS and CS. At this point, the convergence threatens the stratification as the two groups nearly merge. Looking back at table 1.2, we can see that the indices at these points are 87.5 and 89.5 for IS, and 96 and 97.5 for CS.

An important convention that is peculiar to sociolinguistics is that the variants are always counted in such a way that the one that belongs to the standard accent is zero (as in §1.2.2.4 above). Once this convention is understood, it makes the “reading” of graphs very straightforward. In the case of New York (r) taking the variant [ɹ] to be standard and the other variant, Ø, to be non-standard, the index scores are calculated by determining the percentage of Ø in the total. For instance, the MC speakers used variant Ø in the MP style 30 percent of the time: 3 out of 10 times, or 48 out of 160, or something proportional.

The central purpose of making the calculation this way is to fix the archetypical standard accent at zero or, in graphic terms, along the horizontal axis. When you look at a sociolinguistic graph, then, you know that the horizontal axis represents categorically standard speech. You also know that the distance of the lines from the horizontal axis indicates their degree of non-standardness.

It is in this respect that figure 1.1 is slightly unusual. We usually expect the MC to approximate the standard. In numerous graphs later in this book, the MC will be found on a line that closely follows the horizontal axis. This is not so in figure 1.1. Instead, the MC ranges from 30 percentage points to 87.5 above the horizontal axis. The reason for the considerable gap is that Labov’s survey took place at a time of great changes in New York City speech. In the 1930s and 1940s, for reasons not fully understood, the distinctive urban vernacular of New York became the subject of jokes. Features that had formerly been distinctive, regionally prestigious indicators of New York City speech were suddenly subjected to criticism and derision. At the time of Labov’s survey in 1963 a feature like (r) was occurring as [ɹ] quite often in more self-conscious styles but still was heard as Ø more often in the other styles. The distance from the horizontal axis and the sharp rise in styles both indicate an unsettled linguistic situation, with the speakers in a transitional phase.

Apart from this minor oddity, the dependent variable (r) has prototypical attributes. It varies systematically with two independent variables. As

style becomes more casual, the  $\emptyset$  variant becomes more frequent for all speakers in the community. As one goes down the social class hierarchy, the  $\emptyset$  variant becomes more frequent in all styles. (Obviously, these statements could readily be rewritten as their converses, in terms of the [ɹ] variant in more formal styles and up the social hierarchy.)

Correlations like these are crucial. *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. Labov's New York survey demonstrated more convincingly than anything that came before it that linguistic variation correlates with social factors, and hence that language variation is not only amenable to analysis but also linguistically interesting and socially revealing.

### 1.3 Variation and the Tradition of Categoricity

The variable as a structural unit represents a momentous innovation in linguistics. Hitherto, all linguistic units – phones, phonemes, morphemes, phrases, clauses – had been invariant, discrete, and qualitative. The variable is none of these. Instead, it is variant, continuous, and quantitative (Labov 1966b). It is variant in the sense that it is realized differently on different occasions. It is continuous in the sense that certain variants, such as the vowel gradations for (eh) above, take on social significance depending upon their phonetic distance from the standard variant, or, as with the variants for (r), their phonetic differentness from the standard variant. It is quantitative in the sense that its significance is not determined merely by the presence or absence of its variants but by their relative frequency.

#### 1.3.1 *Langue and parole*

The variable can only exist in a theory that abandons the axiom of categoricity. Until the inception of modern sociolinguistics, all major linguistic theories adopted the axiom of categoricity. The domain of linguistic investigation was taken to be, in Saussure's celebrated dichotomy, *langue*, the grammatical system, rather than *parole*, the social uses of language (1916: 9–15). The reason is precisely because *langue* is removed from the

turbulence in which *parole* exists, the quotidian flux, or in Sapir's terms (1929: 166) "the apparent lawlessness of social phenomena." *Langue* is homogeneous, Saussure says, whereas *parole* is heterogeneous. Furthermore, *langue* can be studied in the absence of a community of speakers.

This last point leads to what Labov (1972a: 185–6) calls the "Saussurean Paradox," whereby the data from a single speaker, often linguists relying on their own intuitions, are all that are needed to study the communal *langue*, but in order to study the individualistic *parole* one must study language in the community.

Distinctions similar to *langue* and *parole* can be found both before and after Saussure. Humboldt ([1836] 1972: 129) observed: "Language comprehends in fact two contrasting properties: namely, it is divided up into an infinity as the sole language in one and the same nation; and at the same time these many variants are united into one language having a definite character." The two properties presumably underlie Humboldt's distinction between language as "an *ergon*, that is a mere means of exchange for purposes of communication," and as "an *energeia* which reconstitutes human experience ideally" (Basilius 1952: 98).

Saussure's distinction is paralleled closely by Chomsky's equally well-known distinction between competence, "the speaker-hearer's knowledge of his language," and performance, "the actual use of language in concrete situations" (1965: 4).

### 1.3.2 *The axiom of categoricity*

The decision that the proper domain of linguistics should be homogeneous *langue* rather than heterogeneous *parole* – or the united *energeia* rather than the variable *ergon*, or the speaker-hearer's knowledge rather than actual use – aroused very little debate historically. Saussure put it bluntly: "One might if really necessary apply the term linguistics to each of the two disciplines, and speak of a linguistics of speaking [*parole*], but that science must not be confused with linguistics proper, whose sole object is language [*langue*]" (1916: 19–20). For Saussure, a science of *parole* was simply a contradiction in terms. "As soon as we give language [*langue*] first place among the facts of speech," he says (1916: 9), "we introduce a natural order into a mass that lends itself to no other classification." Chomsky, in his early writings, concurred: "observed use of language . . . surely cannot constitute the actual subject matter of linguistics, if this is to be a serious discipline" (1965: 4, but see the following paragraphs for Chomsky's ma-

ture view). Generations of language students simply assumed that the only way to study language would be by approaching it at some remove from its real-life performance. Therefore the object of study, *langue-energeia*-competence, had to be abstracted from its real-world contexts.

The axiom of categoricity, the basis for the abstraction, has been stated in many ways. We have already noted (in §1.2 above) Joos's characterization of the proper subject matter as language in which "all continuities, all possibilities of infinitesimal gradation, are shoved outside of linguistics in one direction or the other." Here is Hjelmslev (1961: 5–6): "Linguistics must attempt to grasp language, not as a conglomerate of non-linguistic (e.g., physical, physiological, psychological, logical, sociological) phenomena, but as a self-sufficient totality, a structure *sui generis*."

Chomsky, with the kind of clarity that sets him apart from all other linguists, is the author of what is probably the most explicit, and certainly the best-known, statement of the axiom of categoricity:

Linguistic theory is concerned primarily with an ideal speaker–listener, in a completely homogeneous speech-community, who knows its language perfectly and is unaffected by such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interest, and errors (random or characteristic) in applying his knowledge of the language in actual performance. (Chomsky 1965: 3)

Even though this idealization underlies virtually all linguistic research before sociolinguistics – including Bloomfieldian structuralism<sup>3</sup> whose practitioners dissociated themselves from the Chomskyan enterprise – it attracted considerable criticism.

Its basic contention is simply to point out that linguists in this tradition take as their data what Labov once called, in a memorable phrase, "normative edited texts": sentences or pronunciations based on the linguist's intuitions recollected in tranquility, or based on an informant's intuitions, recollected under the linguist's direction.

The removal of linguistic data from its social context was always a requisite in studies of *langue*, and Chomsky's statement simply makes the removal glaringly obvious. Nevertheless, the non-social world Chomsky depicts irked some linguists, including many who were, apparently unwittingly, working in it. Their objections to the passage were so persistent that Chomsky felt enjoined to respond several years later, pointing out that he had intended the statement as "an innocent and uncontroversial statement of an idealization." Innocent it is but not, it turned out, uncontroversial.

sial. In discussing the various objections to the statement, Chomsky makes the following points:

Exactly what is the source of objection? Obviously it cannot be that real speech communities are not homogeneous. That is as obvious as it is irrelevant. Nor can the objection be that this idealization necessarily impedes the study of matters that do not fall within it, say, linguistic variation or what Putnam calls “the social division of linguistic labor.” On the contrary, what is implicitly claimed by someone who adopts the idealization is that these further questions are properly studied within a framework that makes use of the results obtained by inquiring into the idealization. If the idealization does make it possible to unearth real and significant properties of the language faculty, this conclusion would seem to be justified, indeed inescapable. (Chomsky 1980: 24–5)

Chomsky is right, of course. The accomplishments of linguistics carried out under the “idealization” – the axiom of categoricity – are not really disputable. Labov (1972a: 186) concurs: “The science of *parole* never developed [prior to sociolinguistics], but this approach to the science of *langue* has been extremely successful over the past half-century.” Chomsky’s response, however, tacitly revises his earlier pronouncement (cited above) that linguists could not profitably study the “observed use of language.” Between that earlier statement and Chomsky’s tacit revision 15 years later, of course, sociolinguistics had come into being and made tremendous strides in the development of the science of *parole*.

### 1.3.3 *Communicative competence*

*Langue* and *parole* remain useful distinctions today for a reason that Saussure would undoubtedly have found unimaginable, because they now help to define the different objects of inquiry of theoretical linguistics and sociolinguistics. They are separable in theory as natural partitions of the language faculty, or what might plausibly be considered distinct cognitive modules.

Chomsky has argued for the language faculty as comprised of interacting systems as “‘mental organs’ analogous to the heart or the visual system or the system of motor coordination and planning” (1980: 39). Theoretical linguists who adopt the axiom of categoricity are primarily interested in discovering the properties of one of those systems of the language faculty, called *grammar*, conceived as a language-specific bioprogram (Bickerton’s

1984 incisive but under-used term). The grammar is made up of, in Chomsky's terms (1980: 55), "a system of 'computational' rules and representations." Attempts at discovering its innate computational properties have led Chomsky and his followers into minute examinations of surface-structure puzzles involving linguistic coreference, scope, and other structural intricacies. They have produced insights into the grammatical processor as "structure-dependent" rather than strictly linear (cf. Crain and Nakayama 1987) and, crucially for Chomsky's tenacious but disputed stance on the grammatical component's language-specificity, not reducible to other, independently motivated, non-language-processing cognitive systems. (On the dispute, see Sabbagh and Gelman 2000 and references therein.)

The grammar is presumably the module in the language faculty that accounts for the uniquely human attributes of creativity in language production and comprehension, and for the rapidity of language acquisition in infancy. However, it is obviously not autonomous. Linguistic production and comprehension require real-world orientation to express meanings, and the acquisition device requires the stimulus of social interaction to activate learning. Chomsky, of course, recognizes its interdependence with other systems, and he has isolated two of them as follows: "A fuller account of knowledge of language will consider the interactions of grammar and other systems, specifically the system of conceptual structures and pragmatic competence, and perhaps others" (1980: 92). The real-world orientation has its source in what Chomsky calls the *conceptual system*, and the social stimulus has its source in what Chomsky calls "pragmatic competence" but is better known as *communicative competence*.

By the conceptual system, Chomsky means "the system of object-reference and also such relations as 'agent,' 'goal' and 'instrument'; what are sometimes called 'thematic relations'" (1980: 54). It also includes vocabulary items, the most obvious intermediaries between grammar and the world. The conceptual system has received little attention from linguists, but it too reveals uniquely human properties most easily discerned in acquisition. Children master fine semantic distinctions of the sort found in verbs such as *follow* and *chase* relatively early, certainly long before they can consciously define what they mean. They universally develop lexical distinctions in number and color categorization that are unmatched by, say, olfactory categories (Strozer 1994: 40–5). These fine vocabulary distinctions recur in all natural languages. One way of explaining this mastery, Chomsky (1988: 31) says, is by postulating that words "enter into systematic structures based on certain elementary recurrent notions and principles of combination." More generally, he says, "The rate of vocabu-

lary acquisition is so high at certain stages of life, and the precision and delicacy of the concepts acquired so remarkable, that it seems necessary to conclude that in some manner the conceptual system with which lexical items are connected is already substantially in place" (1980: 139). These are stimulating ideas that invite empirical research.

Chomsky's third cognitive module, "pragmatic competence," takes in, in his words, "knowledge of conditions and manner of appropriate use, in conformity with various purposes. . . . We might say that pragmatic competence places language in the institutional setting of its use, relating intentions and purposes to the linguistic means at hand" (1980: 224–5). This notion has long been familiar to sociolinguists. It was influentially described by Hymes as "sociolinguistic competence" or "communicative competence" (1974), as follows:

Within the social matrix in which [a child] acquires a system of grammar, a child acquires also a system of its use, regarding persons, places, purposes, other modes of communication, etc. – all the components of communicative events, together with attitudes and beliefs regarding them. There also develop patterns of the sequential use of language in conversation, address, standard routines, and the like. In such acquisition resides the child's sociolinguistic competence (or, more broadly, communicative competence), its ability to participate in its society as not only a speaking, but also a communicating member. (Hymes 1974: 75)

Hymes adds, "What children so acquire, an integrated theory of sociolinguistic description must be able to describe."

Like the other organs of the language faculty, communicative competence develops early and rapidly in normal children with little or no tutoring. Since most of the conventions governing communicative events are beneath consciousness, explicit teaching is impossible in any case. Evidence for communicative competence as an entity independent of grammatical competence (and presumably the other organs of the language faculty) can be found in extreme social situations and in clinical settings in which people are forced to function with one in the absence of the other.

Such an extreme social situation has come to light in research by Dorian (1982) in East Sutherland in the Scottish Highlands. Dorian's subjects included bilingual Gaelic-English fisherfolk, younger English-speaking monolinguals, and a middle group of English speakers described as "low-proficiency semi-speakers of East Sutherland Gaelic and . . . near-passive bilinguals" (1982: 27). This third group, despite their lack of grammatical competence in Gaelic, interacted freely and comfortably with their Gaelic

neighbors. They were fully integrated in the bilingual community, and their integration depended largely, or perhaps solely, upon their communicative competence. "They knew when it was appropriate to speak and when not," Dorian says (1982: 29), "when a question would show interest and when it would constitute an interruption; when an offer of food or drink was mere verbal routine and was meant to be refused, and when it was meant in earnest and should be accepted; how much verbal response was appropriate to express sympathy in response to a narrative of ill health or ill luck; and so forth."

Their communicative competence was so perfectly attuned, in fact, that neither the fluent bilinguals nor the semi-speakers themselves were aware of the extent of their grammatical shortcomings. In one instance, Dorian inadvertently exposed those shortcomings by testing the language proficiency of one of the semi-speakers in the presence of her bilingual friends, to the considerable embarrassment of everyone, including Dorian. In the East Sutherland speech community, they were all peers by dint not of their shared language but rather of their shared communicative competence.

Further evidence for the independence (or modularity) of communicative competence is revealed by the fact that communicative competence can be disturbed and disrupted in neurological disorders. People suffering from what is called "semantic-pragmatic disorder" tend to interrupt the conversational flow with inappropriate or ill-timed assertions, fail to follow topics, introduce what appear to be digressions or non sequiturs, and speak out of turn (Bishop and Adams 1989, Mogford-Bevan and Sadler 1991). Typically, their speech is phonologically and grammatically well-formed, and not infrequently their speech is remarkably fluent.

Yamada (1990) provides a meticulously detailed, long-term case study of an American girl named Laura with severe cognitive deficiencies. Her speech was generally grammatical but, Yamada says, she "often failed to make use of the necessary discourse conventions." As a result, "conversing with Laura was frequently an odd, unpredictable experience," Yamada (1990: 66) says. "Sometimes she remained silent except for monosyllabic grunts in response to questions and comments, and then suddenly started off on one of her spiels, either in response to someone else's comments or triggered by some unseen, internal stimulus."

What malfunctions linguistically in Laura and other people who are afflicted with semantic-pragmatic disorder is their communicative competence. They provide evidence, in Yamada's terms (1990: 5), "that grammatical knowledge can exceed pragmatic knowledge (or emerge in its absence)." Just as myxedema proves the existence of the thyroid gland in



the endocrine system (if proof were needed), so semantic-pragmatic disorders prove the existence of communicative competence as an autonomous component of the language faculty.

#### 1.3.4 *Linguistics without categoricity*

The possibility of directly studying communicative competence had seldom been considered prior to sociolinguistics. Although Chomsky's critics complained about the lack of realism in his 1965 statement about the ideal speaker–listener in a homogeneous speech community, contemporary proposals for a program of linguistic research *without* the idealization were very rare.

One of the clearest came from Bolinger four years before Chomsky's statement of the axiom:

From the standpoint of what has become traditional in American linguistics, the question is not whether there are such things as continuous phenomena in parts of human behavior that lie close to linguistics – many would grant that there are – but whether such phenomena should be regarded as the object of linguistic study. It comes down more to a scheme of priorities than to a denial of possible ultimate importance: a higher value, for the present, is placed on phenomena that lend themselves to an all-or-none analysis. (Bolinger 1961: 10)

Bolinger's interests lay primarily in phonetic gradience rather than in sociolinguistic variation, but his contention that categorical (all-or-none) linguistics was a matter of priority rather than necessity was ahead of its time. It was borne out a few years later when sociolinguistics came into being.

Chomsky, in contrast to both Saussure who held out no hope for studying *parole* and Bolinger who believed that *parole* could be studied if linguists thought it as interesting as *langue*, felt that there was an ordered relationship between the two in which the study of *langue* was necessarily the precursor of the study of *parole*. He maintained that any “theory of performance” would necessarily be dependent upon a theory of competence. “Performance models,” he says, must “incorporate generative grammars of specific kinds,” or at least incorporate “assumptions about underlying competence” (Chomsky 1965: 10; similarly, 1980: 24–8).

There is a sense in which his claim appears to me to be true, and another sense in which it seems quite false. On the one hand, it is true, I think, that any worthwhile non-categorical theory such as a theory of

linguistic variation must incorporate – or at least must be consistent with and must not contradict – the general properties of linguistic competence that have emerged with wonderful clarity in the Chomskyan paradigm. By “general properties” I mean language as a mental faculty, species-specific and innately predetermined, which when stimulated by social experience creates a grammar by setting parameters on universal principles that may in some instances be structure-dependent and thus independent of non-linguistic cognitive principles.

On the other hand, it is certainly not true – and Chomsky might even agree with this in view of the way categorical linguistics has developed since 1965 when he first stated the idealization – that variation theory must incorporate or in any other way take account of the specific postulates of categorical grammar. That is to say, variation theory need not incorporate notions like – to cite just a few – the affix shift transformation (Chomsky 1957: 39–42), the Katz–Postal principle (Chomsky 1965: 132), the specified-subject condition (Chomsky 1973), the root clause filter (Chomsky and Lasnik 1977: 486), or the antecedent trace chain (Chomsky 1988: 116–17).<sup>4</sup> These postulates gather dust with dozens of others in the generativist scrapyard that is surely one of the most bizarre and tragicomic residues of any intellectual tradition. Any theory of *parole* that adopted those elaborate postulates would necessarily share their evanescence.

The reason that these postulates come and go so quickly is presumably the gulf between the axiom of categoricity as presently formulated and empirical reality. As Naro (1980: 159) put it, “the construction of the sort of elaborate formal theories achieved in the generative school requires many clear-cut decisions on very esoteric sentence types. This is possible because data collection is so easy for this school, and the data sources are malleable enough to conform to the theorist’s wildest desires.” Until grammarians embrace data-gathering methods of greater real-world fidelity, theoretical linguistics is destined to continue proliferating postulates with few effective constraints. Recent movements in the direction of acknowledging variability and attempting to integrate it into theoretical models show some promise in this regard.

It does not seem to me to be problematic for sociolinguistic theory to avoid specific proposals of theoretical linguistics while embracing its general conception. The root clause filter and all the other trappings are in no way integral to Chomsky’s view of the language faculty. Luckily, categorical theory and variation theory are separate enough that they need only share the general view of the language faculty. That shared view marks their common ground as linguistic theories. Beyond that, they have their

own domains and ways of proceeding. Their common ground as interacting systems of the language faculty promises an exciting new domain for future exploration.

### 1.3.5 *Categorical theory and variation theory*

The first attempts by theoretical linguists at incorporating linguistic variation are now under way. They have been inspired by the successes of a generation of sociolinguists in demonstrating that language variation is orderly and interesting. If theoretical studies incorporating phonological and grammatical variation are going to become a productive area of research, they will do so by relaxing the axiom of categoricity, or perhaps by abandoning it altogether. In doing so, they will align themselves, in effect, with aims that were stated by sociolinguists from the beginning.

In the earliest proposals for variable rules, Labov and others clearly conceived of them as a refinement of the optional rules in contemporary generative theory (Labov 1972b: 93–5, Wolfram and Fasold 1974: 99–100). Cedergren and Sankoff (1974: 352) said, “The full importance of variable rules can be appreciated only from a certain paradigmatic viewpoint, one which constitutes a slight but distinct shift from generative theory.” They then expressed the need for generative theory to, first, broaden its notion of competence by including variability, and, second, use actual speech samples as data rather than intuitions. But those changes, far from slight, would amount to generative theory giving up the axiom of categoricity. That, as we have seen, would move the domain of study away from *langue-energeia*–competence.

The fundamental difference between variationist and categorical theories was surprisingly well understood and astutely explained by Gleason (1961: 391–2), in what was the standard introductory textbook in linguistics from its first edition in 1955 until the 1970s. In a chapter called “Variation in Speech,” Gleason outlined how the “descriptive linguist” goes about assembling a “narrowed corpus” by “restricting his attention as far as possible to utterances produced by one speaker under a single set of circumstances.” In the terms of reference we have been using in this chapter, Gleason is describing the effort of a categorical linguist in thinking up data or eliciting it from a native speaker.

Gleason then goes on to contrast that methodology with “another quite valid, but basically very different, approach” which could be put into practice. This approach would take an unregularized corpus and study the

variants in it “by seeking correlations with non-linguistic factors, commonly the speaker and the circumstances.” He continues: “Obviously, the results are predestined to be fundamentally different from those which the descriptive linguist will attain, since the variation under examination is precisely that which the descriptivist will attempt to eliminate.”

Again, in our terms, he is suggesting that a linguist might gather data by, say, recording a conversation and then analyzing the usage of the participants by looking at how the variants in their speech are determined by social conditions. “Here is the basis for a second type of linguistic science,” he says. “Since most workers have restricted their attention to single aspects of the problem, we lack a general term for the discipline as a whole.” The general term, a few years later, came to be sociolinguistics, and its methods took shape very much along the lines that Gleason described. Although he never tried to apply those methods himself, he clearly understood their singularity, their fundamental differentness from the mainstream tradition in linguistics.

Gleason’s chapter, despite the wide distribution of the book it appeared in, was not a direct influence on the new discipline. Neither was Currie’s equally programmatic proposal that coined the word sociolinguistics. Both predated the actual practice, and Gleason’s outline was remarkably prescient. Both were “in the air” when Labov laid the groundwork for discovering patterns where his predecessors had found only noise.

Among the most promising recent developments in theoretical syntax and phonology are attempts at relaxing the axiom of categoricity along the lines promulgated by Cedergren and Sankoff (1974) in order to encompass linguistic variation (for example, Hayes 2001, Henry 2002, Anttila 2002). The motivation for these initiatives, which should ultimately redefine the venerable *langue–parole* distinction as poles on a continuum, comes directly from sociolinguistic breakthroughs in dealing with variable data. In a larger sense, it represents a general re-alignment of goals that has taken place in numerous disciplines in recent times.

### 1.3.6 *Categoricity in other disciplines*

Around the same time that Labov put into practice this second type of linguistic science dispensing with the axiom of categoricity, researchers in several other disciplines were doing exactly the same thing.

The axiom of categoricity is not, of course, unique to linguistics. Mathematics is probably its most celebrated host. It was once known as the

“perfect science” simply because it admitted no exceptions, was perfectly explicit, and was never stuck for an answer. That is no longer so. For almost 50 years, several branches have been inquiring into the mathematical properties of variability. Fuzzy set theory studies the considerable implications if an object is a member of a set not categorically but partially, to some intermediate degree (Zadeh 1965). Many-valued logic computes multidimensional truth tables based on values part-way between “true” and “false” (Rescher 1969). Physicists are tracking the geometry of behavior without such protective idealizations as non-friction and non-gravity in an emerging field called “nonlinear dynamics” (Abraham and Shaw 1992). Scientists interested in the mathematical properties of such diverse natural phenomena as cloud shapes, spider webs, and hurricanes are working out a new branch of mathematics called “chaos” (Stewart 1990, Alligood et al. 1997). Social scientists are working at adapting “catastrophe theory” (Saunders 1980, Schneider 1997), the mathematical model of discontinuity in dynamic systems, to all kinds of formerly unruly phenomena.

If branches of mathematics can give up the axiom of categoricity, then so can branches of other disciplines, and many have, including philosophy, psychology, geography, and statistics as well as linguistics. Looked at from this general perspective, the variationist movement is surely one of the most significant in the intellectual history of our time.

What is under investigation in many of these disciplines is the idea of categories themselves (J. Taylor 1989, Pinker and Prince 1999). What happens if we give up the idealization that categories are intensionally defined and have sharp boundaries? Observations of real-world phenomena show that, in fact, most categories have odd or variable members and fuzzy or indefinite borders. The real revelation of contemporary intellectual history has been the discovery that variability is explicable and fuzziness has form. In other words, once we stop filtering real-world phenomena through the axiom of categoricity it begins to look as if the idealization was merely a convenient fiction, erected in order to simplify reality and make data manageable. Its abandonment by researchers in so many disciplines after so many centuries is undoubtedly fostered partly because we now have better tools – laser beams, microcircuitry, fibroptic scanners, acoustic microscopes, silicon chips – for managing and analyzing data.

One of the more lucid general inquiries into categoricity and variability was made by William James (1911), who distinguished “percepts,” the apprehension of reality, from “concepts,” the idealization of reality. “There must always be a discrepancy between concepts and reality,” he said (1911: 365), “because the former are static and discontinuous, and the latter are

dynamic and flowing.” Although James, like his contemporary Saussure, probably did not foresee that scholars might one day study percepts directly, he, unlike Saussure, recognized their primacy:

The deeper features of reality are found only in perceptual experience. Here alone do we acquaint ourselves with continuity, or the immersion of one thing in another, here alone with self, with substance, with qualities, with activity in its various modes, with time, with cause, with changes, with novelty, with tendency, and with freedom. (James 1911: 96)

James’s rhetoric seems to me to capture something of the much more modern spirit that underlies work in chaos theory and in fuzzy semantics and in language variation.

Once the axiom of categoricity has been abandoned, one of the filters separating the “features of reality” from their investigators is removed. The attraction of studying the universe without the protective shield of the axiom of categoricity is counterbalanced, predictably, by the recalcitrance of the data it yields. As the statistician Moroney (1957: 5) says, “Life and nature may be simple enough to the Almighty who designed them and keeps them going, but to the human mind there is presented an unending stream of problems that cannot be given a clear-cut answer of the type  $p = 1$  or  $p = 0$ .” The answers 1 and 0 are of course categorical, and if the problems we confront have answers that belong in the interval between 1 and 0, then we have an unbounded set of possibilities. The challenges of observing, analyzing, and understanding percepts are imposing. Facing those challenges is a relatively recent development in the natural and social sciences.

Sociolinguistics has met those challenges in interesting, often exciting, ways. The rest of this book presents observations, analyses and interpretations made in the core area of sociolinguistics, the study of the social significance of linguistic variation.