

2 Languages, Dialects, and Varieties

I stated in the introductory chapter that all languages exhibit internal variation, that is, each language exists in a number of varieties and is in one sense the sum of those varieties. But what do we mean by *variety*? Hudson (1996, p. 22) defines a variety of language as ‘*a set of linguistic items with similar distribution*,’ a definition that allows us to say that all of the following are varieties: Canadian English, London English, the English of football commentaries, and so on. According to Hudson, this definition also allows us ‘to treat all the languages of some multilingual speaker, or community, as a single variety, since all the linguistic items concerned have a similar social distribution.’ A variety can therefore be something greater than a single language as well as something less, less even than something traditionally referred to as a dialect. Ferguson (1972, p. 30) offers another definition of variety: ‘any body of human speech patterns which is sufficiently homogeneous to be analyzed by available techniques of synchronic description and which has a sufficiently large repertory of elements and their arrangements or processes with broad enough semantic scope to function in all formal contexts of communication.’ Note the words ‘sufficiently homogeneous’ in this last quotation. Complete homogeneity is not required; there is always some variation whether we consider a language as a whole, a dialect of that language, the speech of a group within that dialect, or, ultimately, each individual in that group. Such variation is a basic fact of linguistic life.

Hudson and Ferguson agree in defining *variety* in terms of a specific set of ‘linguistic items’ or ‘human speech patterns’ (presumably, sounds, words, grammatical features, etc.) which we can uniquely associate with some external factor (presumably, a geographical area or a social group). Consequently, if we can identify such a unique set of items or patterns for each group in question, it might be possible to say there are such varieties as Standard English, Cockney, lower-class New York City speech, Oxford English, legalese, cocktail party talk, and so on. One important task, then, in sociolinguistics is to determine if such unique sets of items or patterns do exist. As we proceed we will encounter certain difficulties, but it is unlikely that we will easily abandon the concept of ‘variety,’ no matter how serious these difficulties prove to be.

Discussion

1. I have just suggested that, although a concept like ‘variety’ is difficult to define, it may still be useful in sociolinguistic work. Linguists have found such concepts as ‘sound,’ ‘syllable,’ ‘word,’ and ‘sentence’ equally difficult to define (in contrast to lay usage, in which they are just assumed to be obvious and uncontroversial). In one sense, linguistics is all about trying to provide adequate definitions for words such as *sound*, *syllable*, *word*, *sentence*, and *language*. What are some of the problems you are aware of concerning the linguist’s difficulty with these words and the associated concepts? What parallels do you see, if any, between these problems and the sociolinguist’s problem with *variety* (and the other terms to be used in the remainder of this chapter)?
2. Hymes (1974, p. 123) has observed that language boundaries between groups are drawn not on the basis of the use of linguistic items alone, because attitudes and social meanings attached to those items also count. He says:

Any enduring social relationship or group may come to define itself by selection and/or creation of linguistic features, and a difference of accent may be as important at one boundary as a difference of grammar at another. Part of the creativity of users of languages lies in the freedom to determine what and how much linguistic difference matters.

How does this inter-relationship between linguistic items and the social evaluations of such items apply in how we regard each of the following pronunciations?

- a. butter, budder, bu’er
- b. fishing, fishin’
- c. farm, fahm
- d. width pronounced like wit, like with
- e. Cuba pronounced as Cuber
- f. ate pronounced like eight, like et
- g. been pronounced like bean, like bin
- h. mischievous pronounced with four syllables (i.e., as mischievious)
- i. aluminum, aluminium
- j. pólice, gúitar, Détroit (with the stress as indicated)

And each of the following utterances?

- a. He hurt hisself.
- b. He done it.
- c. He dove in.
- d. He run away last week.
- e. It looks like it’s going to rain.
- f. To whom did you give it?
- g. She’s taller than me now.

- h. Yesterday he laid down after lunch for an hour.
- i. Can I leave the room?
- j. He ain't got no money left.
- k. Try and do it soon.
- l. Between you and me, I don't like it.
- m. There's twenty dollars for you to spend.
- n. She invited Sally and I to the party.
- o. I wants it.
- p. You done it, did you?
- q. Stand over by them boys.
- r. Is he the one what said it?
- s. They don't learn you nothing there.

Language and Dialect

For many people there can be no confusion at all about what language they speak. For example, they are Chinese, Japanese, or Korean and they speak Chinese, Japanese, and Korean respectively. It is as simple as that; language and ethnicity are virtually synonymous (Coulmas, 1999). A Chinese may be surprised to find that another person who appears to be Chinese does not speak Chinese, and some Japanese have gone so far as to claim not to be able to understand Caucasians who speak fluent Japanese. Just as such a strong connection between language and ethnicity may prove to be invaluable in nation-building, it can also be fraught with problems when individuals and groups seek to realize some other identity, e.g., to be both Chinese and American, or to be Canadian rather than Korean-Canadian. As we will see (p. 368), many Americans seem particularly reluctant to equate language with ethnicity in their own case: although they regard English as the 'natural' language of Americans, they do not consider American to be an ethnic label. The results may be the same; only the reasons differ.

Most speakers can give a name to whatever it is they speak. On occasion, some of these names may appear to be strange to those who take a scientific interest in languages, but we should remember that human naming practices often have a large 'unscientific' component to them. Census-takers in India find themselves confronted with a wide array of language names when they ask people what language or languages they speak. Names are not only ascribed by region, which is what we might expect, but sometimes also by caste, religion, village, and so on. Moreover, they can change from census to census as the political and social climate of the country changes.

While people do usually know what language they speak, they may not always lay claim to be fully qualified speakers of that language. They may experience difficulty in deciding whether what they speak should be called a *language* proper or merely a *dialect* of some language. Such indecision is not surprising: exactly how do you decide what is a language and what is a dialect of a language? What

criteria can you possibly use to determine that, whereas variety X is a language, variety Y is only a dialect of a language? What are the essential differences between a language and a dialect?

Haugen (1966a) has pointed out that *language* and *dialect* are ambiguous terms. Ordinary people use these terms quite freely in speech; for them a dialect is almost certainly no more than a local non-prestigious (therefore powerless) variety of a real language. In contrast, scholars often experience considerable difficulty in deciding whether one term should be used rather than the other in certain situations. As Haugen says, the terms 'represent a simple dichotomy in a situation that is almost infinitely complex.' He points out that the confusion goes back to the Ancient Greeks. The Greek language that we associate with Ancient Greece was actually a group of distinct local varieties (Ionic, Doric, and Attic) descended by divergence from a common spoken source with each variety having its own literary traditions and uses, e.g., Ionic for history, Doric for choral and lyric works, and Attic for tragedy. Later, Athenian Greek, the *koiné* – or 'common' language – became the norm for the spoken language as the various spoken varieties converged on the dialect of the major cultural and administrative center. Haugen points out (p. 923) that the Greek situation has provided the model for all later usages of the two terms with the resulting ambiguity. *Language* can be used to refer either to a single linguistic norm or to a group of related norms, and *dialect* to refer to one of the norms.

The situation is further confused by the distinction the French make between *un dialecte* and *un patois*. The former is a regional variety of a language that has an associated literary tradition, whereas the latter is a regional variety that lacks such a literary tradition. Therefore *patois* tends to be used pejoratively; it is regarded as something less than a dialect because of its lack of an associated literature. Even a language like Breton, a Celtic language still spoken in parts of Brittany, is called a *patois* because of its lack of a strong literary tradition and the fact that it is not some country's language. However, *dialecte* in French, like *Dialekt* in German, cannot be used in connection with the standard language, i.e., no speaker of French considers Standard French to be a dialect of French. In contrast, it is not uncommon to find references to Standard English being a dialect – admittedly a very important one – of English.

Haugen points out that, while speakers of English have never seriously adopted *patois* as a term to be used in the description of language, they have tried to employ both *language* and *dialect* in a number of conflicting senses. *Dialect* is used both for local varieties of English, e.g., Yorkshire dialect, and for various types of informal, lower-class, or rural speech. 'In general usage it therefore remains quite undefined whether such dialects are part of the "language" or not. In fact, the dialect is often thought of as standing outside the language. . . . As a social norm, then, a dialect is a language that is excluded from polite society' (pp. 924–5). It is often equivalent to *nonstandard* or even *substandard*, when such terms are applied to language, and can connote various degrees of inferiority, with that connotation of inferiority carried over to those who speak a dialect.

We can observe too that questions such as 'Which language do you speak?' or 'Which dialect do you speak?' may be answered quite differently by people

who appear to speak in an identical manner. As Gumperz (1982a, p. 20) has pointed out, many regions of the world provide plenty of evidence for what he calls 'a bewildering array of language and dialect divisions.' He adds: 'socio-historical factors play a crucial role in determining boundaries. Hindi and Urdu in India, Serbian and Croatian in Yugoslavia [of that date], Fanti and Twi in West Africa, Bokmål and Nynorsk in Norway, Kechwa and Aimara in Peru, to name just a few, are recognized as discrete languages both popularly and in law, yet they are almost identical at the level of grammar. On the other hand, the literary and colloquial forms of Arabic used in Iraq, Morocco, and Egypt, or the Welsh of North and South Wales, the local dialects of Rajasthan and Bihar in North India are grammatically quite separate, yet only one language is recognized in each case.'

The Hindi-Urdu situation that Gumperz mentions is an interesting one. Hindi and Urdu are the same language, but one in which certain differences are becoming more and more magnified for political and religious reasons. Hindi is written left to right in the Devanagari script, whereas Urdu is written right to left in the Arabic-Persian script. Whereas Hindi draws on Sanskrit for its borrowings, Urdu draws on Arabic and Persian sources. Large religious and political differences make much of small linguistic differences. The written forms of the two varieties, particularly those favored by the elites, also emphasize these differences. They have become highly symbolic of the growing differences between India and Pakistan. (We should note that the situation in India and Pakistan is in almost direct contrast to that which exists in China, where mutually unintelligible Chinese languages (called 'dialects' by the Chinese themselves) are united through a common writing system and tradition.)

Gumperz (1971, pp. 56-7) points out that everyday living in parts of India, particularly in the large cities and among educated segments of those communities, requires some complex choices involving the distinction between Hindi and Urdu:

Since independence Hindi has become compulsory in schools, but Urdu continues to be used extensively in commerce, and the Ghazal, the best known form of Urdu poetry, is universally popular. If we look at the modern realist Hindi writers, we find that they utilize both Sanskrit and Persian borrowings. The juxtaposition of the two styles serves to express subtle shades of meaning and to lend reality to their writings. Similarly on the conversational level the use of Hindi and Urdu forms is not simply a matter of birth and education. But, just as it is customary for individuals to alternate between dialect and standard depending on the social occasion, so when using the standard itself the speaker may select from a range of alternatives. Hindi and Urdu therefore might best be characterized not in terms of actual speech, but as norms or ideal behavior in the sociologist's sense. The extent to which a speaker's performance in a particular communication situation approximates the norm is a function of a combination of factors such as family background, regional origin, education and social attitude and the like.

So far as everyday use is concerned, therefore, it appears that the boundary between the spoken varieties of Hindi and Urdu is somewhat flexible and one that changes with circumstances. This is exactly what we would expect: there

is considerable variety in everyday use but somewhere in the background there is an ideal that can be appealed to, proper Hindi or proper Urdu.

In the first of the two quotations from Gumperz there is a reference to Yugoslavia, a country now brutally dismembered by the instruments of ethnicity, language, and religion. Within the old Yugoslavia Serbs and Croats failed to agree on most things and after the death of President Tito the country, slowly at first and then ever more rapidly later, fell into a fatal divisiveness. Slovenians and Macedonians excised themselves most easily, but the Serbs and the Croats were not so lucky. Linguistically, Serbo-Croatian is a single South Slav language but one used by two groups of people, the Serbs and Croats, with somewhat different historical, cultural, and religious backgrounds. There is a third group in Bosnia, a Muslim group, who also speak Serbo-Croatian, and their existence further compounded the problems and increased the eventual bloodshed. Finally, there is a very small Montenegrin group. The Serbian and Croatian varieties of Serbo-Croatian are known as *srpski* and *srpskohrvatski* respectively. The actual differences between them involve different preferences in vocabulary rather than differences in pronunciation or grammar. That is, Serbs and Croats often use different words for the same concepts, e.g., Serbian *varos* and Croatian *grad* for 'train.' The varieties are written in different scripts (Roman for Croatian and Cyrillic for Serbian), which also reflect the different religious loyalties of Croats and Serbs (Catholic and Orthodox). As conflict grew, differences became more and more important and the country and the language split apart. Now in Serbia people speak Serbian just as they speak Croatian in Croatia. Serbo-Croatian no longer exists as a language of the Balkans. And now that there is a separate Bosnia the Bosnians call their variety *bosanski* and Montenegrins call their variety *crnogorski* (Carmichael, 2002, p. 236, and Greenberg, 2004).

In direct contrast to the above situation, we can observe that the loyalty of a group of people need not necessarily be determined by the language they speak. Although the majority of the people in Alsace are speakers of a variety of German insofar as the language of their home-life is concerned, their loyalty is unquestionably toward France. They look west not east for national leadership and they use French, not German, as the language of mobility and higher education. However, everyday use of Alsatian is a strong marker of local identity; it is an important part of being Alsatian in France. We can contrast this situation with that in another area of France. In Brittany a separatist movement, that is, a movement for local autonomy if not complete independence, is centered on Breton, a language which, unfortunately for those who speak it, is in serious decline. Breton identity no longer has the support of widespread use of the language.

The various relationships among languages and dialects discussed above can be used to show how the concepts of 'power' and 'solidarity' help us understand what is happening. Power requires some kind of asymmetrical relationship between entities: one has more of something that is important, e.g. status, money, influence, etc., than the other or others. A language has more power than any of its dialects. It is the powerful dialect but it has become so because of non-linguistic factors. Standard English and Parisian French are good examples. Solidarity, on the other hand, is a feeling of equality that people have with one

another. They have a common interest around which they will bond. A feeling of solidarity can lead people to preserve a local dialect or an endangered language to resist power, or to insist on independence. It accounts for the persistence of local dialects, the modernization of Hebrew, and the separation of Serbo-Croatian into Serbian and Croatian.

The language–dialect situation along the border between the Netherlands and Germany is an interesting one. Historically, there was a continuum of dialects of one language, but the two that eventually became standardized as the languages of the Netherlands and Germany, Standard Dutch and Standard German, are not *mutually intelligible*, that is, a speaker of one cannot understand a speaker of the other. In the border area speakers of the local varieties of Dutch and German still exist within that *dialect continuum* (see p. 45) and remain largely intelligible to one another, yet the people on one side of the border say they speak a variety of Dutch and those on the other side say they speak a variety of German. The residents of the Netherlands look to Standard Dutch for their model; they read and write Dutch, are educated in Dutch, and watch television in Dutch. Consequently, they say they use a local variety, or dialect, of Dutch in their daily lives. On the other side of the border, German replaces Dutch in all equivalent situations. The interesting linguistic fact, though, is that there are more similarities between the local varieties spoken on each side of the border than between the one dialect (of Dutch?) and Standard Dutch and the other dialect (of German?) and Standard German, and more certainly than between that dialect and the south German, Swiss, and Austrian dialects of German. However, it is also of interest to note (Kremer, 1999) that younger speakers of Dutch in this area of the Netherlands are more conscious of the standard language border than older speakers. Apparently, their Dutch identity triumphs over any linguistic connections they have with speakers of the same dialect over the national border.

Gumperz has suggested some of the confusions that result from popular uses of the terms *language* and *dialect*. To these we can add the situation in Scandinavia as further evidence. Danish, Norwegian (actually two varieties), and Swedish are recognized as different languages, yet if you speak any one of them you will experience little difficulty in communicating while traveling in Scandinavia (excluding, of course, Finland, or at least the non-Swedish-speaking parts of that country). Danish and Norwegian share much vocabulary but differ considerably in pronunciation. In contrast, there are considerable vocabulary differences between Swedish and Norwegian but they are similar in pronunciation. Both Danes and Swedes claim good understanding of Norwegian. However, Danes claim to comprehend Norwegians much better than Norwegians claim to comprehend Danes. The poorest mutual comprehension is between Danes and Swedes and the best is between Norwegians and Swedes. These differences in mutual intelligibility appear to reflect power relationships: Denmark long dominated Norway, and Sweden is today the most influential country in the region and Denmark the least powerful.

A somewhat similar situation exists in the relationship of Thai and Lao. The Laos understand spoken Thai and hear Thai constantly on radio and television. Educated Laos can also read written Thai. However, Thais do not readily

understand spoken Lao nor do they read the written variety. Lao is a low-prestige language so far as Thais are concerned; in contrast, Thai has high prestige in Laos. Thais, therefore, are unwilling to expend effort to understand Lao, whereas Laos are willing to make the extra effort to understand Thai.

If we turn our attention to China, we will find that speakers of Cantonese and Mandarin will tell you that they use the same language. However, if one speaker knows only Cantonese and the other only Mandarin, they will not be able to converse with each other: they actually speak different languages, certainly as different as German and Dutch and even Portuguese and Italian. If the speakers are literate, however, they will be able to communicate with each other through a shared writing system. They will almost certainly insist that they speak different *dialects* of Chinese, not different *languages*, for to the Chinese a shared writing system and a strong tradition of political, social, and cultural unity form essential parts of their definition of *language*.

The situation can become even more confused. A speaker of Cockney, a highly restricted London variety of English, may find it difficult to communicate with natives of the Ozark Mountains in the United States. Do they therefore speak separate languages? Is there one English language spoken in Britain and another, American, spoken in the New World? The American writer Mencken (1919) had very definite views that the varieties spoken on the two sides of the Atlantic were sufficiently distinctive to warrant different appellations. It is also not unusual to find French translations of American books described on their title pages as translations from 'American' rather than 'English.' Is there a bona fide separate Scottish variety of English? There was before the crowns and parliaments were united several centuries ago. However, today there is no clear answer to that question as the power relationship between England and Scotland fluctuates and the issue of language differences is but one of many that must be dealt with. Is the French of Quebec a dialect of Standard (continental) French, or should we regard it as a separate language, particularly after a political separation of well over two centuries? Is Haitian Creole (see p. 84) a variety of French, or is it an entirely separate language, and if so in what ways is it separate and different? How do the different varieties of English spoken in Jamaica (see p. 81) relate to other varieties of English? Or is that question really answerable? What, above all, is English? How can we define it as something apart from what Speaker A uses, or Speaker B, or Speaker C? If it is something A, B, and C share, just what is it that they do share?

We undoubtedly agree that this book is written in English and that English is a language, but we may be less certain that various other things we see written or hear spoken in what is called *English* should properly be regarded as English rather than as dialects or varieties of English, perhaps variously described as Indian English, Australian English, New York English, West Country English, African American Vernacular English, nonstandard English, BBC English, and so on. A language then would be some unitary system of linguistic communication which subsumes a number of mutually intelligible varieties. It would therefore be bigger than a single dialect or a single variety. However, that cannot always be the case, for some such systems used by very small numbers of speakers may have very little internal variation. Yet each must be a language,

for it is quite unlike any other existing system. Actually, neither the requirement that there be internal variation nor the ‘numbers game,’ i.e., that a language must somehow be ‘bigger’ than a dialect, offers much help. Many languages have only a handful of speakers; several have actually been known to have had only a single remaining speaker at a particular point in time and the language has ‘died’ with that speaker.

Still another difficulty arises from the fact that the terms *language* and *dialect* are also used in an historical sense. It is possible to speak of languages such as English, German, French, Russian, and Hindi as Indo-European dialects. In this case the assumption is that there was once a single language, Indo-European, that the speakers of that language (which may have had various dialects) spread to different parts of the world, and that the original language eventually diverged into the various languages we subsume today under the *Indo-European family* of languages. However, we should also be aware that this process of divergence was not as clean-cut as this classical *neo-grammarian* model of language differentiation suggests. (In such a model all breaks are clean, and once two varieties diverge they lose contact with each other.) Processes of convergence must also have occurred, even of convergence among entirely unrelated languages (that is, languages without any ‘family’ resemblance). For example, Indo-European and Dravidian languages have influenced each other in southern India and Sri Lanka, and in the Balkans there is considerable evidence of the spread of common features across languages such as Albanian, Greek, Turkish, and several Slavic languages. In such situations, language and dialect differences become further obscured, particularly when many speakers are also likely to be multilingual.

Perhaps some of the difficulties we have with trying to define the term *language* arise from trying to subsume various different types of systems of communication under that one label. An alternative approach might be to acknowledge that there are different kinds of languages and attempt to discover how languages can differ from one another yet still be entities that most of us would want to call languages rather than dialects. It might then be possible to define a dialect as some sub-variety of one or more of these entities.

One such attempt (see Bell, 1976, pp. 147–57) has listed seven criteria that may be useful in discussing different kinds of languages. According to Bell, these criteria (standardization, vitality, historicity, autonomy, reduction, mixture, and *de facto* norms) may be used to distinguish certain languages from others. They also make it possible to speak of some languages as being more ‘developed’ in certain ways than others, thus addressing a key issue in the language–dialect distinction, since speakers usually feel that languages are generally ‘better’ than dialects in some sense.

Standardization refers to the process by which a language has been codified in some way. That process usually involves the development of such things as grammars, spelling books, and dictionaries, and possibly a literature. We can often associate specific items or events with standardization, e.g., Wycliffe’s and Luther’s translations of the Bible into English and German, respectively, Caxton’s establishment of printing in England, and Dr Johnson’s dictionary of English published in 1755. Standardization also requires that a measure of agreement be achieved about what is in the language and what is not. Once we have such

a codification of the language we tend to see it as almost inevitable, the result of some process come to fruition, one that has also reached a fixed end point. Change, therefore, should be resisted since it can only undo what has been done so laboriously. Milroy (2001, p. 537) characterizes the resulting ideology as follows: 'The canonical form of the language is a precious inheritance that has been built up over the generations, not by the millions of native speakers, but by a select few who have lavished loving care upon it, polishing, refining, and enriching it until it has become a fine instrument of expression (often these are thought to be literary figures, such as Shakespeare). This is a view held by people in many walks of life, including plumbers, politicians and professors of literature. It is believed that if the canonical variety is not universally supported and protected, the language will inevitably decline and decay.'

Once a language is standardized it becomes possible to teach it in a deliberate manner. It takes on ideological dimensions – social, cultural, and sometimes political – beyond the purely linguistic ones. In Fairclough's words (2001, p. 47) it becomes 'part of a much wider process of economic, political and cultural unification . . . of great . . . importance in the establishment of nationhood, and the nation-state is the favoured form of capitalism.' According to these criteria, both English and French are quite obviously standardized, Italian somewhat less so, and the variety known as African American Vernacular English (see chapter 14) not at all.

Haugen (1966a) has indicated certain steps that must be followed if one variety of a language is to become the standard for that language. In addition to what he calls the 'formal' matters of codification and elaboration, the former referring to the development of such things as grammars and dictionaries and the latter referring to the use of the standard in such areas as literature, the courts, education, administration, and commerce, Haugen says there are important matters to do with 'function.' For example, a norm must be selected and accepted because neither codification nor elaboration is likely to proceed very far if the community cannot agree on some kind of model to act as a norm. That norm is also likely to be – or to become – an idealized norm, one that users of the language are asked to aspire to rather than one that actually accords with their observed behavior.

Selection of the norm may prove difficult because choosing one vernacular as a norm means favoring those who speak that variety. It also diminishes all the other varieties and possible competing norms, and those who use those varieties. The chosen norm inevitably becomes associated with power and the rejected alternatives with lack of power. Not surprisingly, it usually happens that a variety associated with an elite is chosen. Attitudes are all-important, however. A group that feels intense solidarity may be willing to overcome great linguistic differences in establishing a norm, whereas one that does not have this feeling may be unable to overcome relatively small differences and be unable to agree on a single variety and norm. Serbs and Croats were never able to agree on a norm, particularly as other differences reinforced linguistic ones. In contrast, we can see how Indonesia and Malaysia are looking for ways to reduce the differences between their languages, with their common Islamic bond a strong incentive.

The standardization process itself performs a variety of functions (Mathiot and Garvin, 1975). It unifies individuals and groups within a larger community

while at the same time separating the community that results from other communities. Therefore, it can be employed to reflect and symbolize some kind of identity: regional, social, ethnic, or religious. A standardized variety can also be used to give prestige to speakers, marking off those who employ it from those who do not, i.e., those who continue to speak a nonstandard variety. It can therefore serve as a kind of goal for those who have somewhat different norms; Standard English and Standard French are such goals for many whose norms are dialects of these languages. However, as we will see (particularly in chapters 6–8), these goals are not always pursued and may even be resisted.

It still may not be at all easy for us to define *Standard English* because of a failure to agree about the norm or norms that should apply. For example, Trudgill (1995, pp. 5–6) defines Standard English as follows (note his use of ‘usually’ and ‘normally’ in this definition):

Standard English is that variety of English which is usually used in print, and which is normally taught in schools and to non-native speakers learning the language. It is also the variety which is normally spoken by educated people and used in news broadcasts and other similar situations. The difference between standard and nonstandard, it should be noted, has nothing in principle to do with differences between formal and colloquial language, or with concepts such as ‘bad language.’ Standard English has colloquial as well as formal variants, and Standard English speakers swear as much as others.

Historically, the standard variety of English is based on the dialect of English that developed after the Norman Conquest resulted in the permanent removal of the Court from Winchester to London. This dialect became the one preferred by the educated, and it was developed and promoted as a model, or norm, for wider and wider segments of society. It was also the norm that was carried overseas, but not one unaffected by such export. Today, Standard English is codified to the extent that the grammar and vocabulary of English are much the same everywhere in the world: variation among local standards is really quite minor, being differences of ‘flavor’ rather than of ‘substance,’ so that the Singapore, South African, and Irish varieties are really very little different from one another so far as grammar and vocabulary are concerned. Indeed, Standard English is so powerful that it exerts a tremendous pressure on all local varieties, to the extent that many of the long-established dialects of England and the Lowlands English of Scotland have lost much of their vigor. There is considerable pressure on them to converge toward the standard. This latter situation is not unique to English: it is also true in other countries in which processes of standardization are under way. It does, however, sometimes create problems for speakers who try to strike some kind of compromise between local norms and national, even supranational, ones.

Governments sometimes very deliberately involve themselves in the standardization process by establishing official bodies of one kind or another to regulate language matters or to encourage changes felt to be desirable. One of the most famous examples of an official body established to promote the language of a country was Richelieu’s establishment of the Académie Française in 1635. Founded

at a time when a variety of languages existed in France, when literacy was confined to a very few, and when there was little national consciousness, the Académie Française faced an unenviable task: the codification of French spelling, vocabulary, and grammar. Its goal was to fashion and reinforce French nationality, a most important task considering that, even two centuries later in the early nineteenth century, the French of Paris was virtually unknown in many parts of the country, particularly in the south. Similar attempts to found academies in England and the United States for the same purpose met with no success, individual dictionary-makers and grammar-writers having performed much the same function for English. Since both French and English are today highly standardized, one might question whether such academies serve a useful purpose, yet it is difficult to imagine France without the Académie Française: it undoubtedly has had a considerable influence on the French people and perhaps on their language.

Standardization is sometimes deliberately undertaken quite rapidly for political reasons. In the nineteenth century Finns developed their spoken language to make it serve a complete set of functions. They needed a standardized language to assert their independence from both Swedes and Russians. They succeeded in their task so that now the Finnish language has become a strong force in the nation's political life and a strong marker of Finnish identity among Germanic tongues on the one side and Slavic tongues on the other. In the twentieth century the Turks under Atatürk were likewise successful in their attempt to both standardize and 'modernize' Turkish. Today, we can see similar attempts at rapid standardization in countries such as India (Hindi), Israel (Hebrew), Papua New Guinea (Tok Pisin), Indonesia (Bahasa Indonesia), and Tanzania (Swahili). In each case a language or a variety of a language had to be selected, developed in its resources and functions, and finally accepted by the larger society. As we have seen, standardization is an ideological matter. Williams (1992, p. 146) calls it 'a sociopolitical process involving the legitimisation and institutionalisation of a language variety as a feature of sanctioning of that variety as socially preferable.' It creates a preferred variety of a language, which then becomes the winner in a struggle for dominance. The dispreferred varieties are losers.

The standardization process occasionally results in some languages actually achieving more than one standardized variety. Norwegian is a good example with its two standards, Nynorsk and Bokmål. In this case there is a special problem, that of trying to unify the two varieties in a way that pleases everyone. Some kind of unification or amalgamation is now official government policy (see pp. 373–4). Countries with two or more competing languages that cannot possibly be unified may tear themselves apart, as we saw in Yugoslavia, or periodically seem to come very close to doing that, as with Belgium and Canada (see chapter 15).

Standardization is also an ongoing matter, for only 'dead' languages like Latin and Classical Greek are standardized for all time. Living languages change and the standardization process is necessarily an ongoing one. It is also one that may be described as more advanced in languages like French or German and less advanced in languages like Bahasa Indonesia and Swahili.

Hindi is still in the process of being standardized in India. That process is hindered by widespread regional resistance to Hindi out of the fear that regional

languages may be submerged or, if not submerged, quite diminished. So far as standardization is concerned, there are problems with accepting local varieties, and with developing and teaching the existing standard as though it were a classical language like Sanskrit and downplaying it as a living language. Hindi is still often taught much like Latin in schools in the West; it is in many places an underused second language at best; children are not encouraged 'to play in Hindi,' and teachers rarely employ Hindi as a language of instruction. Likewise, the kinds of literature available in Hindi are still very limited, there being shortages of everyday reading materials that might appeal to the young, e.g., comic books, mystery stories, and collections of folk tales. Consequently, the process of the standardization of a 'living' Hindi is proving to be a slow one.

The standardization process is also obviously one that attempts either to reduce or to eliminate diversity and variety. However, there may well be a sense in which such diversity and variety are 'natural' to all languages, assuring them of their vitality and enabling them to change (see chapter 8). To that extent, standardization imposes a strain on languages or, if not on the languages themselves, on those who take on the task of standardization. That may be one of the reasons why various national academies have had so many difficulties in their work: they are essentially in a no-win situation, always trying to 'fix' the consequences of changes that they cannot prevent, and continually being compelled to issue new pronouncements on linguistic matters. Unfortunately, those who think you can standardize and 'fix' a language for all time are often quite influential. They often find ready access to the media, there to bewail the fact that English, for example, is becoming 'degenerate' and 'corrupt,' and to advise us to return to what they regard as a more perfect past. They may also resist what they consider to be 'dangerous' innovations, e.g., the translation of a sacred book into a modern idiom or the issue of a new dictionary. Since the existence of internal variation is one aspect of language and the fact that all languages keep changing is another, we cannot be too sympathetic to such views.

Vitality, the second of Bell's seven criteria, refers to the existence of a living community of speakers. This criterion can be used to distinguish languages that are 'alive' from those that are 'dead.' Two Celtic languages of the United Kingdom are now dead: Manx, the old language of the Isle of Man, and Cornish. Manx died out after World War II, and Cornish disappeared at the end of the eighteenth century, one date often cited being 1777, when the last known speaker, Dorothy Pentreath of Mousehole, died. Many of the aboriginal languages of the Americas are also dead. Latin is dead in this sense too for no one speaks it as a native language; it exists only in a written form frozen in time, pronounced rather than spoken, and studied rather than used.

Once a language dies it is gone for all time and not even the so-called revival of Hebrew contradicts that assertion. Hebrew always existed in a spoken form as a liturgical language, as did Latin for centuries. Modern Hebrew is an outgrowth of this liturgical variety. It is after all 'Modern' Hebrew and the necessary secularization of a liturgical language to make it serve the purposes of modern life has not been an easy and uncontroversial matter. Many languages, while not dead yet, nevertheless are palpably dying: the number of people who speak them diminishes drastically each year and the process seems irreversible, so that the

best one can say of their vitality is that it is flagging. For example, the French dialects spoken in the Channel Islands of Jersey, Guernsey, and Sark are rapidly on their way to extinction. Each year that passes brings a decrease in the number of languages spoken in the world (see pp. 378–9).

We should note that a language can remain a considerable force even after it is dead, that is, even after it is no longer spoken as anyone's first language and exists almost exclusively in one or more written forms, knowledge of which is acquired only through formal education. Classical Greek and Latin still have considerable prestige in the Western world, and speakers of many modern languages continue to draw on them in a variety of ways. Sanskrit is important in the same way to speakers of Hindi; Classical Arabic provides a unifying force and set of resources in the Islamic world; and Classical Chinese has considerably influenced not only modern Chinese but also Japanese and Korean. Such influences cannot be ignored, because the speakers of languages subject to such influences are generally quite aware of what is happening: we can even say that such influence is part of their knowledge of the language. We can also periodically observe deliberate attempts to throw off an influence perceived to be alien: for example, Atatürk's largely successful attempt to reduce the Arabic influence on Turkish, and periodic attempts to 'purify' languages such as French and German of borrowings from English. While in the case of Hebrew, a language used only in a very restricted way for religious observances was successfully expanded for everyday use, we should note that a similar attempt to revitalize Gaelic in Ireland has been almost a complete failure.

Historicity refers to the fact that a particular group of people finds a sense of identity through using a particular language: it belongs to them. Social, political, religious, or ethnic ties may also be important for the group, but the bond provided by a common language may prove to be the strongest tie of all. In the nineteenth century a German nation was unified around the German language just as in the previous century Russians had unified around a revitalized Russian language. Historicity can be long-standing: speakers of the different varieties of colloquial Arabic make much of a common linguistic ancestry, as obviously do speakers of Chinese. It can also, as with Hebrew, be appealed to as a unifying force among a threatened people.

Autonomy is an interesting concept because it is really one of feeling. A language must be felt by its speakers to be different from other languages. However, this is a very subjective criterion. Ukrainians say their language is quite different from Russian and deplored its Russification when they were part of the Soviet Union. Some speakers of African American Vernacular English (see chapter 14) maintain that their language is not a variety of English but is a separate language in its own right and refer to it as *Ebonics*. In contrast, speakers of Cantonese and Mandarin deny that they speak different languages: they maintain that Cantonese and Mandarin are not autonomous languages but are just two dialects of Chinese. As we will see (chapter 3), creole and pidgin languages cause us not a few problems when we try to apply this criterion: how autonomous are such languages?

Reduction refers to the fact that a particular variety may be regarded as a sub-variety rather than as an independent entity. Speakers of Cockney will almost certainly say that they speak a variety of English, admit that they are not

representative speakers of English, and recognize the existence of other varieties with equivalent subordinate status. Sometimes the reduction is in the kinds of opportunities afforded to users of the variety. For example, there may be a reduction of resources; that is, the variety may lack a writing system. Or there may be considerable restrictions in use; e.g., pidgin languages are very much reduced in the functions they serve in society in contrast to standardized languages.

Mixture refers to feelings speakers have about the 'purity' of the variety they speak. This criterion appears to be more important to speakers of some languages than of others, e.g., more important to speakers of French and German than to speakers of English. However, it partly explains why speakers of pidgins and creoles have difficulty in classifying what they speak as full languages: these varieties are, in certain respects, quite obviously 'mixed,' and the people who speak them often feel that the varieties are neither one thing nor another, but rather are debased, deficient, degenerate, or marginal varieties of some other standard language.

Finally, having *de facto norms* refers to the feeling that many speakers have that there are both 'good' speakers and 'poor' speakers and that the good speakers represent the norms of proper usage. Sometimes this means focusing on one particular sub-variety as representing the 'best' usage, e.g., Parisian French or the Florentine variety of Italian. Standards must not only be established (by the first criterion above), they must also be observed. When all the speakers of a language feel that it is badly spoken or badly written almost everywhere, that language may have considerable difficulty in surviving; in fact, such a feeling is often associated with a language that is dying. Concern with the norms of linguistic behavior, 'linguistic purism' (see Thomas, 1991), may become very important among specific segments of society. For example, so far as English is concerned, there is a very profitable industry devoted to telling people how they should behave linguistically, what it is 'correct' to say, what to avoid saying, and so on (see Baron, 1982, Cameron, 1995, and Wardhaugh, 1999). As we will see (chapters 7–8), people's feelings about norms have important consequences for an understanding of both variation and change in language.

If we apply the above criteria to the different varieties of speech we observe in the world, we will see that not every variety we may want to call a language has the same status as every other variety. English is a language, but so are Dogrib, Haitian Creole, Ukrainian, Latin, Tok Pisin, and Chinese. Each satisfies a different sub-set of criteria from our list. Although there are important differences among them, we would be loath to deny that any one of them is a language. They are all equals as languages, but that does not necessarily mean that all languages are equal! The first is a linguistic judgment, the second a social one.

As we have just seen, trying to decide whether something is or is not a language or in what ways languages are alike and different can be quite troublesome. However, we usually experience fewer problems of the same kind with regard to dialects. There is usually little controversy over the fact that they are either regional or social varieties of something that is widely acknowledged to be a language. That is true even of the relationship of Cantonese and Mandarin to Chinese if the latter is given a 'generous' interpretation as a language.

Some people are also aware that the standard variety of any language is actually only the preferred dialect of that language: Parisian French, Florentine

Italian, or the Zanzibar variety of Swahili in Tanzania. It is the variety that has been chosen for some reason, perhaps political, social, religious, or economic, or some combination of reasons, to serve as either the model or norm for other varieties. It is the empowered variety. As a result, the standard is often not called a dialect at all, but is regarded as the language itself. It takes on an ideological dimension and becomes the 'right' and 'proper' language of the group of people, the very expression of their being. One consequence is that all other varieties become related to that standard and are regarded as dialects of that standard with none of the power of that standard. Of course, this process usually involves a complete restructuring of the historical facts. If language X^1 differentiates in three areas to become dialects XA, XB, and XC, and then XA is elevated to become a later standard X^2 , then XB, and XC are really historical variants of X^1 , not sub-varieties of X^2 . What happens in practice is that XB and XC undergo pressure to change toward X^2 , and X^2 , the preferred variety or standard, exerts its influence over the other varieties.

We see a good instance of this process in Modern English. The new standard is based on the dialect of the area surrounding London, which was just one of several dialects of Old English, and not the most important for both the western and northern dialects were once at least equally as important. However, in the modern period, having provided the base for Standard English, this dialect exerts a strong influence over all the other dialects of England so that it is not just first among equals but rather represents the modern language itself to the extent that the varieties spoken in the west and north are generally regarded as its local variants. Historically, these varieties arise from different sources, but now they are viewed only in relation to the standardized variety.

A final comment seems called for with regard to the terms *language* and *dialect*. A dialect is a subordinate variety of a language, so that we can say that Texas English and Swiss German are, respectively, dialects of English and German. The language name (i.e., *English* or *German*) is the superordinate term. We can also say of some languages that they contain more than one dialect; e.g., English, French, and Italian are spoken in various dialects. If a language is spoken by so few people, or so uniformly, that it has only one variety, we might be tempted to say that *language* and *dialect* become synonymous in such a case. However, another view is that it is inappropriate to use *dialect* in such a situation because the requirement of subordination is not met. Consequently, to say that we have dialect A of language X must imply also the existence of dialect B of language X, but to say we have language Y is to make no claim about the number of dialect varieties in which it exists: it may exist in only a single variety, or it may have two (or more) subordinate dialects: dialects A, B, and so on.

Finally, two other terms are important in connection with some of the issues discussed above: *vernacular* and *koiné*. Petyt (1980, p. 25) defines the former as 'the speech of a particular country or region,' or, more technically, 'a form of speech transmitted from parent to child as a primary medium of communication.' If that form of speech is Standard English, then Standard English is the vernacular for that particular child; if it is a regional dialect, then that dialect is the child's vernacular. A *koiné* is 'a form of speech shared by people of different vernaculars – though for some of them the *koiné* itself may be their

vernacular.’ A koiné is a common language, but not necessarily a standard one. Petyt’s examples of koinés are Hindi for many people in India and Vulgar Latin (*vulgar*: ‘colloquial’ or ‘spoken’) in the Roman Empire. The original koiné was, of course, the Greek koiné of the Ancient World, a unified version of the Greek dialects, which after Alexander’s conquests (*circa* 330 BCE) became the *lingua franca* of the Western world, a position it held until it was eventually superseded, not without a struggle, by Vulgar Latin.

Discussion

1. A survey of the following kind might prove quite revealing. Ask a variety of people you know questions such as these, and then try to organize their responses in a systematic way:
 - a. Which language(s) do you speak?
 - b. Do you speak a dialect of X?
 - c. Where is the best X spoken?
 - d. What is your native language (or mother tongue)?
 - e. Do you speak X with an accent? If so, what accent?

Try also to get definitions from your informants for each of the terms that you use.

2. A question found on many national census forms concerns the language or languages spoken (or known). It may ask respondents either to check one or more language names or to volunteer a name or names. What problems do you see in collecting data in such a way? Think of countries like China, the United States, Canada, India, France, Spain, and Norway.
3. Is Afrikaans a dialect of Dutch or a different language? To attempt an answer to this question you will have to consider a variety of issues: What is the origin of Afrikaans? Are Afrikaans and Dutch mutually intelligible? How different are the orthographies (i.e., systems of spelling), sounds, vocabularies, and grammars? How important is the factor of the national consciousness of those who speak Afrikaans? Is the initial question clearly answerable from the kinds of theories and data that are currently available to us?
4. Speakers of Faroese are said to understand speakers of Icelandic but not vice versa. Danes seem to understand Norwegians better than Norwegians understand Danes. Monolingual speakers of Mandarin and Cantonese cannot communicate with each other in speech. What do such facts have to say about using the criterion of mutual intelligibility in deciding whether we are dealing with a single language, with two dialects of one language, or with two separate languages? Consider the following pieces of evidence in arriving at your answer. Speakers of Isoko in Nigeria say they cannot understand those who speak other Urhobo languages/dialects; but these others apparently understand them. This situation seems to have developed concurrently with demands for greater political autonomy and ethnic self-sufficiency.

5. Standard languages are usually based on an existing dialect of the language. For example, the British variety of English is based, historically at least, on the dialect of the area surrounding London, Continental French on the dialect of Paris, and Italian on the dialect of Florence or Tuscany (although Rome and Milan became important influences in the late twentieth century). In other countries the situation is not so clear-cut. What can you find out about the difficulties of choosing a variety for standardization in Denmark, Indonesia, Greece, China, Haiti, and the Arab world?
6. Old English, the language spoken a thousand years ago in England, was a west-country variety of English, West Saxon. The court was located at Winchester and the literature and documents of the period were written in West Saxon (or sometimes in Latin). By 1400 the English court was well established in London, which became the center of social, political, and economic power. It also became the literary center of the country, particularly after the development of printing. The variety of English spoken in and around London, including Oxford and Cambridge (which were important intellectual centers), became predominant. How would you use facts such as these to argue that no variety of a language is intrinsically better than another and that what happens to a language is largely the result of the chance interplay of external forces? Can you think of other examples which might support such a conclusion?
7. Mencken wrote a series of books under the general title *The American Language*. Why did he choose this particular title? Why not *The English Language in America*? If the English of the United States is properly regarded as a separate language, how about the varieties found in Canada, Australia, South Africa, and Singapore? You might read Lilles (2000) for a strongly expressed dismissal of 'Canadian English,' as a 'fiction [without] any value linguistically, pragmatically, socially, or politically' (p. 9). (See Clyne, 1992, for a discussion of what he calls 'pluricentric languages'.)
8. One of the goals Dr Johnson set himself in compiling his *Dictionary* of 1755 was to 'fix,' i.e., standardize, English. What does Johnson say in the Preface to that dictionary about his success in meeting that goal?
9. The publication in 1961 of *Webster's Third New International Dictionary* caused a tremendous stir in North America, being regarded by many critics as an attack on prevailing language standards. What were the issues? (See Sledd and Ebbitt, 1962, Finegan, 1980, and Wardhaugh, 1999.)
10. Writing of the codification of Standard English, Leith (1997, pp. 56–7) says that 'by analyzing "correct" usage in terms that only a tiny minority of educated people could command, the codifiers ensured that correctness remained the preserve of an elite. The usage of most people was wrong, precisely because it was the usage of the majority.' There appear to be both advantages and disadvantages to having a 'standard language.' Is it possible to make an objective assessment of these? Or is any judgment inherently ideological?
11. If Scotland continues to devolve from England, what might this mean for the variety of English spoken there? How might Scots become unequivocally a distinctive variety of English?

12. Arabs have a particular historical view of Arabic and Turks of Turkish. Try to find out something about these views. How do they help Arabs and Turks to maintain their languages? Hindi and Urdu are now viewed as rather different by those who speak these languages. How is each language being reshaped to conform to these views?
13. How would you evaluate each of the following languages according to the criteria stated above (standardization, vitality, historicity, autonomy, reduction, mixture, and *de facto* norms); that is, for each criterion, does the language possess the stated characteristic or lack it: Haitian Creole, Provençal, Singapore English, Old English, Pitcairnese, African American Vernacular English, Tok Pisin, Cockney, Ukrainian, and the language of Shakespeare's plays?
14. Find out what you can about Basic English. In what ways is it a reduced form of Standard English? Do the kinds of reductions introduced into Basic English make it 'simpler' to learn and use? (You will have to define 'simpler'.)
15. From time to time certain users of languages such as French and German have objected to borrowings, in particular borrowings from English. What Anglicisms have been objected to? What kinds of native resources have been suggested as suitable alternative sources of exploitation in order to develop and/or purify the language? What motivates the objections?
16. Some Chinese scholars are concerned with developing the vocabulary of Chinese to make it usable for every kind of scientific and technical endeavor. They reject the idea that such vocabulary should be borrowed from other languages. What do you think they hope to gain by doing this? Do they lose anything if they are successful?
17. 'A language is a dialect with an army and a navy' is a well-known observation. (Today we would add an 'airforce'!) True? And, if so, what are the consequences?
18. In the *UNESCO Courier* of April, 2000, a writer makes the following observation: 'Languages usually have a relatively short life span as well as a very high death rate. Only a few, including Basque, Egyptian, Chinese, Greek, Hebrew, Latin, Persian, Sanskrit, and Tamil have lasted more than 2000 years.' How is this statement at best a half-truth?
19. Are the Australian, New Zealand, Canadian, and other national varieties of English 'new dialects' of English, or autonomous languages, or possibly even both? (See Hickey, 2004, Gordon et al., 2004, and Trudgill, 2004.)

Regional Dialects

Regional variation in the way a language is spoken is likely to provide one of the easiest ways of observing variety in language. As you travel throughout a wide geographical area in which a language is spoken, and particularly if that language has been spoken in that area for many hundreds of years, you are almost certain to notice differences in pronunciation, in the choices and forms

of words, and in syntax. There may even be very distinctive local colorings in the language which you notice as you move from one location to another. Such distinctive varieties are usually called *regional dialects* of the language. As we saw earlier (p. 28), the term *dialect* is sometimes used only if there is a strong tradition of writing in the local variety. Old English and to a lesser extent Middle English had dialects in this sense. In the absence of such a tradition of writing the term *patois* may be used to describe the variety. However, many linguists writing in English tend to use *dialect* to describe both situations and rarely, if at all, use *patois* as a scientific term. You are likely to encounter it only as a kind of anachronism, as in its use by Jamaicans, who often refer to the variety of English spoken on the island as a 'patois.'

The *dialect-patois* distinction actually seems to make more sense in some situations, e.g., France, than in others. In medieval France, a number of languages flourished and several were associated with strong literary traditions. However, as the language of Paris asserted itself from the fourteenth century on, these traditions withered. Parisian French spread throughout France, and, even though that spread is still not yet complete (as visits to such parts of France as Brittany, Provence, Corsica, and Alsace will confirm), it drastically reduced the importance of the local varieties: they continue to exist largely in spoken forms only; they are disfavored socially and politically; they are merely *patois* to those who extol the virtues of Standard French. However, even as these varieties have faded, there have been countervailing moves to revive them as many younger residents of the areas in which they are spoken see them as strong indicators of identities they wish to preserve.

There are some further interesting differences in the use of the terms *dialect* and *patois* (Petyt, 1980, pp. 24–5). *Patois* is usually used to describe only rural forms of speech; we may talk about an *urban dialect*, but to talk about an *urban patois* seems strange. *Patois* also seems to refer only to the speech of the lower strata in society; again, we may talk about a *middle-class dialect* but not, apparently, about a *middle-class patois*. Finally, a dialect usually has a wider geographical distribution than a *patois*, so that, whereas *regional dialect* and *village patois* seem unobjectionable, the same cannot be said for *regional patois* and *village dialect*. However, as I indicated above, many Jamaicans refer to the popular spoken variety of Jamaican English as a *patois* rather than as a dialect. So again the distinction is in no way an absolute one.

This use of the term *dialect* to differentiate among regional varieties of specific languages is perhaps more readily applicable to contemporary conditions in Europe and some other developed countries than it would have been in medieval or Renaissance Europe or today in certain other parts of the world, where it was (and still is) possible to travel long distances and, by making only small changes in speech from location to location, continue to communicate with the inhabitants. (You might have to travel somewhat slowly, however, because of the necessary learning that would be involved!) It has been said that at one time a person could travel from the south of Italy to the north of France in this manner. It is quite clear that such a person began the journey speaking one language and ended it speaking something entirely different; however, there was no one point at which the changeover occurred, nor is there actually any way of determining