popular press. Some scholars have proposed that several language families that are generally regarded as distinct should be viewed as having a common source further back in time. The languages hypothesized to belong to Nostratic differ slightly from scholar to scholar, but most scholars espousing this theory include Indo-European, Afroasiatic, Uralic, Altaic, Dravidian, and Eskimo-Aleut. Assuming that detailed comparative reconstruction confirmed this hypothesis, the Nostratic macrofamily would then make distant cousins of English (Indo-European); Hebrew, Arabic, Somali, and Hausa (Afroasiatic); Finnish and Hungarian (Uralic); perhaps Korean and Turkish (Altaic); Tamil (Dravidian); and Inuktitut (Eskimo-Aleut).

Although the links among these far-flung languages are not widely accepted among scholars, the hypothesis is provocative in an important way. As demonstrated in this chapter, the principal method for establishing genealogical relations among languages is by comparative reconstruction, whereby the forms of a parent language are hypothesized and the forms of the various daughter languages are derived by regular rules. Before any comparative reconstruction can be attempted, there must be hypotheses about which languages are and are not related. Without such hypotheses, just which languages would constitute the bases for establishing the sound correspondences that make the stuff of comparative reconstruction? With the Nostratic hypothesis in mind, you may find it thought provoking to reexamine the tables of common words for those Nostratic languages illustrated in this chapter: Tables 13-7 through 13-10 for four Indo-European groups, Table 13-12 for Afroasiatic, and Table 13-13 for Dravidian. Bear in mind that the sound correspondences among these languages would not be between the sounds of the daughter languages directly but between the sounds of the reconstructed parent languages, so any immediate correspondences that you might spy may be deceptive.

Languages in Contact

At no other time in history have there been such extensive contacts between language communities as in the last few centuries. As a result of the exploratory and colonizing enterprises of the English, French, Dutch, Spanish, and Portuguese, European languages have come into contact with languages of Africa, Native America, Asia, and the Pacific. These colonizing efforts put members of different speech communities in contact with each other. For example, the importing of slaves from Africa to the Americas forced speakers of different African languages to live side by side. Several language contact phenomena can take place when speakers of different languages interact.

Multilingualism

Bilingualism The first of these phenomena is bilingualism or multilingualism, in which members of a community acquire more than one language natively. In a multilingual community, children grow up speaking several languages. Use of each language is often compartmentalized, as when one is used at home and another at school or at work. Multilingualism is such a natural solution to the problem of language contact that it is extremely widespread throughout the world. In this respect, industrialized societies such as the United States and Japan, in which bilingualism is not widespread, are exceptional. In the United States, bilingualism is mostly relegated to immigrant communities, whose members are expected to learn English upon arrival. This adaptation is one-sided in contrast to what is found in most areas of the globe, where neighboring communities learn each other's languages with little ado. In central Africa, India, and Papua New Guinea, it is commonplace for small children to grow up speaking four or five languages. In Papua New Guinea, multilingualism is a highly valued attribute that enhances a person's status in the community.

Nativization A possible side effect of multilingualism is nativization, which takes place when a community adopts a new language (in addition to its native language) and modifies the structure of that new language, thus developing a dialect that becomes characteristic of the community. That is precisely what has happened with English in India, where Indian English is recognized as a separate dialect of English with some of its own structural characteristics. Indeed, it has become one of India's two national languages (along with Hindi, the most widely spoken indigenous language) and is used in education, government, and communications within India and with the rest of the world.

Pidgins Another process that may take place in language contact situations is pidginization. Although it is probably derived from the word business, the origin of the word pidgin is unclear, but the term refers to a contact language that develops where groups are in a dominant/subordinate situation, often in the context of colonization. Pidgins arise when members of a politically or economically dominant group do not learn the native language of the people they interact with as political or economic subordinates. To communicate, members of the subordinate community create a simplified variety of the language of the dominant group as their own second language. These simplified varieties then become the language of interaction between the colonizer and the colonized. Pidgins are thus defined in terms of sociological and linguistic characteristics. They are based on the language of the dominant group but are structurally simpler. They have no native speakers and are typically used for a restricted range of purposes.

Pidgins have arisen in many areas of the world, including West Africa, the Caribbean, the Far East, and the Pacific. Many pidgins have been based on English and French, the languages of the two most active colonial powers in the eighteenth and nineteenth centuries. Portuguese, Spanish, Dutch, Swedish, German, Arabic, and Russian, among others, have also served as a base for the development of pidgins. Today, most pidgins have given way to creole languages.

From Pidgin to Creole At some point, a pidgin may begin to fulfill a greater number of roles in social life. Instead of using the pidgin language only in the workplace to communicate with traders or colonizers, speakers may begin to use it at home or among themselves. Such situations frequently arise when the colonized population is linguistically diversified. Members of that community may find it convenient to adopt the new language as a lingua franca—a means to communicate across language boundaries. As a result, small children begin to grow up speaking the new language, and as greater demands are put onto that language its structure becomes more complex in a process called creolization. A creole language is thus a former pidgin that has "acquired" native speakers. Creoles are structurally complex, eventually as complex as any other language, and they differ from pidgins in that they exhibit less variability from speaker to speaker than pidgins do.

The boundary between pidgin and creole is often difficult to establish. Creolization is a gradual process, and in many places pidgins are undergoing creolization. In such situations,
there will be much variability from speaker to speaker and from situation to situation. For some speakers and in some contexts, the language will clearly be at the pidgin stage; for speakers whose language is more advanced in the creolization process, or in contexts that call for a more elaborated variety, the language will be structurally more complex. Furthermore, as a creole gains wider usage and becomes structurally more complex, it often comes to resemble the language on which it is based. For example, in the Caribbean and in Hawaii, English-based creoles are very similar to standard English for many speakers. Typically in such situations we find a continuum from speaker to speaker and from situation to situation—from a nonstandard dialect of the parent language to a very basic pidgin.

Figure 13-14 on page 455 shows the location of the more important creoles in the world. Note that in common parlance many creoles are called pidgins. Such is the case with Hawaiian Pidgin and Papua New Guinea Tok Pisin (from ‘talk Pidgin’), both of which are actually creoles.

Some creoles have low status where they are spoken. Hawaiian creole, or Da Kine Talk, is often referred to as a “bastardized” version of English or as “broken English.” The fact is that Hawaiian creole has its own structure, different from that of English, and you could not pretend to speak Da Kine Talk by speaking “broken” English.

In contrast, in many areas of the world creoles have become national languages used in government proceedings, education, and the media. In Papua New Guinea, Tok Pisin is one of the three national languages (along with English and Kiri Motu, also a creole) and has become a symbol of national identity. Some creoles have become the language of important bodies of literature, particularly in West Africa. Elsewhere, creoles are used in newspapers and on the radio for various purposes, including cartoons and commercials. Figure 13-15 on page 456 is a publicity cartoon in Papua New Guinea Tok Pisin; the English translation of the captions is given underneath. Tok Pisin is even used to write about linguistics, as illustrated by the following discussion of relative clause formation in Tok Pisin; it begins with three example sentences.

1. 01 ikilim pik bipo.
2. Na pik bai ikamap olosem draipela ston.
3. Na pik ia [01 ikilim bipo ia] bai ikamap olosem draipela ston.

Sapos yumi tingting gut long dispela tripela tok, yumi ken k1ia long tupela samting. Nambawan samting, sapos pik istap long (1) em inarapela pik, na pik istap long (2) em inarapela, orait, yumi no ken wokim (3). Tasol sapos wampela pik tasol istap long (1) na (2), em orait long wokim (3). Na tu, tingting istap long (1) ia, mi bin banisim insait long tupela banis long (3), long wowem, em bilong k1iaim yumi long wowem pik Elena itok en.

[Translation]
1. They killed the pig.
2. The pig looks like a big rock.
3. The pig [that they killed] looks like a big rock.

If we think carefully about these three sentences, we can obtain two interpretations. First, if the pig of sentence (1) is one pig, and the pig of sentence (2) is another pig, then we cannot construct (3). However, if the pig in (1) and (2) is the same, then we can construct (3). Thus, I have bracketed in (3) the meaning corresponding to (1) with two brackets, because it has the purpose of identifying for us...
The structural similarities among creoles worldwide are striking. Many creoles, for example, lack indefinite articles and a distinction between the future and other tenses. These similarities have led some researchers to propose that the development of pidgins and creoles follows a "program" that is genetically innate in humans. There are, however, many differences among the world's creoles, in which the imprint of various native languages is clear. In many South Pacific creoles, for example, a distinction is made in the pronoun system between dual and plural and between inclusive-first-person dual and plural and exclusive-first-person dual and plural (see Chapter 7, page 219 where the Tok Pisin pronoun system is given). These distinctions are not found in West African creoles, and their presence in South Pacific creoles reflects the fact that many languages spoken in the South Pacific make these distinctions. In Nigerian creole, on the other hand, we find honorific terms of address (Mom and Dad) that are used when addressing high-status individuals. These honorifics are not found in any other creole; again, they are transferred from local languages. Thus there is both homogeneity and heterogeneity among the creoles of the world.

In the study of historical linguistics and language change, computers have been particularly helpful in their ability to manipulate large quantities of data accurately and efficiently. Several major historical corpora have been compiled over the past couple of decades, and their ability to aid researchers in tracing lexical, morphological, semantic, and syntactic change in language has proven impressive and interesting.

Among the influential historical corpora is the Helsinki Corpus of English Texts: Diachronic and Dialectal (called the Helsinki Corpus for short). Here we concentrate on the historical (diachronic) part. Compiled by researchers at the University of Helsinki, this corpus contains texts of English from the Old English period (starting at about A.D. 800) and continuing through the early eighteenth century in the period known as Early Modern English. Unlike the LOB and Brown corpora, which contain 2000-word extracts of texts, the Helsinki Corpus contains texts varying in length from 2500 to almost 20,000 words. Altogether, there are 242 text files totaling about 1.5 million words of running text. Like many corpora, for each text the Helsinki Corpus includes information about the author's name, sex, education, origin, and social status, as well as information about the date of composition and the genre of the text (which is related to what we have been calling register). Using the Helsinki Corpus, researchers have been able to investigate patterns of development with certain genres across time, across genres within a given period of time, between male and female writers, and between British and American English, to mention just some of the dimensions along which it is possible to explore.

**Computers and the History of Languages**

In short, creoles can fulfill all the demands that are commonly imposed on a language. The structural similarities among creoles worldwide are striking. Many creoles, for example, lack indefinite articles and a distinction between the future and other tenses, and many have preposition stranding (as in the English expression the house I live in). Such similarities have led some researchers to propose that the development of pidgins and creoles follows a "program" that is genetically innate in humans. There are, however, many differences among the world's creoles, in which the imprint of various native

which pig Elena [the speaker who produced these sentences] is talking about. [Gillian Sankoff, "Sampela Nupela lo Ikamp Long Tok Pisin," 1975.]

In short, creoles can fulfill all the demands that are commonly imposed on a language. The structural similarities among creoles worldwide are striking. Many creoles, for example, lack indefinite articles and a distinction between the future and other tenses, and many have preposition stranding (as in the English expression the house I live in). Such similarities have led some researchers to propose that the development of pidgins and creoles follows a "program" that is genetically innate in humans. There are, however, many differences among the world's creoles, in which the imprint of various native

languages is clear. In many South Pacific creoles, for example, a distinction is made in the pronoun system between dual and plural and between inclusive-first-person dual and plural and exclusive-first-person dual and plural (see Chapter 7, page 219 where the Tok Pisin pronoun system is given). These distinctions are not found in West African creoles, and their presence in South Pacific creoles reflects the fact that many languages spoken in the South Pacific make these distinctions. In Nigerian creole, on the other hand, we find honorific terms of address (Mom and Dad) that are used when addressing high-status individuals. These honorifics are not found in any other creole; again, they are transferred from local languages. Thus there is both homogeneity and heterogeneity among the creoles of the world.

**Computers and the History of Languages**

In the study of historical linguistics and language change, computers have been particularly helpful in their ability to manipulate large quantities of data accurately and efficiently. Several major historical corpora have been compiled over the past couple of decades, and their ability to aid researchers in tracing lexical, morphological, semantic, and syntactic change in language has proven impressive and interesting.

Among the influential historical corpora is the Helsinki Corpus of English Texts: Diachronic and Dialectal (called the Helsinki Corpus for short). Here we concentrate on the historical (diachronic) part. Compiled by researchers at the University of Helsinki, this corpus contains texts of English from the Old English period (starting at about A.D. 800) and continuing through the early eighteenth century in the period known as Early Modern English. Unlike the LOB and Brown corpora, which contain 2000-word extracts of texts, the Helsinki Corpus contains texts varying in length from 2500 to almost 20,000 words. Altogether, there are 242 text files totaling about 1.5 million words of running text. Like many corpora, for each text the Helsinki Corpus includes information about the author's name, sex, education, origin, and social status, as well as information about the date of composition and the genre of the text (which is related to what we have been calling register). Using the Helsinki Corpus, researchers have been able to investigate patterns of development with certain genres across time, across genres within a given period of time, between male and female writers, and between British and American English, to mention just some of the dimensions along which it is possible to explore.

**Computers and the History of Languages**

In short, creoles can fulfill all the demands that are commonly imposed on a language. The structural similarities among creoles worldwide are striking. Many creoles, for example, lack indefinite articles and a distinction between the future and other tenses, and many have preposition stranding (as in the English expression the house I live in). Such similarities have led some researchers to propose that the development of pidgins and creoles follows a "program" that is genetically innate in humans. There are, however, many differences among the world's creoles, in which the imprint of various native