

**FIGURE 5-8** Soft-drink dialects. Soft drinks are called “soda” in the Northeast and Southwest, “pop” in the Midwest and Northwest, and “Coke” in the South. Map reflects voting at [www.popssoda.com](http://www.popssoda.com).

and reflects the place of origin of most New England colonists.

It also reflects the relatively high degree of contact between the two groups. Residents of Boston, New England’s main port city, maintained especially close ties to the important ports of southern England, such as London, Plymouth, and Bristol. Compared to other colonists, New Englanders received more exposure to changes in pronunciation that occurred in Britain during the eighteenth century.

The New England and southern accents sound unusual to the majority of Americans because the standard pronunciation throughout the American West comes from the Middle Atlantic states rather than the New England and Southern regions. This pattern occurred because most western settlers came from the Middle Atlantic states.

The diffusion of particular English dialects into the middle and western parts of the United States is a result of the westward movement of colonists from the three dialect regions of the East. The area of the Midwest south of the Ohio River was settled first by colonists from Virginia and the other southern areas. The Middle Atlantic colonies sent most of the early settlers north of the Ohio River, although some New Englanders moved to the Great Lakes area. The pattern by which dialects diffused westward resembles the diffusion of East Coast house types discussed in Chapter 4 (compare Figure 5-7 with Figure 4-12).

As more of the West was opened to settlement during the nineteenth century, people migrated from all parts of the East Coast. The California gold rush attracted people from throughout the East, many of whom subsequently moved to other parts of the West. The mobility of Americans has been a major reason for the relatively uniform language that exists throughout much of the West.

## KEY ISSUE 2

# Why Is English Related to Other Languages?

- Indo-European Branches
- Origin and Diffusion of Indo-European

English is part of the Indo-European language family. A **language family** is a collection of languages related through a common ancestral language that existed long before recorded history. Indo-European is the world’s most extensively spoken language family by a wide margin. ■

## Indo-European Branches

Within a language family, a **language branch** is a collection of languages related through a common ancestral language that existed several thousand years ago. Differences are not as extensive or as old as with language families, and archaeological evidence can confirm that the branches derived from the same family.

Indo-European is divided into eight branches (Figure 5-9). Four of the branches—Indo-Iranian, Romance, Germanic, and Balto-Slavic—are spoken by large numbers of people. Indo-Iranian languages are clustered in South Asia, Romance languages in southwestern Europe and Latin America, Germanic languages in northwestern Europe and North America, and Balto-Slavic languages in Eastern Europe. The four less extensively used Indo-European language branches are Albanian, Armenian, Greek, and Celtic.



**FIGURE 5-9** Branches of the Indo-European language family. Most Europeans speak languages from the Indo-European language family. In Europe, the three most important branches are Germanic (north and west), Romance (south and west), and Slavic (east). The fourth major branch, Indo-Iranian, clustered in southern and western Asia, has more than 1 billion speakers, the greatest number of any Indo-European branch.

## Germanic Branch of Indo-European

German may seem a difficult language for many English speakers to learn, but the two languages are actually closely related. Both belong to the Germanic language branch of Indo-European. English is part of the Germanic branch of the Indo-European family due to the language spoken by the Germanic tribes that invaded England 1,500 years ago.

A **language group** is a collection of languages within a branch that share a common origin in the relatively recent past and display relatively few differences in grammar and vocabulary. West Germanic is the group within the Germanic branch of Indo-European to which English belongs. Although they sound very different, English and German are both languages in the West Germanic group because they are structurally similar and have many words in common (Figure 5-10).

West Germanic is further divided into High Germanic and Low Germanic subgroups, so named because they are found in high and low elevations within present-day Germany. High German, spoken in the southern mountains of Germany, is the basis for the modern standard German language. English is classified in the Low Germanic subgroup of the West Germanic

group. Other Low Germanic languages include Dutch, which is spoken in the Netherlands, as well as Flemish, which is generally considered a dialect of Dutch spoken in northern Belgium. Afrikaans, a language of South Africa, is similar to Dutch, because Dutch settlers migrated to South Africa 300 years ago. Frisian is spoken by a few residents in northeastern Netherlands. A dialect of German spoken in the northern lowlands of Germany is also classified as Low Germanic.

The Germanic language branch also includes North Germanic languages, spoken in Scandinavia. The four Scandinavian languages—Swedish, Danish, Norwegian, and Icelandic—all derive from Old Norse, which was the principal language spoken throughout Scandinavia before A.D. 1000. Four distinct languages emerged after that time because of migration and the political organization of the region into four independent and isolated countries.

## Indo-Iranian Branch of Indo-European

The branch of the Indo-European language family with the most speakers is Indo-Iranian. This branch includes more than 100 individual languages. The branch is divided into an eastern group (Indic) and a western group (Iranian).



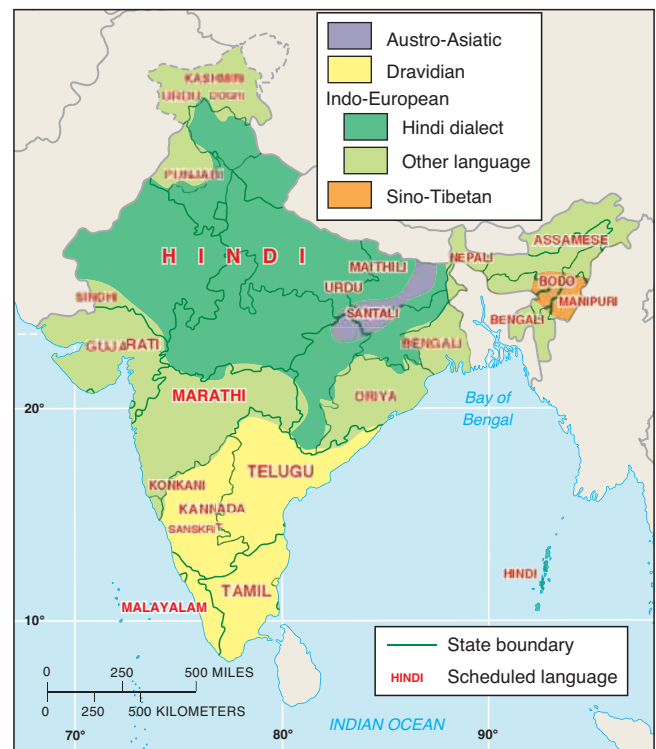
**FIGURE 5-10** Germanic branch of the Indo-European language family. Germanic languages predominate in Northern and Western Europe.

**INDIC (EASTERN) GROUP OF INDO-IRANIAN LANGUAGE BRANCH.** The most widely used languages in India, as well as in the neighboring countries of Pakistan and Bangladesh, belong to the Indo-European language family and, more specifically, to the Indic group of the Indo-Iranian branch of Indo-European.

One of the main elements of cultural diversity among the 1 billion plus residents of India is language (Figure 5-11). *Ethnologue* identifies 438 languages currently spoken in India, including 29 by at least one million people.

The official language of India is Hindi, which is an Indo-European language. Originally a variety of Hindustani spoken in the area of New Delhi, Hindi grew into a national language in the nineteenth century when the British encouraged its use in government.

After India became an independent state in 1947, Hindi was proposed as the official language, but speakers of other languages strongly objected. Consequently, English—the language



**FIGURE 5-11** Languages and language families in India. India's official language is Hindi, which has many dialects. The country has 22 scheduled languages that the government is required to protect.

of the British colonial rulers—has been retained as an official language. Speakers of different Indian languages who wish to communicate with each other sometimes are forced to turn to English as a common language.

India also recognizes 22 so-called scheduled languages, including 15 Indo-European (Assamese, Bengali, Dogri, Gujarati, Hindi, Kashmiri, Konkani, Maithili, Marathi, Nepali, Oriya, Panjabi, Sanskrit, Sindhi, and Urdu), four Dravidian (Kannada, Malayalam, Tamil, and Telugu), two Sino-Tibetan (Bodo and Manipuri), and one Austro-Asiatic (Santali). The government of India is obligated to encourage the use of these languages.

Hindi is spoken many different ways—and therefore could be regarded as a collection of many individual languages. But there is only one official way to write Hindi, using a script called Devanagari, which has been used in India since the seventh century A.D. For example, the word for *sun* is written in Hindi as सूर्य pronounced “surya.” Local differences arose in the spoken forms of Hindi but not in the written form because until recently few speakers of that language could read or write it.

Adding to the complexity, Urdu is spoken very much like Hindi, but it is recognized as a distinct language. Urdu is written with the Arabic alphabet, a legacy of the fact that most of its speakers are Muslims and their holiest book (the Quran) is written in Arabic.

**IRANIAN (WESTERN) GROUP OF INDO-IRANIAN LANGUAGE BRANCH.** Indo-Iranian languages are also spoken in Iran and neighboring countries in southwestern Asia.

These form a separate group from Indic within the Indo-Iranian branch of the Indo-European family. The major Iranian group languages include Persian (sometimes called Farsi) in Iran, Pashto in eastern Afghanistan and western Pakistan, and Kurdish, used by the Kurds of western Iran, northern Iraq, and eastern Turkey. These languages are written in the Arabic alphabet.

## Balto-Slavic Branch of Indo-European

The other Indo-European language branch with large numbers of speakers is Balto-Slavic. Slavic was once a single language, but differences developed in the seventh century A.D. when several groups of Slavs migrated from Asia to different areas of Eastern Europe and thereafter lived in isolation from one another. As a result, this branch can be divided into East, West, and South Slavic groups as well as a Baltic group. Figure 7–30 shows the widespread area populated with Balto-Slavic speakers.

**EAST SLAVIC AND BALTIC GROUPS OF THE BALTO-SLAVIC LANGUAGE BRANCH.** The most widely used Slavic languages are the eastern ones, primarily Russian, which is spoken by more than 80 percent of Russian people. Russian is one of the six official languages of the United Nations.

The importance of Russian increased with the Soviet Union's rise to power after the end of World War II in 1945. Soviet officials forced native speakers of other languages to learn Russian as a way of fostering cultural unity among the country's diverse peoples. In Eastern European countries that were dominated politically and economically by the Soviet Union, Russian was taught as the second language. The presence of so many non-Russian speakers was a measure of cultural diversity in the Soviet Union, and the desire to use languages other than Russian was a major drive in its breakup. With the demise of the Soviet Union, the newly independent republics adopted official languages other than Russian, although Russian remains the language for communications among officials in the countries that were formerly part of the Soviet Union.

After Russian, Ukrainian and Belarusian are the two most important East Slavic languages and are the official languages in Ukraine and Belarus. *Ukraine* is a Slavic word meaning "border," and *Bela-* is translated as "white."

**WEST AND SOUTH SLAVIC GROUPS OF THE BALTO-SLAVIC LANGUAGE BRANCH.** The most spoken West Slavic language is Polish, followed by Czech and Slovak. The latter two are quite similar, and speakers of one can understand the other.

The government of the former state of Czechoslovakia tried to balance the use of the two languages, even though the country contained twice as many Czechs as Slovaks. For example, the announcers on televised sports events used one of the languages during the first half and switched to the other for the second half. These balancing measures were effective in

promoting national unity during the Communist era, but in 1993, four years after the fall of communism, Slovakia split from the Czech Republic. Slovaks rekindled their long-suppressed resentment of perceived dominance of the national culture by the Czech ethnic group.

The most important South Slavic language is the one spoken in Bosnia and Herzegovina, Croatia, Montenegro, and Serbia. Bosnians and Croats write the language in the Roman alphabet (what you are reading now), whereas Montenegrans and Serbs use the Cyrillic alphabet (for example, *Yugoslavia* is written ЈУГОСЛАВИЈА).

When Bosnia and Herzegovina, Croatia, Montenegro, and Serbia were all part of Yugoslavia, the language was called Serbo-Croatian. This name now offends Bosnians and Croats because it recalls when they were once in a country that was dominated by Serbs. Instead, the names *Bosnian*, *Croatian*, and *Serbian* are preferred by people in these countries, to demonstrate that each language is unique, even though linguists consider them one.

Differences have crept into the language of the South Slavs. Bosnian Muslims have introduced Arabic words used in their religion, and Croats have replaced words regarded as having a Serbian origin with words considered to be purely Croatian. For example, the Serbo-Croatian word for martyr or hero—*junak*—has been changed to *heroj* by Croats and *shahid* by Bosnian Muslims. In the future, after a generation of isolation and hostility among Bosnians, Croats, and Serbs, the languages spoken by the three may be sufficiently different to justify their classification as distinct languages.

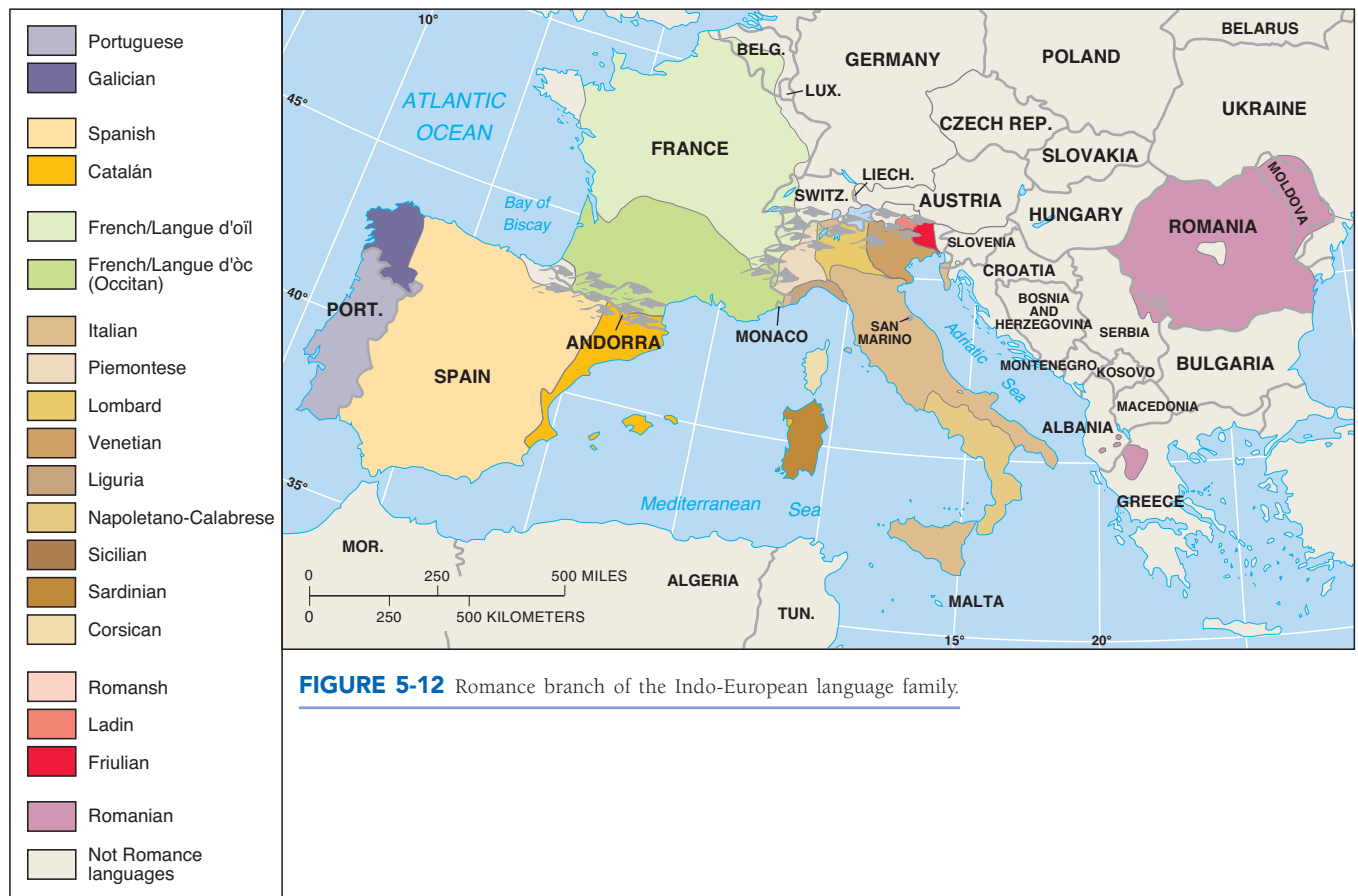
In general, differences among all of the Slavic languages are relatively small. A Czech, for example, can understand most of what is said or written in Slovak and could become fluent without much difficulty. However, because language is a major element in a people's cultural identity, relatively small differences among Slavic as well as other languages are being preserved and even accentuated in recent independence movements.

## Romance Branch of Indo-European

The Romance language branch evolved from the Latin language spoken by the Romans 2,000 years ago. The four most widely used contemporary Romance languages are Spanish, Portuguese, French, and Italian (Figure 5-12). Spanish and French are two of the six official languages of the United Nations.

The European regions in which these four languages are spoken correspond somewhat to the boundaries of the modern states of Spain, Portugal, France, and Italy. Rugged mountains serve as boundaries among these four countries. France is separated from Italy by the Alps and from Spain by the Pyrenees, and several mountain ranges mark the border between Spain and Portugal. Physical boundaries such as mountains are strong intervening obstacles, creating barriers to communication between people living on opposite sides.

The fifth most important Romance language, Romanian, is the principal language of Romania and Moldova. It is separated



from the other Romance-speaking European countries by Slavic-speaking peoples.

The distribution of Romance languages shows the difficulty in trying to establish the number of distinct languages in the world. In addition to the five languages already mentioned, two other official Romance languages are Romansh and Catalán. Romansh is one of four official languages of Switzerland, although it is spoken by only 40,000 people. Catalán is the official language of Andorra, a tiny country of 70,000 inhabitants situated in the Pyrenees Mountains between Spain and France. Catalán is also spoken by 6 million people in eastern Spain and is the official language of Spain's highly autonomous Catalonia province, centered on the city of Barcelona. A third Romance language, Sardinian—a mixture of Italian, Spanish, and Arabic—was once the official language of the Mediterranean island of Sardinia.

In addition to these official languages, several other Romance languages have individual literary traditions. In Italy, Ladin (not Latin) is spoken by 30,000 people living in the South Tyrol, and Friulian is spoken by 800,000 people in the northeast. Ladin and Friulian (along with the official Romansh) are dialects of Rhaeto-Romanic.

A Romance tongue called Ladino—a mixture of Spanish, Greek, Turkish, and Hebrew—is spoken by 100,000 Sephardic Jews, most of whom now live in Israel. None of these languages have an official status in any country, although they are used in literature.

### ORIGIN AND DIFFUSION OF ROMANCE LANGUAGES.

The Romance languages, including Spanish, Portuguese, French, Italian, and Romanian, are part of the same branch because they all developed from Latin, the “Romans’ language.” The rise in importance of the city of Rome 2,000 years ago brought a diffusion of its Latin language.

At its height in the second century A.D., the Roman Empire extended from the Atlantic Ocean on the west to the Black Sea on the east and encompassed all lands bordering the Mediterranean Sea (the empire’s boundary is shown in Figure 6-9). As the conquering Roman armies occupied the provinces of this vast empire, they brought the Latin language with them. In the process, the languages spoken by the natives of the provinces were either extinguished or suppressed in favor of the language of the conquerors.

Even during the period of the Roman Empire, Latin varied to some extent from one province to another. The empire grew over a period of several hundred years, so the Latin used in each province was based on that spoken by the Roman army at the time of occupation. The Latin spoken in each province also integrated words from the language formerly spoken in the area.

The Latin that people in the provinces learned was not the standard literary form but a spoken form, known as **Vulgar Latin** from the Latin word referring to “the masses” of the populace. Vulgar Latin was introduced to the provinces by the soldiers stationed throughout the empire. For example, the

literary term for “horse” was *equus*, from which English has derived such words as *equine* and *equestrian*. The Vulgar term, used by the common people, was *caballus*, from which are derived the modern terms for “horse” in Italian (*cavallor*), Spanish (*caballo*), Portuguese (*cavalo*), French (*cheval*), and Romanian (*cal*).

Following the collapse of the Roman Empire in the fifth century, communication among the former provinces declined, creating still greater regional variation in spoken Latin. By the eighth century, regions of the former empire had been isolated from each other long enough for distinct languages to evolve. But Latin persisted in parts of the former empire. People in some areas reverted to former languages; others adopted the languages of conquering groups of people from the north and east who spoke Germanic and Slavic.

In the past, when migrants were unable to communicate with speakers of the same language back home, major differences emerged between the languages spoken in the old and new locations, leading to the emergence of distinct, separate languages. This was the case with the migration of Latin speakers 2,000 years ago.

**ROMANCE LANGUAGE DIALECTS.** Distinct Romance languages did not suddenly appear in the former Roman Empire. As with other languages, they evolved over time. Numerous dialects existed within each province, many of which are still spoken today. The creation of standard national languages, such as French and Spanish, was relatively recent.

The dialect of the Île-de-France region, known as *Francien*, became the standard form of French because the region included Paris, which became the capital and largest city of the country. *Francien* French became the country’s official language in the sixteenth century, and local dialects tended to disappear as a result of the capital’s longtime dominance over French political, economic, and social life.

The most important surviving dialect difference within France is between the north and the south (refer to Figure 5-11). The northern dialect is known as *langue d’oïl* and the southern as *langue d’oc*. It is worth exploring these names, for they provide insight into how languages evolve.

These terms derive from different ways in which the word for “yes” was said. One Roman term for “yes” was *hoc illud est*, meaning “that is so.” In the south, the phrase was shortened to *hoc*, or *oc*, because the /h/ sound was generally dropped, just as we drop it on the word *honor* today. Northerners shortened the phrase to *o-il* after the first sound in the first two words of the phrase, again with the initial /h/ suppressed. If the two syllables of *o-il* are spoken very rapidly, they are combined into a sound like the English word *wheel*. Eventually, the final consonant was eliminated, as in many French words, giving a sound for “yes” like the English *we*, spelled in French *oui*.

A province where the southern dialect is spoken in south-western France is known as Languedoc. The southern French dialect is now sometimes called Occitan, derived from the French region of Aquitaine, which in French has a similar pronunciation to Occitan. About 2 million people in southern

France speak one of a number of Occitan dialects, including Auvergnat, Gascon, and Provençal.

Spain, like France, contained many dialects during the Middle Ages. One dialect, known as Castilian, arose during the ninth century in Old Castile, located in the north-central part of the country. The dialect spread southward over the next several hundred years as independent kingdoms were unified into one large country.

Spain grew to its approximate present boundaries in the fifteenth century, when the Kingdom of Castile and León merged with the Kingdom of Aragón. At that time, Castilian became the official language for the entire country. Regional dialects, such as Aragón, Navarre, León, Asturias, and Santander, survived only in secluded rural areas. The official language of Spain is now called Spanish, although the term *Castilian* is still used in Latin America.

Spanish and Portuguese have achieved worldwide importance because of the colonial activities of their European speakers. Approximately 90 percent of the speakers of these two languages live outside Europe, mainly in Central and South America. Spanish is the official language of 18 Latin American states, and Portuguese is spoken in Brazil, which has as many people as all the other South American countries combined and 18 times more than Portugal itself.

These two Romance languages were diffused to the Americas by Spanish and Portuguese explorers. The division of Central and South America into Portuguese- and Spanish-speaking regions is the result of a 1493 decision by Pope Alexander VI to give the western portion of the New World to Spain and the eastern part to Portugal. The Treaty of Tordesillas, signed a year later, carried out the papal decision.

The Portuguese and Spanish languages spoken in the Western Hemisphere differ somewhat from their European versions, as is the case with English. The members of the Spanish Royal Academy meet every week in a mansion in Madrid to clarify rules for the vocabulary, spelling, and pronunciation of the Spanish language around the world. The Academy’s official dictionary, published in 1992, has added hundreds of “Spanish” words that originated either in the regional dialects of Spain or the Indian languages of Latin America.

Brazil, Portugal, and several Portuguese-speaking countries in Africa agreed in 1994 to standardize the way their common language is written. Many people in Portugal are upset that the new standard language more closely resembles the Brazilian version, which eliminates most of the accent marks—such as tildes (São Paulo), cedillas (Alcobaça), circumflexes (Estância), and hyphens—and the agreement recognizes as standard thousands of words that Brazilians have added to the language.

The standardization of Portuguese is a reflection of the level of interaction that is possible in the modern world between groups of people who live tens of thousands of kilometers apart. Books and television programs produced in one country diffuse rapidly to other countries where the same language is used.

**DISTINGUISHING BETWEEN DIALECTS AND LANGUAGES.** Difficulties arise in determining whether two languages are distinct or whether they are merely two dialects of the same language:

- Galician, spoken in northwestern Spain and northeastern Portugal, is as distinct from Portuguese as, say, Catalán is from Spanish. However, Catalán is generally classified as a distinct language, and Galician is classified as a dialect of Portuguese.
- Moldovan is the official language of Moldova yet is generally classified as a dialect of Romanian.
- Flemish, the official language of northern Belgium, is generally considered a dialect of Dutch.

Several languages of Italy are viewed as different enough to merit consideration as languages distinct from Italian according to *Ethnologue*. In southern Italy, the most widespread of those possible distinct languages are Neapolitano-Calabrese, spoken by 7 million people, and Sicilian, spoken by 5 million. In the north, the most widespread are Lombard, spoken by 9 million people; Piemontese, spoken by 3 million; and Emiliano-Romagnolo, Liguria, and Venetian, spoken by 2 million each. Distinguishing individual languages from dialects is difficult, because many speakers choose to regard their languages as distinct.

Romance languages spoken in some former colonies can also be classified as separate languages because they differ substantially from the original introduced by European colonizers. Examples include French Creole in Haiti, Papiamentu (creolized Spanish) in Netherlands Antilles (West Indies), and Portuguese Creole in the Cape Verde Islands off the African coast.

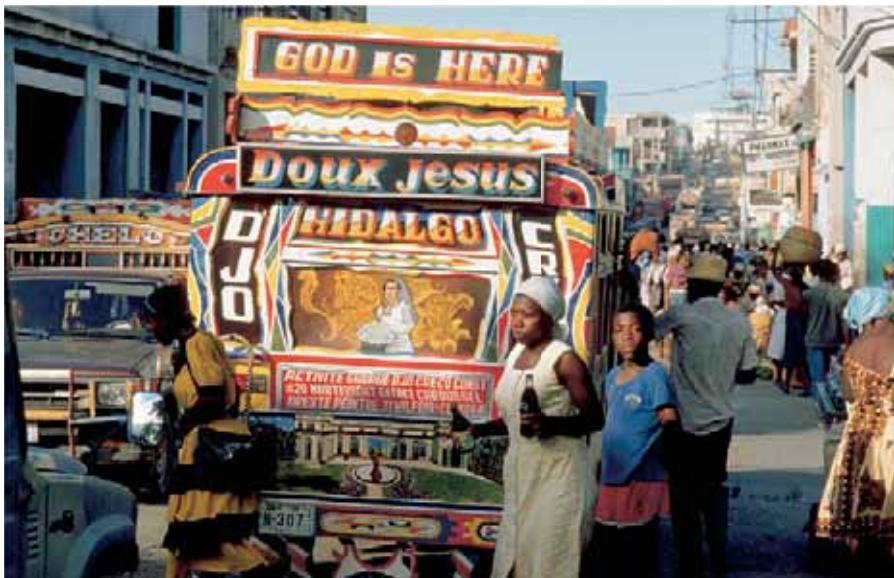
A **creole or creolized language** is defined as a language that results from the mixing of the colonizer's language with the indigenous language of the people being dominated (Figure 5-13). A creolized language forms when the colonized group adopts the language of the dominant group but makes some changes, such as simplifying the grammar and adding words from their former language. The word *creole* derives from a word in several Romance languages for a slave who is born in the master's house.

## Origin and Diffusion of Indo-European

If Germanic, Romance, Balto-Slavic, and Indo-Iranian languages are all part of the same Indo-European language family, then they must be descended from a single common ancestral language. Unfortunately, the existence of a single ancestor—which can be called Proto-Indo-European—cannot be proved with certainty, because it would have existed thousands of years before the invention of writing or recorded history.

The evidence that Proto-Indo-European once existed is “internal,” derived from the physical attributes of words themselves in various Indo-European languages. For example, the words for some animals and trees in modern Indo-European languages have common roots, including *beech*, *oak*, *bear*, *deer*, *pheasant*, and *bee*. Because all Indo-European languages share these similar words, linguists believe the words must represent things experienced in the daily lives of the original Proto-Indo-European speakers. In contrast, words for other features, such as *elephant*, *camel*, *rice*, and *bamboo*, have different roots in the various Indo-European languages. Such words therefore cannot be traced back to a common Proto-Indo-European ancestor and must have been added later, after the root language split into many branches. Individual Indo-European languages share common root words for *winter* and *snow* but not for *ocean*. Therefore, linguists conclude that original Proto-Indo-European speakers probably lived in a cold climate, or one that had a winter season, but did not come in contact with oceans.

Linguists and anthropologists generally accept that Proto-Indo-European must have existed, but they disagree on when and where the language originated and the process and routes by which it diffused. The debate over place of origin and paths of diffusion is significant, because one theory argues that language diffused primarily through warfare and conquest, and the other theory argues that the diffusion resulted from peaceful sharing of food. So where did Indo-European originate? Not surprisingly, scholars disagree on



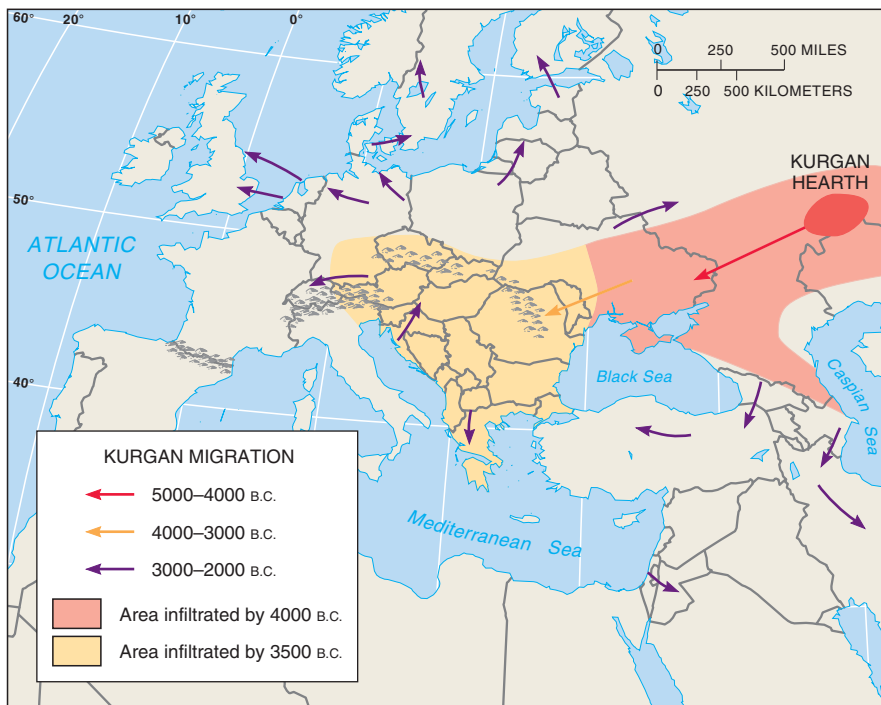
**FIGURE 5-13** Creole. A mix of French and English adorns this public bus in Port-au-Prince, capital of Haiti. French and Haitian Creole, a dialect of French, are both official languages in Haiti, although English is the lingua franca in Haiti because of the country's proximity to the United States.

where and when the first speakers of Proto-Indo-European lived.

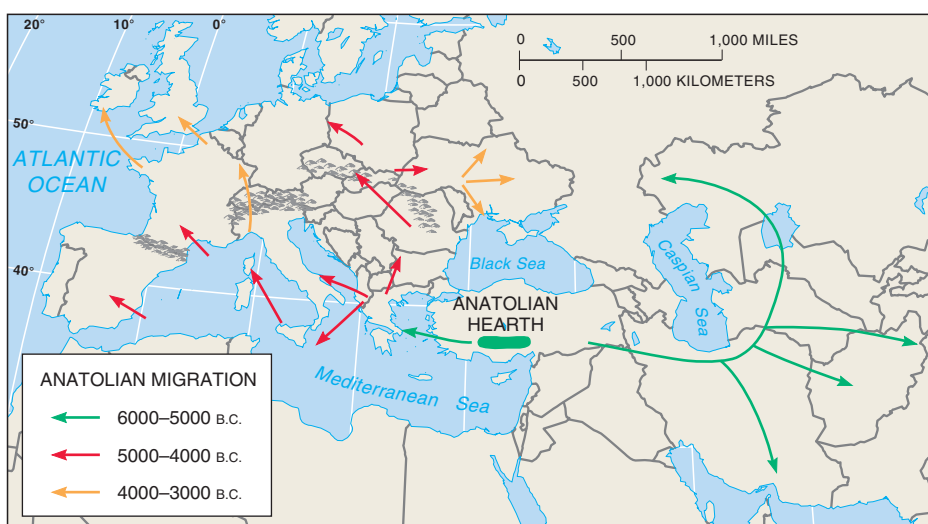
- **Nomadic Warrior Thesis.** One influential hypothesis, espoused by Marija Gimbutas, is that the first Proto-Indo-European speakers were the Kurgan people, whose homeland was in the steppes near the border between present-day Russia and Kazakhstan. The earliest

archaeological evidence of the Kurgans dates to around 4300 B.C.

The Kurgans were nomadic herders. Among the first to domesticate horses and cattle, they migrated in search of grasslands for their animals. This took them westward through Europe, eastward to Siberia, and southward to Iran and South Asia. Between 3500 and 2500 B.C., Kurgan warriors, using their domesticated horses as weapons, conquered much of Europe and South Asia (Figure 5-14).



**FIGURE 5-14** Origin and diffusion of Indo-European (nomadic warrior theory). The Kurgan homeland was north of the Caspian Sea, near the present-day border between Russia and Kazakhstan. According to this theory, the Kurgans may have infiltrated into Eastern Europe beginning around 4000 B.C. and into central Europe and southwestern Asia beginning around 2500 B.C.



**FIGURE 5-15** Origin and diffusion of Indo-European (sedentary farmer theory). Indo-European may have originated in present-day Turkey 2,000 years before the Kurgans. According to this theory, the language diffused along with agricultural innovations west into Europe and east into Asia.

- **Sedentary Farmer Thesis.** Archaeologist Colin Renfrew argues that the first speakers of Proto-Indo-European lived 2,000 years before the Kurgans, in eastern Anatolia, part of present-day Turkey (Figure 5-15). Biologist Russell D. Gray supports the Renfrew position but dates the first speakers even earlier, at around 6700 B.C.

Renfrew believes they diffused from Anatolia westward to Greece (the origin of the Greek language branch) and from Greece westward toward Italy, Sicily, Corsica, the Mediterranean coast of France, Spain, and Portugal (the origin of the Romance language branch). From the Mediterranean coast, the speakers migrated northward toward central and northern France and on to the British Isles (perhaps the origin of the Celtic language branch).

Indo-European is also said to have diffused northward from Greece toward the Danube River (Romania) and westward to central Europe, according to Renfrew. From there the language diffused northward toward the Baltic Sea (the origin of the Germanic language branch) and eastward toward the Dnepr River near Ukraine (the origin of the Slavic language branch). From the Dnepr River, speakers migrated eastward to the Dnepr River (the homeland of the Kurgans).

The Indo-Iranian branch of the Indo-European language family originated either directly through migration from Anatolia along the south shores of the Black and Caspian seas by way of Iran and Pakistan, or indirectly by way of Russia north of the Black and Caspian seas.

Renfrew argues that Indo-European diffused into Europe and South Asia along with agricultural practices rather than by



military conquest. The language triumphed because its speakers became more numerous and prosperous by growing their own food instead of relying on hunting.

Regardless of how Indo-European diffused, communication was poor among different peoples, whether warriors or farmers. After many generations of complete isolation, individual groups evolved increasingly distinct languages.

### KEY ISSUE 3

## Where Are Other Language Families Distributed?

- Classification of Languages
- Distribution of Language Families

This section describes where different languages are found around the world. The several thousand spoken languages can be organized logically into a small number of language families. Larger language families can be further divided into language branches and language groups. ■

## Classification of Languages

Figure 5-16 shows the world's language families:

- A language in the *Indo-European family*, such as English, is spoken by 46 percent of the world's people.
- A language in the *Sino-Tibetan family*, such as Mandarin, is spoken by 21 percent of the world, mostly in China.
- A language in the *Afro-Asiatic family*, including Arabic, is spoken by 6 percent, mostly in the Middle East
- A language in the *Austronesian family* is spoken by 6 percent, mostly in Southeast Asia.
- A language in the *Niger-Congo family* is spoken by 6 percent, mostly in Africa.
- A language in the *Dravidian family* is spoken by 4 percent, mostly in India.
- A language in the *Altaic family* is spoken by 2 percent, mostly in Asia.
- A language in the *Austro-Asiatic family* is spoken by 2 percent, mostly in Southeast Asia.
- *Japanese*, a separate language family, is spoken by 2 percent.
- The remaining 5 percent of the world's people speak a language belonging to one of 100 smaller families.

Figure 5-17 attempts to depict differences among language families, branches, and groups. Language families form the trunks of the trees, whereas individual languages are displayed as leaves. The larger the trunks and leaves are, the greater the number of speakers of those families and languages. Some trunks divide into several branches, which logically represent

language branches. The branches representing Germanic, Balto-Slavic, and Indo-Iranian in Figure 5-17 divide a second time into language groups.

Figure 5-17 displays each language family as a separate tree at ground level because differences among families predate recorded history. Linguists speculate that language families were joined together as a handful of superfamilies tens of thousands of years ago. Superfamilies are shown as roots below the surface because their existence is highly controversial and speculative.

## Distribution of Language Families

Nearly one-half the people in the world speak an Indo-European language. The second-largest family is Sino-Tibetan, spoken by one-fifth of the world. Another half-dozen families account for most of the remainder.

### Sino-Tibetan Family

The Sino-Tibetan family encompasses languages spoken in the People's Republic of China—the world's most populous state at more than 1 billion—as well as several smaller countries in Southeast Asia. The languages of China generally belong to the Sinitic branch of the Sino-Tibetan family.

There is no single Chinese language. Rather, the most important is Mandarin (or, as the Chinese call it, *pu tong hua*—“common speech”). Spoken by approximately three-fourths of the Chinese people, Mandarin is by a wide margin the most used language in the world. Once the language of emperors in Beijing, Mandarin is now the official language of both the People's Republic of China and Taiwan, as well as one of the six official languages of the United Nations. Other Sinitic branch languages are spoken by tens of millions of people in China, mostly in the southern and eastern parts of the country—Wu, Yue (also known as Cantonese), Min, Jinyu, Xiang, Hakka, and Gan. However, the Chinese government is imposing Mandarin countrywide.

The relatively small number of languages in China (compared to India, for example) is a source of national strength and unity. Unity is also fostered by a consistent written form for all Chinese languages. Although the words are pronounced differently in each language, they are written the same way.

You already know the general structure of Indo-European quite well because you are a fluent speaker of at least one Indo-European language. But the structure of Chinese languages is quite different (Figure 5-18). They are based on 420 one-syllable words. This number far exceeds the possible one-syllable sounds that humans can make, so Chinese languages use each sound to denote more than one thing. The sound *shi*, for example, may mean “lion,” “corpse,” “house,” “poetry,” “ten,” “swear,” or “die.” The sound *jian* has more than 20 meanings, including “to see.” The listener must infer the meaning from the context in the sentence and the intonation the speaker uses. In addition, two one-syllable words can be combined into two syllables, forming a new word. For example, the two-syllable word “Shanghai” is a