

## KEY ISSUE 1

## Where Are Languages Distributed?

- Classification of Languages
- Distribution of Language Families

How many languages do you speak? If you are Dutch, you were required to learn at least two foreign languages in high school. For those of you who do not happen to be Dutch, the number is probably a bit lower.

In fact, most people in the United States know only English. Fewer than one-half of American high school students have studied a foreign language. In contrast, nearly two-thirds of graduates from Dutch high schools have learned at least three foreign languages. Even in other English-speaking countries, foreign languages are studied more frequently than in the United States. For example, two-thirds of 10-year-olds in the United Kingdom are learning a foreign language in school.

**Language** is a system of communication through speech, a collection of sounds that a group of people understands to have the same meaning. Many languages also have a **literary tradition**, or a system of written communication. However, hundreds of spoken languages lack a literary tradition. The lack of written records makes it difficult to document the distribution of many languages.

Many countries designate at least one language as their **official language**, which is the one used by the government for laws, reports, and public objects, such as road signs, money, and stamps. A country with more than one official language may require all public documents to be in all languages. Logically, an official language would be understood by most if not all of the country's citizens, but some countries



▲ **FIGURE 5-2 THE WORLD'S THREE MOST WIDELY SPOKEN LANGUAGES** (left) This sign in New York's Chinatown includes English, Mandarin, and Spanish. (right) Mandarin on an iPhone.

that were once British colonies designate English as an official language, even though few of their citizens can speak it.

Earth's heterogeneous collection of languages is one of its most obvious examples of cultural diversity. *Ethnologue*, one of the most authoritative sources of languages (see [www.ethnologue.com](http://www.ethnologue.com)), estimates that the world has 6,909 languages. Only 11 of these languages, including English, are spoken by at least 100 million people (Figure 5-2). Four of these are relatively familiar to North Americans (German, Portuguese, Spanish, and Russian), but others are less familiar (Arabic, Bengali, Hindi, Japanese, Lahnda, and Mandarin).

Approximately 85 languages are spoken by at least 10 million people, and approximately 300 languages by between 1 million and 10 million. The remaining 6,524 languages are spoken by fewer than 1 million people each. The distribution of some of these languages is easy for geographers to document, whereas others—especially in Africa and Asia—are difficult, if not impossible.

The world's languages can be organized into families, branches, and groups:

- A **language family** is a collection of languages related through a common ancestral language that existed long before recorded history.
- A **language branch** is a collection of languages within a family related through a common ancestral language that existed several thousand years ago; differences are not as extensive or as old as between language families, and archaeological evidence can confirm that the branches derived from the same family.
- A **language group** is a collection of languages within a branch that share a common origin in the relatively recent past and display many similarities in grammar and vocabulary.



# Classification of Languages

## Learning Outcome 5.1.1

### Name the largest language families.

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.

Figure 5-3 depicts differences among language families, branches, groups, and individual languages:

- Language families form the trunks of the trees.
- Individual languages are displayed as leaves.
- Some trunks divide into several branches, which logically represent language branches.
- The branches representing Germanic, Balto-Slavic, and Indo-Iranian in Figure 5-3 divide a second time into language groups.

The larger the trunks and leaves are, the greater the number of speakers of those families and languages.

Two-thirds of the people in the world speak a language that belongs to the Indo-European or Sino-Tibetan language family. Seven other language families are used by between 2 and 6 percent of the world (Figure 5-4). The remaining 5 percent of the world's people speak a language belonging to one of 100 smaller families.

Figure 5-3 displays each language family as a separate tree at ground level because differences among families predate recorded history. Some 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. A researcher in New Zealand, Quentin Atkinson, carries the speculation further, arguing that all languages can be ultimately traced to Africa. According to Atkinson, languages are most complex and diverse in Africa. Atkinson thinks humans outside Africa display less linguistic diversity because their languages have had a shorter time in which to evolve into new languages than have African languages.

### Pause and Reflect 5.1.1

Based on Figure 5-3, what are the language family, branch, and group to which English belongs?

► **FIGURE 5-3 LANGUAGE FAMILY TREE** Language families with at least 10 million speakers according to *Ethnologue* are shown as trunks of trees. Some language families are divided into branches and groups. Individual languages that have more than 5 million speakers are shown as leaves. Below ground level, the language tree's "roots" are shown, but these are speculative because they predated recorded history.

Printed by Jonathan Linne (jonathan.linne@mpls.k12.mn.us) on 10/2/2014 from 205.215.177.155 authorized to use until 8/27/2020. Use beyond the authorized user or valid subscription date represents a copyright violation.

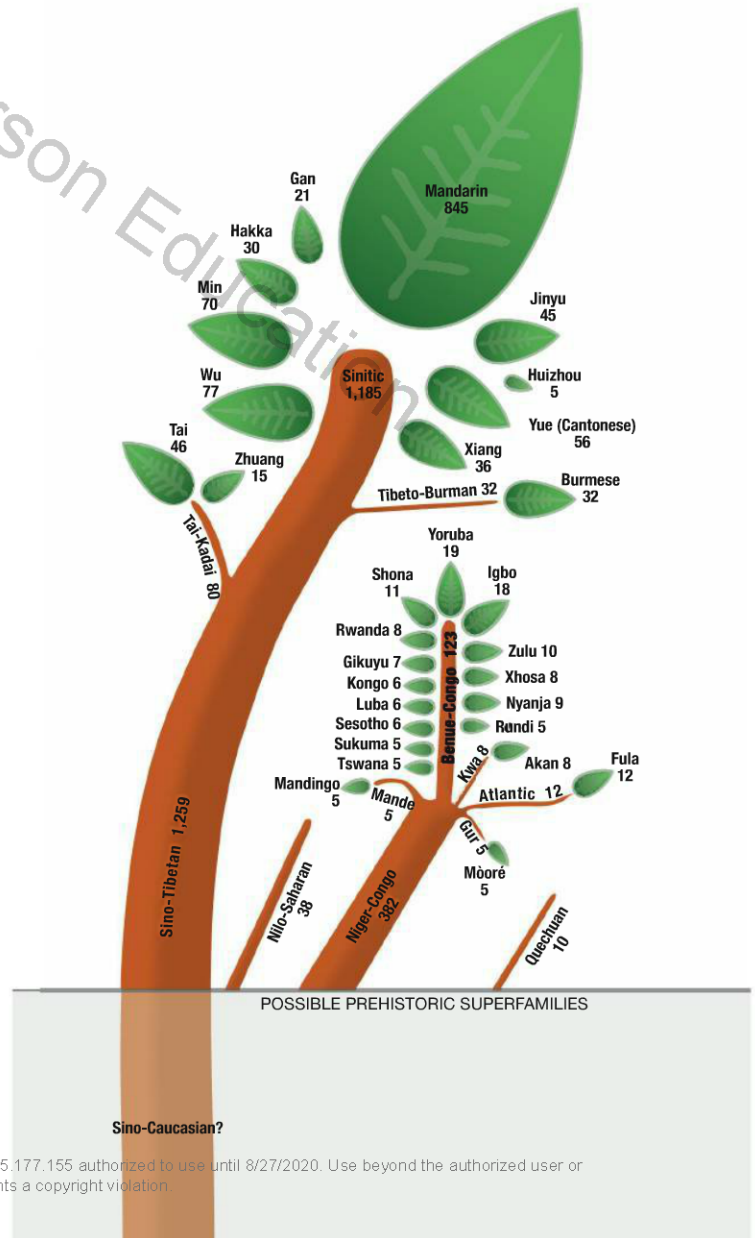
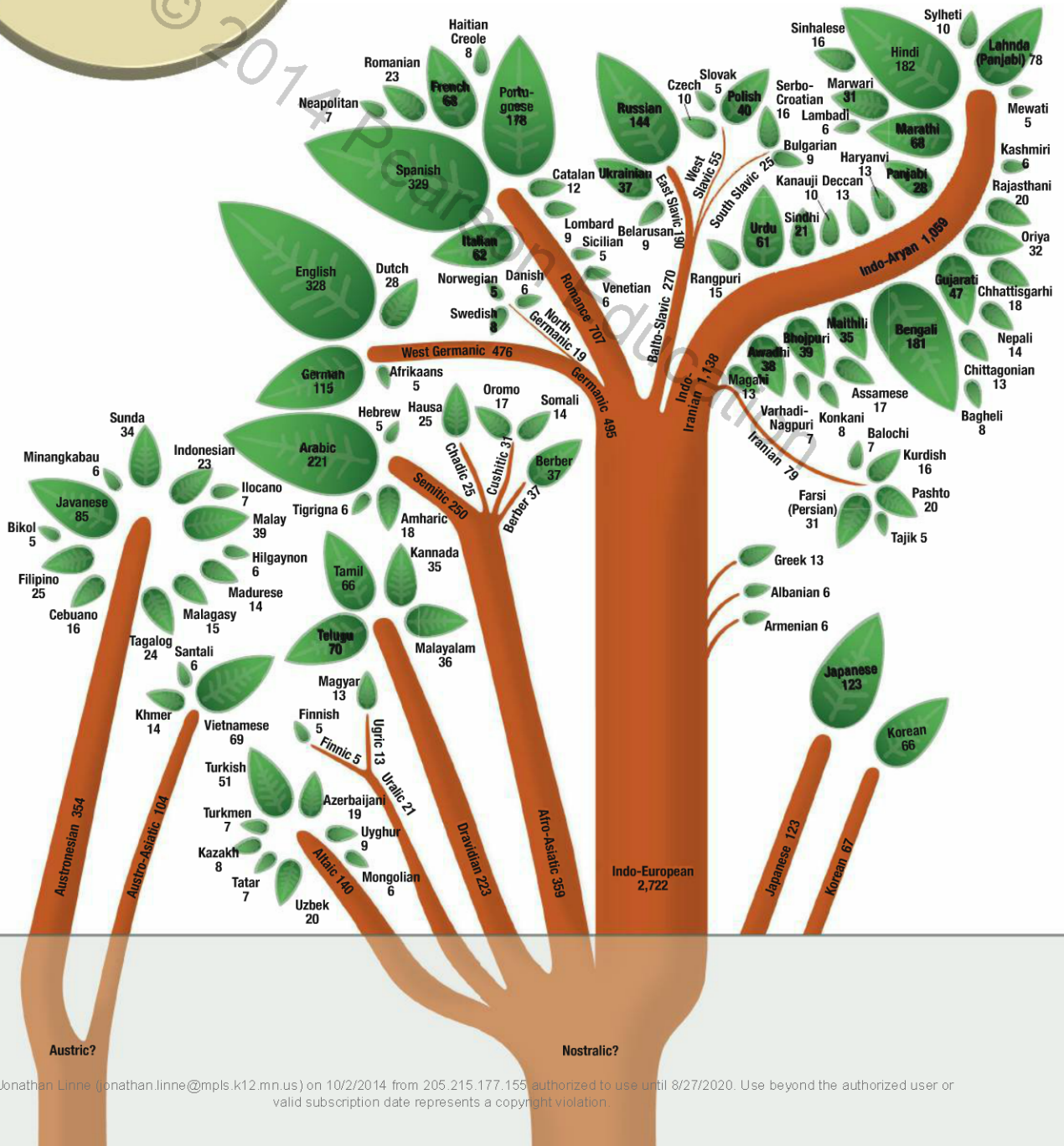
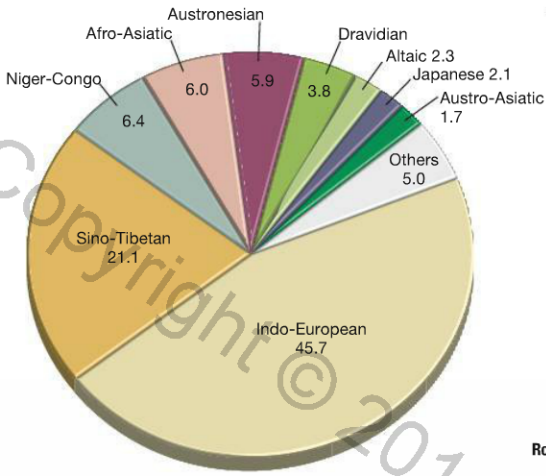


FIGURE 5-4 SHARE OF EACH LANGUAGE FAMILY The chart shows the percentage of people who speak a language from each major family.



## Distribution of Language Families

### Learning Outcome 5.1.2

Identify the names and distribution of the two largest language families.

Language families with at least 10 million native speakers are shown in Figure 5-5. Individual languages with at least 50 million speakers are named on the map.

### INDO-EUROPEAN

Indo-European, the most widely used language family, is the predominant one in Europe, South Asia, and North and Latin America. Its origin and distribution are discussed in more detail in the next key issue.

### SINO-TIBETAN

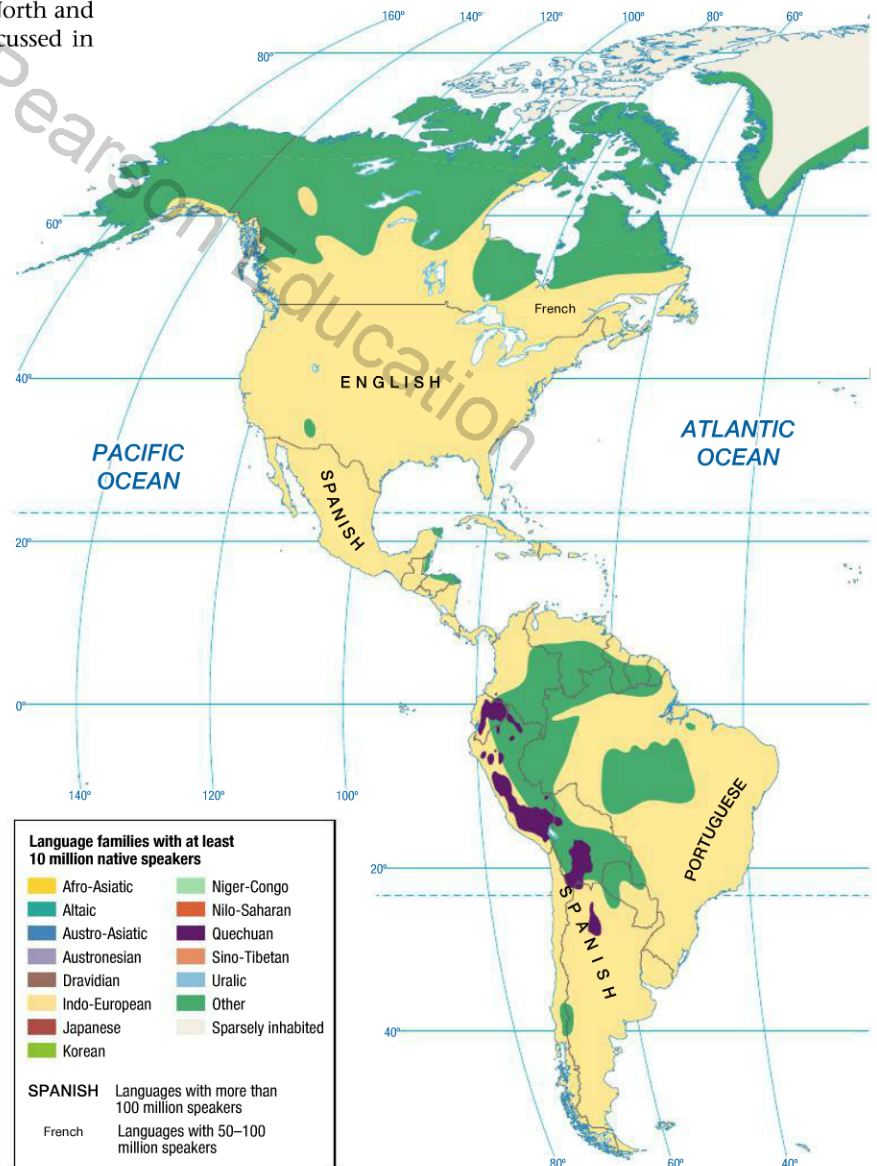
The Sino-Tibetan family encompasses languages spoken in the People's Republic of China—the world's most populous state, at 1.3 billion—and in several smaller countries in Southeast Asia. The languages of China generally belong to the Sinitic branch of the Sino-Tibetan family.

There is no single Chinese language. Rather, the most commonly used is Mandarin (or, as the Chinese call it, Putonghua—"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, and it is one of the six official languages of the United Nations. Seven other Sinitic branch languages are spoken by at least 20 million each in China, mostly in the southern and eastern parts of the country—Wu, Min, Yue (also known as Cantonese), 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. They are written primarily with **logograms**, which are symbols that represent words, or meaningful parts of words, rather than sounds (as in English). Ability to read a book requires understanding several thousand logograms (Figure 5-6). Most logograms are compounds; words related to bodies of water, for example, include a symbol that represents a river, plus additional strokes that alter the river in some way.



► **FIGURE 5-5 DISTRIBUTION OF LANGUAGE FAMILIES** Most language can be classified into one of a handful of language families.

River 河  
 River bed 河床  
 Lake 湖  
 Stream 流  
 Riptide 冲

FIGURE 5-6 CHINESE Similar logograms represent various water-related words.



Printed by Jonathan Linne (jonathan.linne@mpls.k12.mn.us) on 10/2/2014 from 205.215.177.155 authorized to use until 8/27/2020. Use beyond the authorized user or valid subscription date represents a copyright violation.

## OTHER ASIAN LANGUAGE FAMILIES

### Learning Outcome 5.1.3

Identify the names and distribution of the largest language families in addition to Indo-European and Sino-Tibetan.

In addition to Sino-Tibetan, several other language families spoken by large numbers of people can be found in East and Southeast Asia. If you look at their distribution in Figure 5-5, you can see a physical reason for their independent development: These language families are clustered on either islands or peninsulas.

**AUSTRONESIAN.** Austronesian languages are spoken by about 6 percent of the world's people, who are mostly in Indonesia, the world's fourth-most-populous country. With its inhabitants dispersed among thousands of islands, Indonesia has an extremely large number of distinct languages and dialects; *Ethnologue* identifies 722 actively used languages in Indonesia. Indonesia's most widely used first language is Javanese, spoken by 85 million people, mostly on the island of Java, where two-thirds of the country's population is clustered. As Figures 5-1, 5-3, and 5-5 show, Malagasy also belongs to the Austronesian family because of migration from Indonesia to Madagascar 2,000 years ago.

**AUSTRO-ASIATIC.** Spoken by about 2 percent of the world's population, Austro-Asiatic is based in Southeast Asia. Vietnamese, the most-spoken tongue of the Austro-Asiatic language family, is written with our familiar Roman alphabet, with the addition of a large number of diacritical marks above the vowels. The Vietnamese alphabet was devised in the seventeenth century by Roman Catholic missionaries.

**TAI KADAI.** The Tai Kadai family was once classified as a branch of Sino-Tibetan. The principal languages of this family are spoken in Thailand and neighboring portions of China. Similarities with the Austronesian family have led some linguistic scholars to speculate that people speaking these languages may have migrated from the Philippines.

**JAPANESE.** Written in part with Chinese logograms, Japanese also uses two systems of phonetic symbols, like Western languages, used either in place of the logograms or alongside them. Chinese cultural traits have diffused into Japanese society, including the original form of writing the Japanese language. But the structures of the two languages differ. Foreign terms may be written with one of these sets of phonetic symbols.

**KOREAN.** Unlike Sino-Tibetan languages and Japanese, Korean is written not with logograms but in a system known as hankul (also called hangul or onmun). In this system, each letter represents a sound, as in Western languages. More than half of the Korean vocabulary

derives from Chinese words. In fact, Chinese and Japanese words are the principal sources for creating new words to describe new technology and concepts.

### Pause and Reflect 5.1.3

If you are trying to recall where in the world language families are distributed, keep in mind that most of them are named for regions or countries. Based on their names, how would you expect the language families Austronesian and Austro-Asiatic to differ in their distribution?

## LANGUAGES OF SOUTHWEST ASIA & NORTH AFRICA AND CENTRAL ASIA

The two largest language families in Southwest Asia & North Africa and Central Asia are Afro-Asiatic and Altaic. Uralic languages were once classified with Altaic.

**AFRO-ASIATIC.** Arabic is the major language of the Afro-Asiatic family, an official language in two dozen countries of Southwest Asia & North Africa, and one of six official languages of the United Nations. In addition to the 200-million-plus native speakers of Arabic, a large percentage of the world's Muslims have at least some knowledge of Arabic because Islam's holiest book, the Quran (Koran), was written in that language in the seventh century. The Afro-Asiatic family also includes Hebrew, the language of the Bible.

**ALTAIC.** Altaic languages are thought to have originated in the steppes bordering the Qilian Shan and Altai mountains between Tibet and China. Present distribution covers an 8,000-kilometer (5,000-mile) band of Asia. The Altaic language with by far the most speakers is Turkish (Figure 5-7).

When the Soviet Union governed most of the Altaic-speaking region of Central Asia, use of Altaic languages was suppressed to create a homogeneous national culture. With the dissolution of the Soviet Union in the early 1990s, Altaic languages became official in several newly independent countries, including Azerbaijan, Kazakhstan, Kyrgyzstan, Turkmenistan, and Uzbekistan.

**URALIC.** Every European country is dominated by Indo-European speakers, except for three—Estonia, Finland, and Hungary. The Estonians, Finns, and Hungarians speak languages that belong to the Uralic family.

The Altaic and Uralic language families were once thought to be linked as one family, but recent studies point to geographically distinct origins. Uralic languages are traceable back to a common language, Proto-Uralic, first used 7,000 years ago by people living in the Ural Mountains of present-day Russia, north of the Kurgan homeland. Migrants carried the Uralic languages to Europe, carving out homelands for themselves in the midst of Germanic- and Slavic-speaking peoples and retaining their language as a major element of cultural identity.

## AFRICAN LANGUAGE FAMILIES

No one knows the precise number of languages spoken in Africa, and scholars disagree on classifying those known into families. In the 1800s, European missionaries and colonial officers began to record African languages using the Roman or Arabic alphabet. More than 1,000 distinct languages and several thousand named dialects have been documented. Most lack a written tradition.

**NIGER-CONGO.** More than 95 percent of the people in sub-Saharan Africa speak languages of the Niger-Congo family (Figure 5-8). One of these languages—Swahili—is the first language of only 800,000 people and an official language in only one country (Tanzania), but it is spoken as a second language by approximately 30 million Africans.

Especially in rural areas, the local language is used to communicate with others from the same village, and Swahili is used to communicate with outsiders. Swahili originally developed through interaction among African groups and Arab traders, so its vocabulary has strong Arabic influences. Also, Swahili is one of the few African languages with an extensive literature.

**NILO-SAHARAN.** Languages of the Nilo-Saharan family are spoken by a few million people in north-central Africa, immediately north of the Niger-Congo language region.

▼ **FIGURE 5-7 TURKISH** In 1928, Turkey's leader Kemal Ataturk ordered Turks to write Turkish with Roman letters instead of Arabic. Ataturk believed that using Roman letters would help modernize Turkey's economy and culture through increased communications with European countries. This painting depicts Ataturk writing with Roman letters.



Printed by Jonathan Linne (jonathan.linne@mpls.k12.mn.us) on 10/2/2014 from 205.215.177.155 authorized to use until 8/27/2020. Use beyond the authorized user or valid subscription date represents a copyright violation.

Divisions within the Nilo-Saharan family exemplify the problem of classifying African languages. Despite having relatively few speakers, the Nilo-Saharan family is divided into six branches, plus numerous groups and subgroups. The total number of speakers of each individual Nilo-Saharan language is extremely small.

**KHOISAN.** A distinctive characteristic of the Khoisan languages is the use of clicking sounds. Upon hearing this, whites in southern Africa derisively and onomatopoeically named the most important Khoisan language Hottentot.

▼ **FIGURE 5-8 AFRICA'S LANGUAGE FAMILIES** More than 1,000 languages have been identified in Africa, and experts do not agree on how to classify them into families, especially languages in central Africa. Languages with more than 5 million speakers are named on the map. The great number of languages results from at least 5,000 years of minimal interaction among the thousands of cultural groups inhabiting the African continent. Each group developed its own language, religion, and other cultural traditions in isolation from other groups.



### CHECK-IN: KEY ISSUE 1

#### Where Are Languages Distributed?

- ✓ Languages can be classified into families and branches.
- ✓ The two largest families are Indo-European, which is found primarily in Europe, North America, Latin America, and South Asia, and Sino-Tibetan, which is clustered primarily in East Asia.

## KEY ISSUE 2

# Why Is English Related to Other Languages?

- Distribution of Indo-European Branches
- Origin and Diffusion of Language Families

### Learning Outcome 5.2.1

Learn the distribution of the Germanic and Indo-Iranian branches of Indo-European.

English belongs to Indo-European, the world's most widely spoken language family. Indo-European is divided into eight branches (Figure 5-9).

## Distribution of Indo-European Branches

This section discusses the distribution of the four most widely spoken Indo-European branches.

### GERMANIC BRANCH

German may seem like a difficult language for many English speakers to learn, but the two languages are structurally similar and have many words in common. Both belong to the West Germanic group

**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 widely used 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.



Printed by Jonathan Linne (jonathan.linne@mpls.k12.mn.us) on 10/2/2014 from 205.215.177.155 authorized to use until 8/27/2020. Use beyond the authorized user or valid subscription date represents a copyright violation.

of the Germanic language branch of the Indo-European family (Figure 5-10).

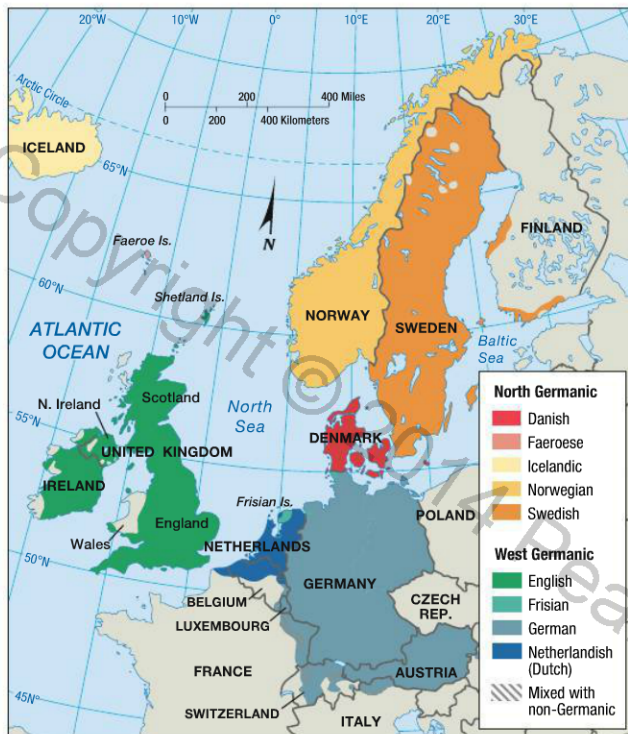
West Germanic is further divided into High Germanic and Low Germanic subgroups, so named because they are found in high and low elevations within present-day Germany. High German, spoken in the southern mountains of Germany, is the basis for the modern standard German language. English is classified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

The branch of the Indo-European language family with the most speakers is Indo-Iranian. This branch includes more





▲ **FIGURE 5-10 LANGUAGE GROUPS OF THE GERMANIC BRANCH** Germanic languages predominate in Northern and Western Europe.

than 100 individual languages. The branch is divided into an eastern group (Indic) and a western group (Iranian).

**INDIC (EASTERN) GROUP.** 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 languages spoken by at least 1 million people. 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.

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

India also recognizes 22 so-called scheduled languages, including 15 Indo-European (Assamese, Bengali, Dogri,



▲ **FIGURE 5-11 LANGUAGE FAMILIES IN INDIA** India's principal official language is Hindi, which has many dialects. The country has 22 scheduled languages that the government is required to protect.

Gujarati, Hindi, Kashmiri, Konkani, Maithili, Marathi, Nepali, Oriya, Panjabi, Sanskrit, Sindhi, and Urdu), 4 Dravidian (Kannada, Malayalam, Tamil, and Telugu), 2 Sino-Tibetan (Bodo and Manipuri), and 1 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 and pronounced “surya.”

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.** Indo-Iranian languages are spoken in Iran and neighboring countries in Central 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

### Learning Outcome 5.2.2

Learn the distribution of the Balto-Slavic and Romance branches of Indo-European.

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 other. As a result, this branch can be divided into East, West, and South Slavic groups as well as a Baltic group. Figure 5-9 shows the widespread area populated with Balto-Slavic speakers.

**EAST SLAVIC AND BALTIC GROUPS.** The most widely used Slavic languages are the eastern ones, primarily Russian, which is spoken by more than 80 percent of Russian people (Figure 5-12). 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

▼ **FIGURE 5-12 RUSSIAN** New York City's Brighton Beach neighborhood is home to many Russian immigrants, including this bookseller. The red words say "Christian Library," with the text below announcing Saturday services at the Russian Evangelical Baptist Church.



other than Russian, although Russian remains the language for communication among officials in the countries that were formerly part of the Soviet Union.

After Russian, Ukrainian and Belarusian are the two most commonly used East Slavic languages and are the official languages in Ukraine and Belarus. Ukraine is a Slavic word meaning "border," and *bela-* means "white."

**WEST AND SOUTH SLAVIC GROUPS.** 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 widely used South Slavic language is the one spoken in Bosnia & Herzegovina, Croatia, Montenegro, and Serbia. When Bosnia & 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. 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, Serbia is written Србија).

Differences have crept into the South Slavic languages. 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.



▲ **FIGURE 5-13 ROMANCE BRANCH OF THE INDO-EUROPEAN LANGUAGE FAMILY** Romance branch languages predominate in southwestern Europe.

### Pause and Reflect 5.2.2

On the map of Europe, which branch predominates to the north, which to the south, and which to the east?

## ROMANCE BRANCH

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 5-13). 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 widely used 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 (Figure 5-14). 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-Romantic.

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.



▲ **FIGURE 5-14 CATALÁN** The sign says "Passage is restricted to workers." The sign is in front of Sagrada Família church, designed by Antoni Gaudí, in Barcelona, Spain.

## Origin and Diffusion of Language Families

### Learning Outcome 5.2.3

#### Understand the origin and diffusion of English.

Like other cultural elements, the contemporary distribution of languages exists because of patterns of origin and diffusion. The origin and initial diffusion of language families predate recorded history, so we can only speculate about them. On the other hand, language branches and individual languages have originated and diffused since recorded history began. The origin and diffusion of the English language and the Romance language branches are examples.

### ORIGIN AND DIFFUSION OF ENGLISH

The British Isles had been inhabited for thousands of years, but we know nothing of their early languages until tribes called the Celts arrived around 2000 B.C., speaking languages we call Celtic. Around A.D. 450, tribes from mainland Europe invaded, pushing the Celts into the remote northern and western parts of Britain, including Cornwall and the highlands of Scotland and Wales.

**GERMAN INVASION.** The tribes invading the British Isles were known as the Angles, Jutes, and Saxons. All three were Germanic tribes—the Jutes from northern Denmark, the Angles from southern Denmark, and the Saxons from northwestern Germany (Figure 5-15). Modern English has evolved primarily from the language spoken by the Angles, Jutes, and Saxons.

The three tribes who brought the beginnings of English to the British Isles shared a language similar to that of other peoples in the region from which they came. Today, English people and others who trace their cultural heritage back to England are often called Anglo-Saxons, after the two larger tribes.

The name England comes from *Angles' land*. In Old English, *Angles* was spelled *Engles*, and the Angles' language was known as *englisc*. The Angles came from a corner, or *angle*, of Germany known as Schleswig-Holstein.

At some time in history, all Germanic people spoke a common language, but that time predates written records. The common origin of English with other Germanic languages can be reconstructed by analyzing language differences that emerged after Germanic groups migrated to separate territories and lived in isolation from each other, allowing their languages to continue evolving independently.

Other peoples subsequently invaded England and added their languages to the basic English. Vikings from present-day Norway landed on the northeastern coast of England in the ninth century. Although defeated in their effort to conquer the islands, many Vikings remained in the country and enriched the language with new words.



**▲ FIGURE 5-15 INVASIONS OF ENGLAND** The first speakers of the language that became known as English were tribes that lived in present-day Germany and Denmark. They invaded England in the fifth century. The Jutes settled primarily in southeastern England, the Saxons in the south and west, and the Angles in the north, eventually giving the country its name—Angles' Land, or England. Invasions by Vikings in the ninth century and Normans in the eleventh century brought new words to the language spoken in the British Isles. The Normans were the last successful invaders of England.

**NORMAN INVASION.** English is quite different from German today primarily because England was conquered by the Normans in 1066. The Normans, who came from present-day Normandy in France, spoke French, which they established as England's official language for the next 300 years. The leaders of England, including the royal family, nobles, judges, and clergy, therefore spoke French. However, the majority of the people, who had little education, did not know French, so they continued to speak English to each other.

England lost control of Normandy in 1204, during the reign of King John, and entered a long period of conflict with France. As a result, fewer people in England wished to speak French, and English again became the country's unchallenged dominant language. Recognizing that nearly everyone in England was speaking English, Parliament enacted the Statute of Pleading in 1362 to change the official language of court business from French to English. However, Parliament continued to conduct business in French until 1489.

During the 300-year period that French was the official language of England, the Germanic language used by the common people and the French used by the leaders mingled to form a new language. Modern English owes its simpler, straightforward words, such as *sky*, *horse*, *man*, and *woman*, to its Germanic roots, and fancy, more elegant words, such as *celestial*, *equestrian*, *masculine*, and *feminine*, to its French invaders.

The contemporary distribution of English speakers around the world exists because the people of England migrated with their language when they established colonies during over the course of four centuries.

### Pause and Reflect 5.2.3

**School and university: which word is Germanic and which is Romance?**

**DIFFUSION TO NORTH AMERICA.** English first diffused west from England to North America in the seventeenth century. The first successful English colonies were built in North America, beginning with Jamestown, Virginia, in 1607, and Plymouth, Massachusetts, in 1620. After England defeated France in a battle to dominate the North American colonies during the eighteenth century, the position of English as the principal language of North America was assured.

Similarly, the British took control of Ireland in the seventeenth century, South Asia in the mid-eighteenth century, the South Pacific in the late eighteenth and early nineteenth centuries, and southern Africa in the late nineteenth century. In each case, English became an official language, even if only the colonial rulers and a handful of elite local residents could speak it.

More recently, the United States has been responsible for diffusing English to several places, most notably the Philippines, which Spain ceded to the United States in 1899, a year after losing the Spanish-American War. After gaining full independence in 1946, the Philippines retained English as one of its official languages, along with Filipino.

## 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 the 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 8-11.) 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 (Figure 5-16). 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.



▲ **FIGURE 5-16 PARIS LATIN QUARTER** The university area of Paris gets its name from the Latin language, which was the language of instruction when the universities were founded during the Middle Ages.

## Origin and Diffusion of Indo-European

### Learning Outcome 5.2.4

Understand the two theories of the origin and diffusion of Indo-European.

If Germanic, Romance (Figure 5-17), 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; one theory argues that language diffused primarily through warfare and conquest, and another theory argues that the diffusion resulted from peaceful sharing of food. So where did Indo-European originate? Not surprisingly, scholars disagree on where and when the first speakers of Proto-Indo-European lived.

**NOMADIC WARRIOR HYPOTHESIS.** 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 people to domesticate horses and cattle, they migrated in search of grasslands for their animals. This took them westward through Europe, eastward to Siberia, and southeastward 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 5-18).

**SEDENTARY FARMER HYPOTHESIS.** 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 5-19). Biologist Russell D. Gray supports the Renfrew position but dates the first speakers even earlier, at around 6700 B.C.

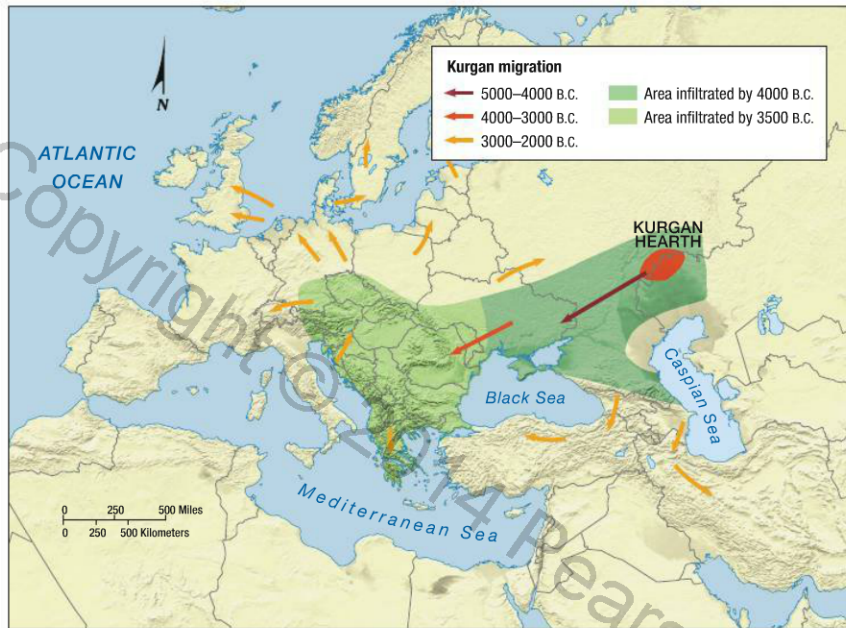
Renfrew believes they diffused from Anatolia westward to Greece (the origin of the Greek language branch) and from Greece westward toward Italy, Sicily, Corsica, the Mediterranean coast of France, Spain, and Portugal (the origin of the Romance language branch). From the Mediterranean coast, the speakers 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).

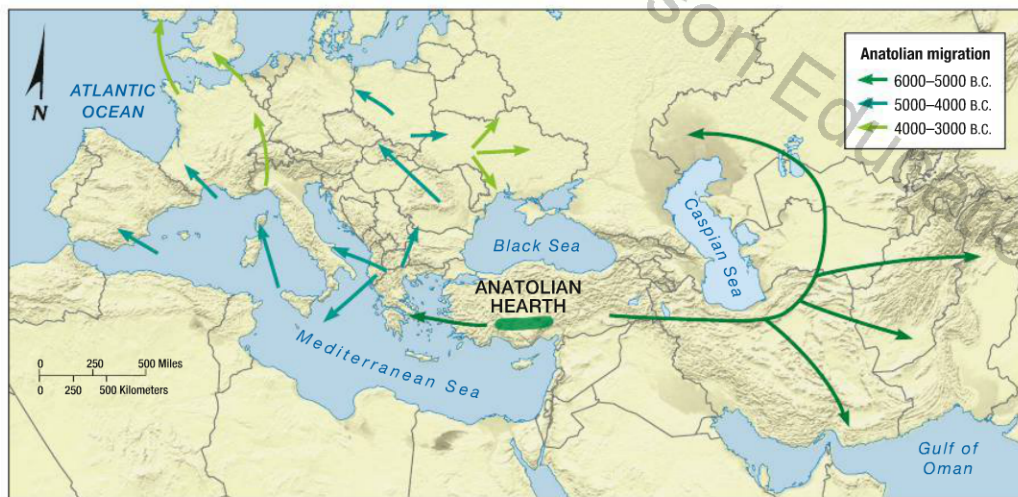
The Indo-Iranian branch of the Indo-European language family originated either directly through migration from Anatolia along the

▼ **FIGURE 5-17 A ROMANCE LANGUAGE: PORTUGUESE** The Museum of Portuguese Language in São Paulo, Brazil, has exhibits related to the Portuguese language, such as authors who have written in Portuguese.





**FIGURE 5-18 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 Southwest Asia beginning around 2500 B.C.



**FIGURE 5-19 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.

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.

#### Pause and Reflect 5.2.4

Which hypothesis appeals more to you: the “war” or the “peace” hypothesis? Why?

#### CHECK-IN: KEY ISSUE 2

##### Why Is English Related to Other Languages?

- ✓ The Indo-European family has four widely spoken branches.
- ✓ Individual languages, such as English and languages of the Romance branch, have documented places of origin and patterns of diffusion.
- ✓ The origin and early diffusion of language families such as Indo-European is speculative because these language families existed before recorded history.

## KEY ISSUE 3

## Why Do Individual Languages Vary among Places?

- Dialects of English
- Distinguishing between Languages and Dialects

### Learning Outcome 5.3.1

Describe the main dialects in the United States.

A **dialect** is a regional variation of a language distinguished by distinctive vocabulary, spelling, and pronunciation. Generally, speakers of one dialect can understand speakers of another dialect. Geographers are especially interested in differences in dialects because they reflect distinctive features of the environments in which groups live.

The distribution of dialects is documented through the study of particular words. Every word that is not used nationally has some geographic extent within the country and therefore has boundaries. Such a word-usage boundary, known as an **isogloss**, can be constructed for each word. Isoglosses are determined by collecting data directly from people, particularly natives of rural areas. People are shown pictures to identify or are given sentences to complete with a particular word. Although every word has a unique isogloss, boundary lines of different words coalesce in some locations to form regions.

## Dialects of English

When speakers of a language migrate to other locations, various dialects of that language may develop. This was the case with the migration of English speakers to North America several hundred years ago. Because of its large number of speakers and widespread distribution, English has an especially large number of dialects. North Americans are well aware that they speak English differently from the British, not to mention people living in India, Pakistan, Australia, and other English-speaking countries. Further, English varies by regions within individual countries. In both the United States and England, northerners sound different from southerners.

## DIALECTS IN THE UNITED STATES

Major differences in U.S. dialects originated because of differences in dialects among the original settlers. The English dialect spoken by the first colonists, who arrived in

the seventeenth century, determined the future speech patterns for their communities because later immigrants adopted the language used in their new homes when they arrived. The language may have been modified somewhat by the new arrivals, but the distinctive elements brought over by the original settlers continued to dominate.

**SETTLEMENT IN THE EAST.** The original American settlements stretched along the Atlantic Coast in 13 separate colonies. The settlements can be grouped into three dialect regions (Figure 5-20):

- **New England.** These colonies were established and inhabited almost entirely by settlers from England. Two-thirds of the New England colonists were Puritans from East Anglia in southeastern England, and only a few came from the north of England.
- **Southeastern.** About half came from southeastern England, although they represented a diversity of



▲ **FIGURE 5-20 DIALECTS IN THE EASTERN UNITED STATES** The most comprehensive classification of dialects in the United States was made by Hans Kurath in 1949. He found the greatest diversity of dialects in the eastern part of the country, especially in vocabulary used on farms. Kurath divided the eastern United States into three major dialect regions—Northern, Midlands, and Southern—each of which contained a number of important subareas. Compare this to the map of source areas of U.S. house types (Figure 4-25). As Americans migrated west, they took with them distinctive house types as well as distinctive dialects.