

## 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.

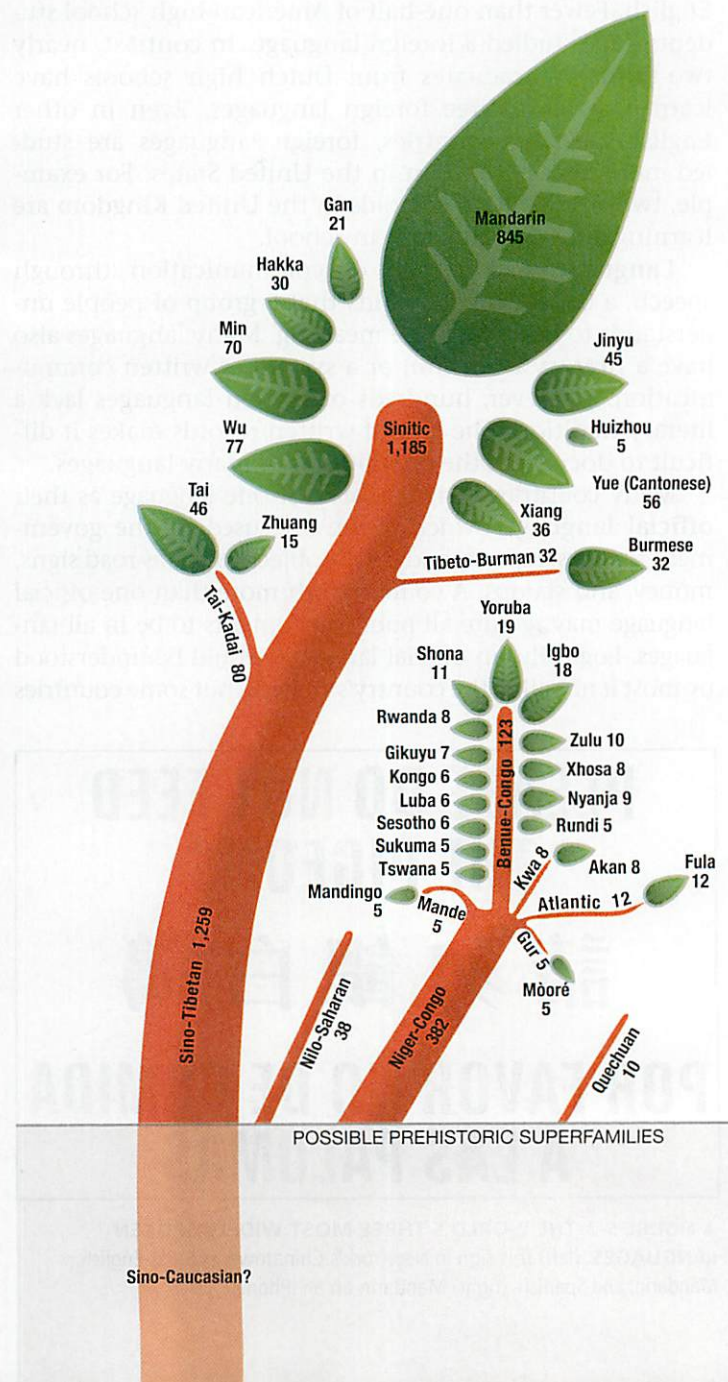
Two-thirds of the people in the world speak a language that belongs to the Indo-European or Sino-Tibetan language family. Seven other language families are used by between 2 and 6 percent of the world (Figure 5-4). The remaining 5 percent of the world's people speak a language belonging to one of 100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

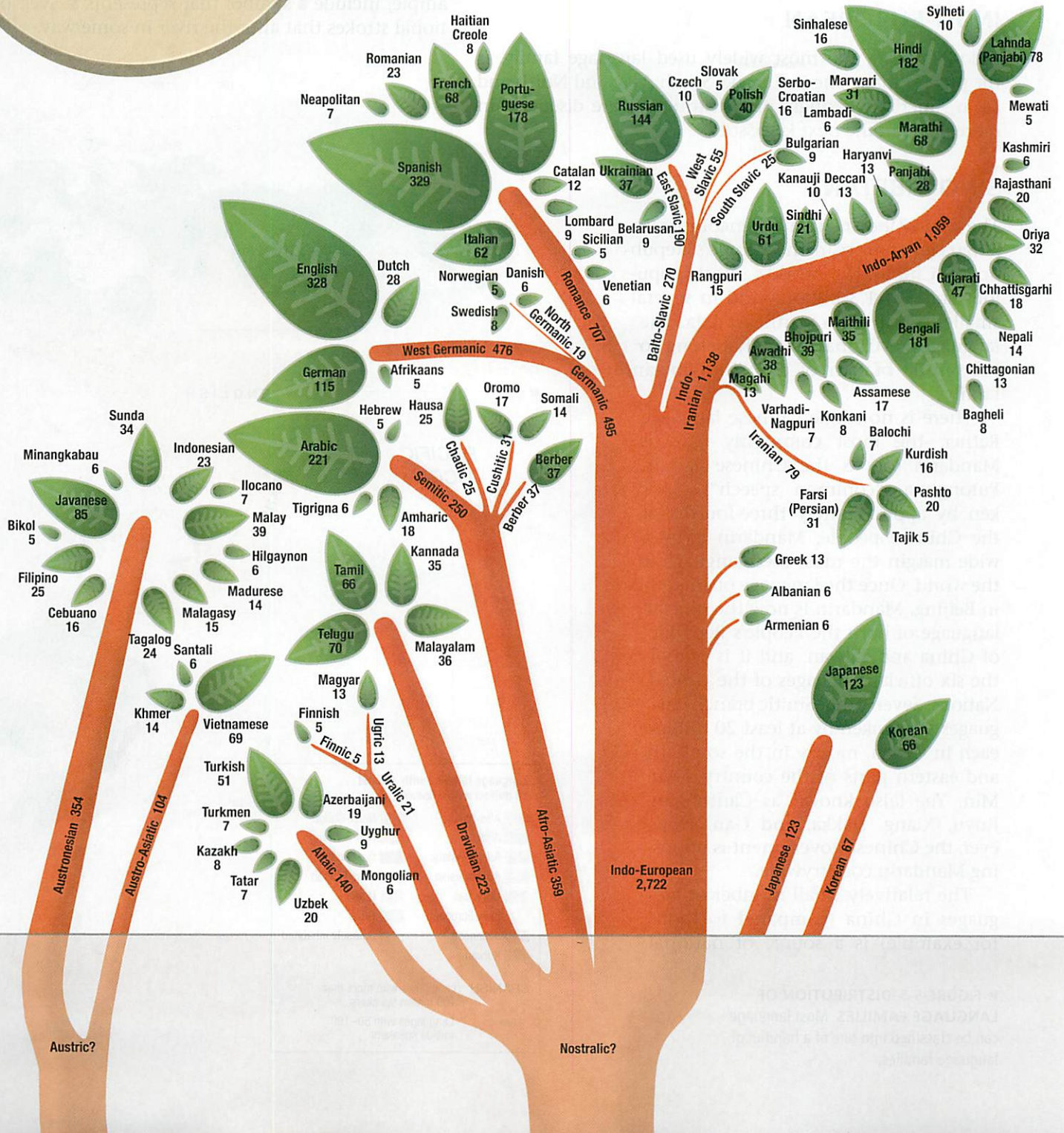
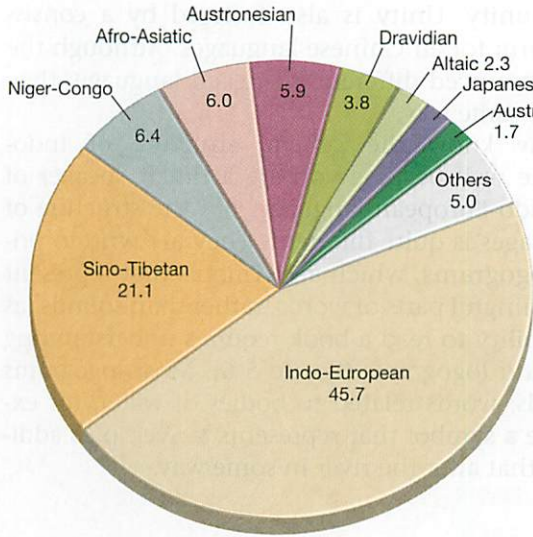
### Pause and Reflect 5.1.1

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.



◀ **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

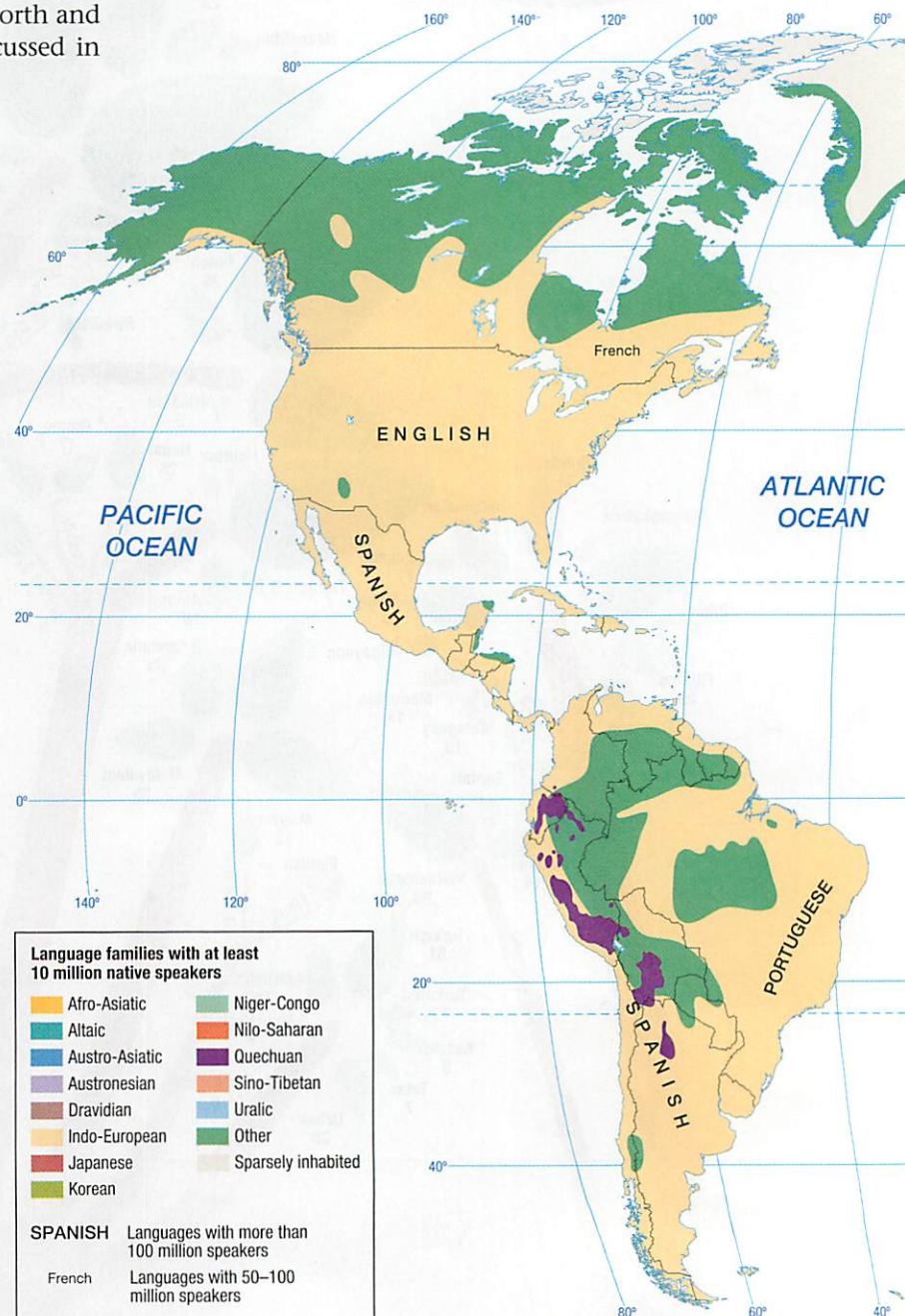
The Sino-Tibetan family encompasses languages spoken in the People's Republic of China—the world's most populous state, at 1.3 billion—and in several smaller countries in Southeast Asia. The languages of China generally belong to the Sinitic branch of the Sino-Tibetan family.

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.

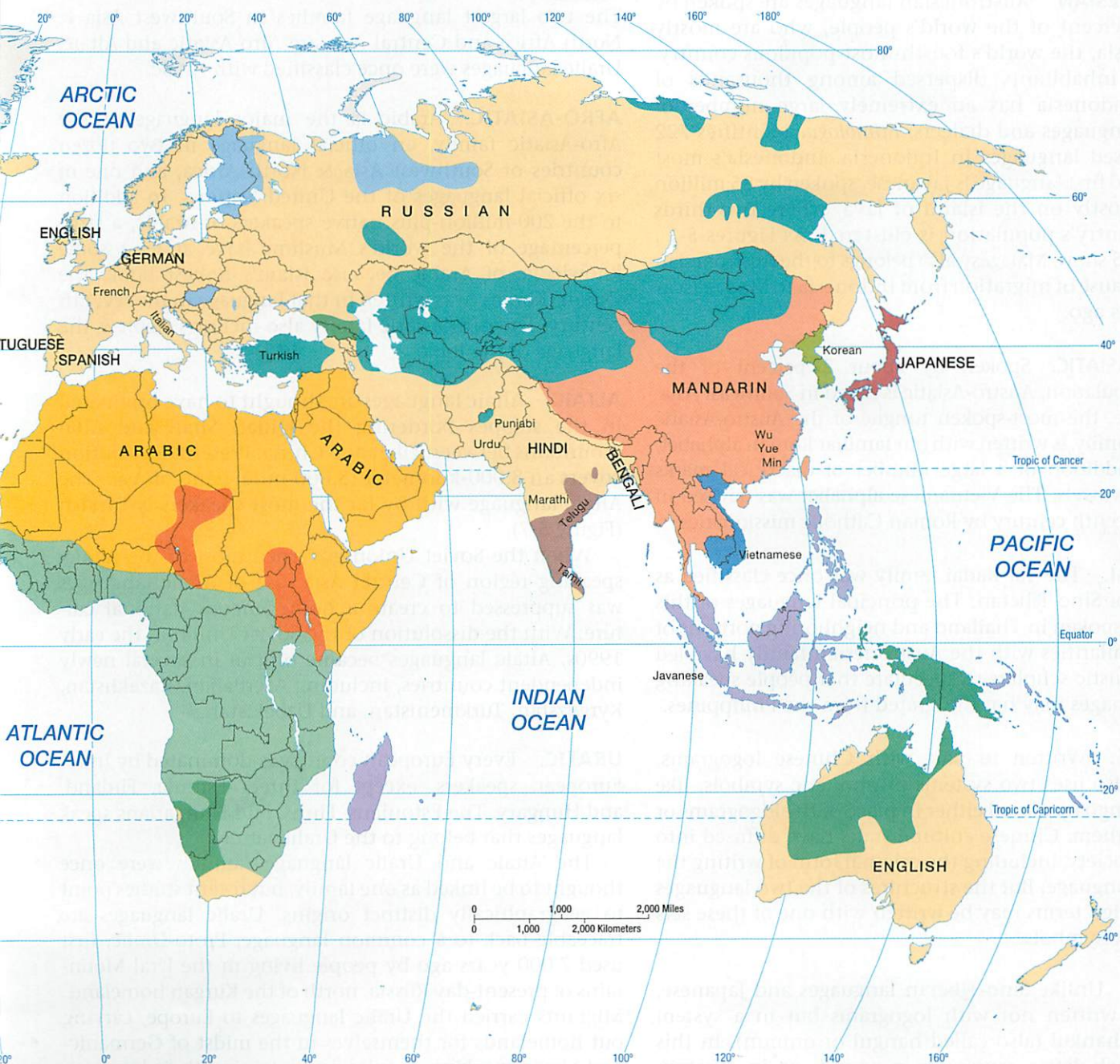
You already know the general structure of Indo-European quite well because you are a fluent speaker of at least one Indo-European language. But the structure of Chinese languages is quite different. They are written primarily with **logograms**, which are symbols that represent words, or meaningful parts 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.



## 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.

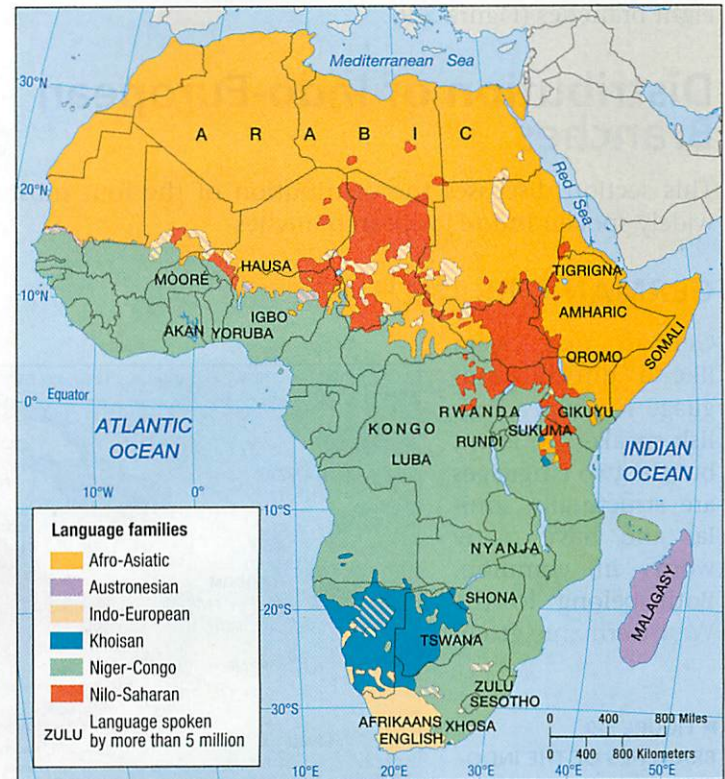
▼ **FIGURE 5-7 TURKISH** In 1928, Turkey's leader Kemal Ataturk ordered Turks to write Turkish with Roman letters instead of Arabic. Ataturk believed that using Roman letters would help modernize Turkey's economy and culture through increased communications with European countries. This painting depicts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



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.