

1

Languages and Linguistics



What Do You Think?

- Rob and Rod, college roommates who'll argue over almost anything, are debating the number of languages in the world. Rob says thousands. Rod says there's no way to count 'em. What do you say?
- Lauren in Los Angeles opens her utility bill and says with alarm as she glances at an insert, "Look at this! Five different languages! Spanish and Chinese and who knows what else! Isn't English the official language of the USA?" Is it?
- Reading a magazine, sixth-grade Sydney looks up and asks what *note* means. Figuring she knows its meaning in expressions like *thank-you note* and *musical note*, you ask her to read the sentence aloud. She reads it, and you say it means 'paper money' or 'bill,' as in "\$20 bill." She asks how hearing the whole sentence helped you. Your explanation?
- Fifteen-year-old Felix is teasing seven-year-old Seth at a family picnic and asks, "Do you know when your birthday is?" When Seth says, "May ninth," Felix retorts, "I didn't ask *when* your birthday is, dude! I asked if you *knew* when it is!" What does Seth understand about the question that Felix pretends not to know?

How Many Languages Are There?

Some dictionaries include language names among their entries, and you've probably seen lists that provide information about the number of speakers of various languages. When the U.S. Census Bureau compiles its data, it records which languages residents speak and makes that information publicly available. At the United Nations, most countries are represented, and their ambassadors would likely know which languages are spoken in their home countries. With all that information, it should be easy to answer the question, "How many languages are there in the world?"

Actually, enumerating the languages of the world is not a straightforward task. First, it's not always clear whether to call two language varieties *dialects* of one language or *different languages* altogether. Then, too, languages previously unknown to outsiders are sometimes discovered in the Amazon, Papua New Guinea, and other remote parts of the world. Some compilations of languages may be limited to spoken tongues, while others include signed languages. Finally, languages die when their last speaker dies, and that happens more often than you'd think.

Even when criteria are established for inclusion on a language list, compiling the information may be tough. For one thing, a given language may have different names, as with *Hebrew/Ivrit* and *Irish/Erse/Gaëlge/Irish Gaelic*. For another, a name may be spelled in different ways. One language spoken mainly in China (but not related to Chinese) can be spelled *Uyghur*, *Uighur*, *Uighar*, *Uygur*, *Uigur*, *Uighuir*, *Uiguir*, *Weiwuer*, and *Wiga*. And that's if you're using the Roman alphabet! By its speakers, Uyghur (we had to choose one!) is spelled with Arabic script, and it is also sometimes represented in Cyrillic letters or Chinese characters.

In the course of a century, many languages die and a few may be born. Occasionally, a dead language may be revived, as Hebrew has been. Similarly, the last speaker of Cornish, a Celtic language, died in 1777, but in the southwest of England it is being revived and is in use now among several hundred speakers, including some native speakers under 20 years of age. Manx, also Celtic and formerly spoken on the Isle of Man in the Irish Sea, is now extinct as a first language, but some second-language speakers are endeavoring to revive it. In Worcester, Massachusetts, Red Thunder Cloud died in 1996—and with him died Catawba, a Siouan language. In 2010 Boa Sr., the last native speaker of Bo, died in the Andaman Islands in the Bay of Bengal, and with her died her language. It is estimated that one natural language dies every couple of weeks.

New languages can also be born. Pidgins are spoken in some places as a second language, usually for limited functions such as bartering or commerce between speakers of different languages, but when circumstances in those places allow children to acquire a pidgin as their first language it will undergo a process called **creolization**, through which it develops into a full-blown language. Creoles such as Nigerian Pidgin and Saramaccan must be counted among the world's languages (even when their speakers call them pidgins).

One useful source of information, *The Ethnologue*, lists 7,105 languages. But don't take that number to be exact. Consider that in this book we sometimes refer to "Chinese" and that the U.S. Census Bureau allows residents to identify themselves as speaking "Chinese," whereas *The Ethnologue* cites Chinese

only as a “macrolanguage,” with 13 distinct language members carrying names such as Hakka Chinese, Mandarin Chinese, Wu Chinese, Xiang Chinese, and Yue Chinese, each of which may have dialects of its own. In the English-speaking world, Mandarin Chinese is known simply as Mandarin and Yue Chinese as Cantonese. Included among the *Ethnologue*’s 7,105 languages are some sign languages. Except for their channel of expression, sign languages are like spoken languages and share with them the challenges of being identified and counted. In the city of Chiangmai in Thailand, Chiangmai Sign Language is known only among older signers in the deaf community, while younger signers use a distinct language called Thai Sign Language.

It seems safe to stick with the conventional wisdom that there are about 7,000 languages in use in the world. Of those, only six—Arabic, Chinese, English, French, Russian, and Spanish—have official status at the United Nations, and only five of those rank among the top 10 languages in the world in number of speakers. French does not. On the other hand, Hindi, Bengali, Portuguese, German, and Japanese have greater numbers of speakers than some official U.N. languages. The U.S. Census Bureau identified well over 300 individual languages in use in the United States between 2006 and 2008. The *New York Times* reported in 2010 that New York City’s public schools were home to speakers of 176 languages and that residents of Queens, one of the city’s five boroughs, spoke 138 languages. In all, the Big Apple was home to some 800 languages. Given that the U.S. Census Bureau reports over 300 and the *New York Times* about 800, the difficulty of counting languages becomes apparent. Consider, too, that *The Ethnologue* lists over 40,000 primary names, alternate names, and dialect names for language varieties. Determining an exact number of languages in the entire world is a significant challenge!

Does the United States Have an Official Language?

Many people, including many U.S. residents, think the United States has English as its official language. It does not. In fact, the United States does *not* have an official language—and has never had one. Like the nation as a whole, some states have no official language; other states designate English as their official language; the only state with more than one official language is Hawaii, where English and Hawaiian are both official.

Some Americans also tend to think of the United States as essentially a monolingual nation, albeit with large numbers of Spanish speakers in the Southwest, Southeast, and Northeast. Actually, 55 million U.S. residents 5 years old or older speak a home language other than English (that’s almost 20 percent). Thirty-four million of those speak Spanish, and more than half of those Spanish speakers report that they also speak English very well. In seven heavily populated states at least one of every four residents over the age of 5 speaks a home language other than English. All 50 states have speakers of Arabic, Hindi, Hungarian, Korean, Tagalog, Thai, Urdu, and Vietnamese. All 50 states also have Native American speakers of indigenous North American languages. Navajo, with about 170,000 speakers,



© Edward Finegan

Los Angeles, California. Polling place notice posted in seven languages: English, Chinese, Japanese, Korean, Spanish, Tagalog, and Vietnamese.

is spoken in homes in New Mexico, Arizona, and several other states. According to figures in *The Ethnologue*, 176 living languages are used in the United States today. They range from Achumawi, Alabama, and Louisiana Creole French to Hawaiian, Lakota, Maricopa, Uyghur, Vietnamese, and Yatzachi Zapoteco. The California Department of Motor Vehicles administers its driver's license written examination in no fewer than 32 languages, including Amharic, Armenian, Cambodian, Hindi, Hmong, Hungarian, Indonesian, Laotian, Persian, Punjabi, Samoan, Tongan, and Turkish. At the 2010 U.S. Census website, instructions for filling out the census form were available in 59 languages, including Cebuano, Chamorro, Chuukese, Dinka, Haitian Creole, Marshallese, Navajo, Nepali, Swahili, Tamil, and Telugu. Wherever useful, election ballots are printed in multiple languages.

Still, the linguistic richness of the United States is not stable. The survival of most Native American languages is threatened. Because speakers tend to be older and because insufficient resources support these heritage languages, they are yielding to English among younger Native Americans. After English, no language spoken in the United States comes close to Spanish in number of speakers. Its 34.5 million U.S. speakers far exceed the 2.5 million who speak Chinese

and the 2.0 million who speak French, the two next popular. With few exceptions the children or grandchildren of immigrants cannot comfortably speak or readily understand the language of their grandparents, and this is true even of Spanish. Moreover, for all the richness of languages other than English throughout the United States, 225 million U.S. residents above the age of 5 speak only English at home—that's a whopping 80 percent.



Try It Yourself Using your knowledge about current and past immigration patterns, identify the eight most popular non-English languages spoken among U.S. residents aged 5 and older. Spanish, Chinese, and French fill the first three slots, and Italian and Arabic are ranked ninth and tenth. Which languages are in between?

What Is Human Language?

The modern study of language is rooted in questions first raised millennia ago. As old as speculation on any subject, inquiry into the nature of language occupied Plato and Aristotle and other Greek philosophers, as well philosophers from India. In some areas of grammatical analysis, the ancients made contributions that have remained useful for 2,000 years and established some of the categories still used today. In the nineteenth and twentieth centuries,

the field of linguistics emerged to address certain age-old questions, including these:

- ◆ What is the nature of the relationship between signs and what they signify?
- ◆ What are the basic elements of a language?
- ◆ How are the basic elements of language organized into words, sentences, and discourse?
- ◆ What enables us to produce and understand countless sentences never encountered before?
- ◆ How do languages achieve their communicative goals?
- ◆ What is the origin of language?
- ◆ In what ways do languages change and develop?
- ◆ What does it mean to say two languages are related?
- ◆ What is the relationship between a language and a dialect?
- ◆ What enables a young child to learn a language so efficiently?
- ◆ What makes it so challenging for an adult to learn a language?
- ◆ Are there right and wrong ways to express things, and—if so—who decides?

This book provides a modern context for asking and addressing these and other questions about language.

Three Faces of a Language System

The fundamental function of every language system is to link meaning and expression—to provide verbal expression for thought and feeling and for that expression to be comprehensible to others. A grammar can be viewed as a coin whose two sides are *expression* and *meaning* and whose function is to systematically link the two. But language is not simply a two-sided coin; it has a third face so important in producing and interpreting utterances that it can override almost all else. That face is *context*, and only in context can an expression convey a speaker's intended meaning and be correctly interpreted by a hearer.

Imagine a New England dinner-table conversation about, say, the cost of living. In the course of the conversation, a guest asks the host, "Is there a state income tax in Connecticut?" The question could elicit replies of "Yes," "No," or "I don't know" because in this context it is likely to be understood as a request for information. But consider an equally straightforward inquiry made of the host on the same occasion: "Is there any salt on the table?" To this question earnest replies of "Yes," "No," or "I don't know" would mark the host as inconsiderate.

Is there a state income tax in Connecticut?

Is there any salt on the table?

The form of the salt question resembles the form of the income tax question, but the *point* of the questions—their *intended* meaning—and the expected responses could hardly be more different. At a dinner table, a guest inquiring about salt expects a host to recognize that it's *salt* that's wanted, not *information*! By contrast, in a related context—with the host in the kitchen, pepper

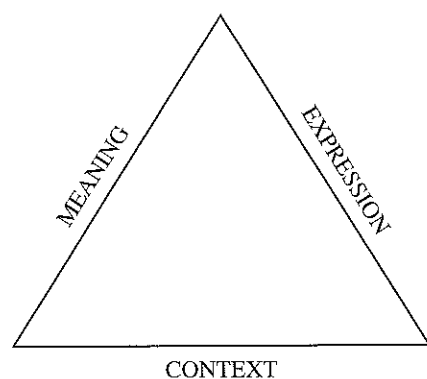


Figure 1.1 Three Faces of Language

mill in hand, and facing a guest who's just come from the dining room—the question, “Is there any salt on the table?” is likely to be understood as seeking information even though *the form of the question* is exactly the same as the one asked by the guest at the table. In answer to the question asked in the dining room, a reply of “Yes” or “No” or “I don’t know” would seem bizarre. In the kitchen, it would be altogether appropriate.

You can see, then, that conversationalists cannot interpret the point of an utterance from expression alone. To grasp the intended meaning of an expression, hearers must consider it *in its context*. Likewise, speakers routinely rely on a hearer’s ability to recognize their intentions in uttering an expression in a specific context.

Besides meaning and expression, then, at the base of language use is context, and language can be better viewed as a three-sided triangle comprising expression, meaning, and context, as shown in Figure 1.1.

Expression encompasses words, phrases, and sentences, including intonation and stress. **Meaning** refers to the senses and referents of those elements of expression. **Context** refers to the social situation in which expression is uttered and includes whatever has been expressed earlier in that situation and any shared knowledge between speaker and hearer. What links expression and meaning is grammar. What links grammar and interpretation is context. Without attention to both grammar and context, language utterances cannot be understood.

Language: Mental and Social

Language is often viewed as a vehicle of thought, a system of expression that mediates the transfer of thought from one person to another. It does—and in everyday life it also serves equally important social and emotional functions.

Linguists are interested in models of how language is organized in the mind and how it is shaped by the social structures of human communities, reflecting those structures in expression and interpretation.

Signs: Arbitrary and Nonarbitrary

In everyday conversation we talk about signs—*signs* of trouble with the economy, no *sign* of a train arriving at a railway station, someone’s vital *signs*, and so on. **Signs** are indicators of something else. In the examples just mentioned, the indicator is inherently related to the thing indicated, and such nonarbitrary signs have a direct, often causal relationship to the things they indicate. Smoke is a nonarbitrary sign of fire, clouds a nonarbitrary sign of impending rain.

Arbitrary Signs

Nonarbitrary signs such as clouds and smoke differ from partly or wholly arbitrary signs. With arbitrary signs like traffic lights, railroad crossing indicators, wedding rings, and national flags, no causal or inherent connection exists between the sign and what it signifies. Arbitrary indicators can be present even when the thing indicated is absent (as with a bachelor wearing a wedding ring). Because they are conventional representations, arbitrary signs can be changed. If a national transportation department decided to use the color blue as the signal to stop traffic, it could do so. Or it could use red or, for that matter, green. The relationship between words and what they signify is generally arbitrary, and we say that language is a system of *arbitrary* signs.

Representational Signs

Some essentially arbitrary signs are not entirely arbitrary and may suggest their meaning. Poison may be suggested by a skull and crossbones ☠, while an icon such as ☀ may suggest the sun, and the roman numerals II and III, with two and three strokes respectively, represent the numbers two and three. Because these signs suggest what they indicate, they are partly iconic. Signs that are basically arbitrary but partly iconic are called **representational**. Linguistic examples include the English *meow* and *boom*, insofar as those words suggest what they signify, but even words that mimic natural noises are cross-linguistically different.

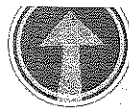
Iconic expression can also appear spontaneously in ordinary speech. I once telephoned the home of a friend, and her 4-year-old son answered. He reported that his mother was showering, and when I said I'd call back in a few minutes he indicated that calling back soon would do no good. His explanation was this: *My mother is taking a long, loong, looong shower*. By stretching out his pronunciation of the vowel sound in *long*, the boy demonstrated the potential for iconicity in human language. By making his vowels longer, he directly signaled length of time and thus iconically emphasized the salient part of his meaning. Representational language (also called iconic) is expression that in any fashion mimics or directly suggests its content.

Iconicity can also be expressed in grammar. Consider that English has two ways of organizing conditional sentences—such as those with *if*. The condition (the “if” part) can precede the consequence or follow it.

If you behave, I'll give you some M&Ms. (condition precedes consequence)

I'll give you some M&Ms if you behave. (condition follows consequence)

English permits placing the condition (*if you behave*) before or after the consequence (*I'll give you some M&Ms*). While contextual factors may influence the



Try It Yourself Besides the boy's stretching out the vowel in *long* to represent length of time, identify a second way in which “My mother is taking a long, loong, looong shower” is iconic. Now, identify another very, very common example in English in which the second way conveys a meaning different from extended length. (Hint: Examine the preceding sentence attentively.)

choice, speakers and writers show a strong preference for the condition to precede the consequence, a preference shared with many languages. The reason has to do with the order of occurrence of real-world events described by conditional sentences. In our example the addressee must *first* behave; *only after that* will the speaker provide some M&M candies. These real-world events are ordered in time, and the real-world order is reflected in the preferred linguistic order of condition before consequence. With condition before consequence, the expression mimics the sequencing of real-world events. Some languages allow *only* the condition-before-consequence pattern; others permit both patterns. But no language appears to limit conditional sentences to the non-iconic order of consequence before condition.

Language—A System of Arbitrary Signs

Despite occasional iconic characteristics, human language is essentially arbitrary. Except for the associations established by convention, the form of an expression is generally independent of its meaning. Imagine a parent trying to catch a few minutes of the televised evening news while preparing dinner. Suddenly a strong aroma of, say, burning rice wafts into the TV room. This *nonarbitrary sign* will send the parent scurrying to salvage dinner. The aroma is *caused* by the burning rice and would convey its message to speakers of any language. There is nothing conventionalized about the message. Now consider the words of a youngster in the kitchen who shouts, “The rice is burning!” That utterance is just as likely to send the parent hurtling to the kitchen, but the words are arbitrary. It is a set of facts about *English* (not about burning rice) that enables the utterance to alert the parent. The utterance is thus an arbitrary sign.

Other languages express the same meaning differently: Korean by *pap t^handa*, Swahili by *wali inaunguwi*, Arabic by *yahtariqu aluzzu*. The forms of these utterances are not iconic and have nothing to do with rice or the manner in which it is cooking. Instead, they reflect only the language systems of Korean, Swahili, or Arabic. As you see, a central characteristic of human language is that the connection between words and what they mean—between signifier and signified—is largely arbitrary.

Languages as Patterned Structures

Grammatical Competence

Given the arbitrary relationship between linguistic signs and what they represent, languages must be highly organized systems in order to function as reliable means of communication.

What underlies the observable patterns that languages follow are called “rules.” Not rules imposed from the outside like traffic regulations, linguistic rules do not specify how something *should* be done. Instead, the rules described in this book are based on the observed regularities of language behavior and the underlying systems that can be inferred from that behavior and

certain kinds of intuitions. They are the rules that even very young children acquire unconsciously and put to use when they display mastery of their native tongue.

A language is a set of elements and a system of rules for combining those elements into patterned expressions that serve to accomplish specific tasks in specific contexts. Utterances report news, greet relatives, invite friends to lunch, request the time of day from strangers. With language, we make wisecracks, poke fun, express admiration, propose marriage, create fictional worlds, argue for a course of action, and so on. And a language accomplishes its work with a finite system that a child masters in a few years. The mental capacity that enables speakers to form grammatical sentences such as *My mother is taking a long shower* rather than “A taking long my shower is mother” (or thousands of other possible ill-formed strings of those seven words) is **grammatical competence**. It enables speakers to produce and understand an infinite number of sentences they haven’t heard before. Besides arbitrariness, then, four other hallmarks of human language systems deserve highlighting.

Discreteness

Speakers can identify the sound segments in the words of their language. Most speakers of English would agree that the words *sat*, *pat*, and *kid* have three sounds each, while *spat*, *pats*, and *kids* have four each, and *spats* and *skids* five each. Likewise for the three sounds in *then*: the initial consonant sound represented by *th*, the vowel sound represented by *e*, and the final sound represented by *n*. It is a structural feature of language that words are made up of discrete (separate) elemental sounds.

Duality

Human languages can be analyzed on two levels. At one level, a language can be viewed as having meaningful parts. For example, *tabletop* has *table* and *top* as meaningful parts. At a lower level the elements that make up the meaningful parts do not themselves carry meaning. The sounds of *top* don’t individually mean anything but form a meaningful unit only when combined in a particular order. Precisely because the individual sounds in *top* don’t carry meaning, they can be formed into other combinations with different meanings, as in *pot*, *opt*, *topped*, and *popped*. This two-level analysis—of meaningless elements combined into meaningful ones—is what we call the *duality* hallmark.

Displacement

Human languages are capable of representing things and events that are not present but are spatially or temporally distant. Speakers aren’t confined to discussing events of the here and now but can talk of faraway places and events of yesterday, yesteryear, and even yet—or never—to occur. We call this feature **displacement**.

Recursion and Productivity

Relying on relatively few elements and rules in a language system, humans can produce and understand a limitless number of sentences. A single week's issue of a magazine is not likely to repeat any sentences (other than a quoted one), and the same is true over months and years of publication. The human capacity for linguistic inventiveness makes repeated sentences unlikely, and a literate speaker of any language is capable of understanding countless sentences in a lifetime of reading. Consider that you can add "I said" to the beginning of any sentence, no matter how long it already is. That property of incorporating structures within similar linguistic structures is known as **recursion**, and recursion underlies linguistic **productivity**.

Consider the sentence below, a more complex example than "I said . . .":

- ◆ It's a weight-loss program THAT PEOPLE CAN SUCCEED AT.

That sentence contains a relative clause (in small caps for visibility). For now, think of a relative clause as a sentence-like structure that modifies a noun phrase. Here, THAT PEOPLE CAN SUCCEED AT modifies "a weight-loss program." We can go further. The relative clause THAT PEOPLE CAN SUCCEED AT could itself contain a relative clause.

- ◆ It's a weight-loss program [THAT PEOPLE [*who can't easily stick to a diet*] CAN SUCCEED AT].

The relative clause *who can't easily stick to a diet* modifies PEOPLE—and it too could contain a relative clause:

- ◆ It's a weight-loss program [THAT PEOPLE [*who can't easily stick to a diet* [*that requires giving up chocolate*]] CAN SUCCEED AT].

In that sentence the relative clause that requires giving up chocolate modifies *a diet*.

In principle, recursion allows structures to be incorporated into structures of the same kind without limit. We say "in principle" because human memory limits the number of relative clauses that we can actually process in a sentence, but memory is a separate issue altogether. Some researchers view recursion—the ability to generate and understand an infinite number of sentences by combining and recombining the same few elements and structures—as the singular hallmark of human language, the characteristic that theories about language must centrally address.

Speech as Patterned Language Use: Communicative Competence

Knowing the elements of a language and the rules for putting them together into well-formed sentences still falls short of knowing how to accomplish the work that speakers accomplish with language. That certainly requires mastery of grammatical rules, but it also requires competence in the appropriate *use* of the

sentences produced by the rules. Among other things, it requires knowing how to link sentences appropriately in conversations and how to rely appropriately on context to shape utterances and interpret them.

The capacity to use language appropriately is called **communicative competence**. It enables us to weave utterances together into apologies, requests, directions, recipes, sermons, scoldings, jokes, prayers, stories, and all the other things we do with language. Being a fluent speaker requires communicative competence and grammatical competence.

Grammatical competence is the language user's unconscious, or implicit, knowledge of vocabulary, pronunciation, sentence structure, and meaning. *Communicative competence* is the implicit knowledge that underlies the appropriate use of grammatical competence in communicative situations. Because the patterns that govern the appropriate use of language differ from one speech community to the next, even a shared grammatical competence in a language such as English may not be adequate to make you a fluent speaker across other English-speaking communities. For example, members of one culture may find jokes about other people's misadventures funny, whereas members of another culture may find them offensive. In fact, the very concept of telling jokes (*Did you hear the one about . . . ?*) as distinct from telling funny stories seems not to exist in certain societies. Likewise, what is considered impolite in one community might be routine interaction in another community. Differences in interactional customs explain why even some American visitors to the Big Apple judge New Yorkers brusque or impolite when giving directions, though the same directions may be interpreted by a fellow New Yorker as routine or routinely polite.

Languages and Dialects

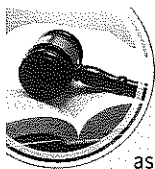
Along with physical appearance and cultural characteristics, language contributes to defining nationality. But even within a nation's borders, people may speak different languages. In Quebec, ethnic French-Canadians maintain allegiance to the French language, while ethnic Anglos maintain loyalty to English. Citizens of Switzerland may speak French, German, Italian, or Romansch. Across India, scores and scores of languages are spoken, some confined to villages, others used regionally or nationally. In Papua New Guinea one may hear hundreds of languages, and an English-based language called Tok Pisin is used for communication across groups.

Wherever speakers of a language become separated by geographical or social distance, linguistic variation is likely to arise. Striking differences can be noted between the French varieties of Quebec and Paris, between the Spanish varieties of Madrid and Mexico City, and between the English varieties of Sydney, London, Dublin, and Chicago. Every language variety reflects the social identity of its speakers.



Try It Yourself Identify a characteristic of your *pronunciation* that others have commented about when you've traveled outside your region. What about a *vocabulary* item that others have found unusual? Do the same for a roommate or classmate whose speech differs from yours in pronunciation or vocabulary.

Some people seem to believe that only *other people* speak a **dialect** but that *they themselves* don't. Instead, they think of themselves as speaking a *language* or even *the language*. The truth is, everyone speaks a dialect and speaking a dialect means speaking a language. Australian English, British English, Irish English, and American English are national dialects, and speakers of those national dialects also speak a regional dialect. Anyone who speaks a dialect of English speaks the English language, and anyone who speaks the English language does so only by using one of its dialects. When we speak of a "language"—Chinese, Spanish, Arabic, Navajo, or any other—we are speaking of a set of dialects. Languages do not exist independently of their dialects.



AT THE BAR Jack the Ripper

Linguistic experts have long served as consultants to law courts, defense and prosecution attorneys, police investigators, and a wide range of individuals and companies in matters at the intersection of language and the law. Conferences on "forensic linguistics" are now mounted regularly, and the field has its own journal. You know, at least from television programs, that ballistics specialists, psychiatrists, coroners, and others serve as expert consultants and may testify at trial. Linguists do the same, and as a regular feature of this book we will exemplify the work of linguists in legal matters both civil and criminal—in other words, at the bar! Here we tell the story of the hoaxer *Jack the Ripper*.

In the late 1970s Britain was gripped by a series of brutal murders of women in the West Yorkshire area. Assistant Chief Constable George Oldfield, the officer in charge of the investigation, was under pressure to find the murderer, and he thought he had the clue he needed when he received a cassette recording that opened with these words:

I'm Jack, I see you are still having no luck catching me. I have the greatest respect for you George, but, Lord, you are no nearer catching me now than four years ago when I started. I reckon your boys are letting you down, George, they can't be much good can they?

The tape's taunting sign-off was "Jack the Ripper," and Oldfield sought help from linguists at a nearby

university to identify Jack's dialect. After analyzing the recording and several letters judged to be sent by the same person, dialect geographer Stanley Ellis and phonetician Jack Windsor Lewis reported that the voice belonged to someone raised in the north-east of England, likely in or very near the town of Sunderland, some 75 miles from the murder sites in West Yorkshire. They also warned the police that the tape was likely a hoax and that the actual killer might well speak with a different accent. It was therefore unwise, they cautioned, to focus the investigation on speakers with a Sunderland accent. But Oldfield became convinced the killer was from Sunderland (where the cassette and letters had been mailed) and lavished investigative resources on that assumption. Concerned about continued killing, Windsor Lewis spoke to a reporter for a newspaper in the region of the murders, and in late 1980 a story appeared in the *Yorkshire Post*, under the headline VOICE EXPERT DISMISSES VITAL CLUE AS JUST A RED HERRING/RIPPER TAPE MADE BY SUPER-HOAXER, SAYS POLICE ADVISER. The reporter wrote, "If Mr. Windsor Lewis . . . is right, then hundreds of thousands of police man-hours have been wasted, particularly in Sunderland, searching for a Geordie-accented Ripper."

A few months later, Peter Sutcliffe—living in West Yorkshire, where he was born and raised—was caught and admitted the killings. All of Britain felt a sense of relief—but the police were determined to catch the hoaxer because with the police investigation focused on men with a Sunderland accent, Sutcliffe—a West Yorkshireman—felt safe

in murdering three more women. Gradually, over a quarter of a century, the hoaxer tape investigation went cold. Then, in 2005, a cold-case investigation matched DNA from one of the envelopes licked by the hoaxer to the DNA of a man named John Humble. Just as Ellis and Windsor Lewis had reported, Humble was born and raised within walking distance of the Sunderland locale they originally identified.

After Humble's arrest, two linguists were asked to carry out an acoustic analysis of the voices on the hoax tape and a new recording of Humble reading a transcript of the original hoax message. Using

spectrographic and other analyses, Peter French and Philip Harrison concluded that the voices were so similar that the likelihood of their belonging to different men was remote. In the face of that evidence, Humble's defense team consulted none other than Jack Windsor Lewis, one of the original two linguists, to see whether features of the original letters and of versions that John Humble willingly handwrote from dictation in jail were sufficiently different to point to different authors. Windsor Lewis concluded in the negative. Humble then confessed and was sentenced to prison for perverting the course of justice! ■

Different Dialects or Different Languages?

As is widely known, the Romance languages developed from regional dialects of Latin spoken in parts of the Roman empire. Those dialects eventually gave rise to Italian, French, Spanish, Portuguese, Romanian, now the distinct languages of independent nations. While these tongues share certain inherited features of grammar, pronunciation, and vocabulary, the nationalistic pride taken by Italian, French, Spanish, Portuguese, and Romanian speakers supports the view that they speak distinct languages rather than dialects of a single language. The opposite situation characterizes Chinese. Not all Chinese dialects are mutually intelligible (for example, monolingual speakers of Cantonese and Mandarin can't understand one another), but speakers regard themselves as sharing a single language and highlight that unity with a shared writing system.

Whether two varieties are regarded as dialects of one language or as distinct languages is a social as much as a linguistic matter, and the perception may be influenced by nationalistic and religious attitudes. Hindus in northern India speak Hindi, while Muslims there and in neighboring Pakistan speak Urdu. Opinions differ as to how well the groups understand one another. Until a few decades ago, Hindi and Urdu constituted a single language called Hindustani, and the fact that professional linguists wrote grammars of "Hindi-Urdu" reflects a judgment that the two varieties required only a single grammatical description. With the passage of time, of course, Hindi and Urdu—whose different names proclaim that their speakers choose to belong to different social, political, and religious groups—will become increasingly differentiated, as the Romance languages did.

Modes of Linguistic Communication

There are three basic **modes** of linguistic communication, corresponding to different modes of production and perception: oral communication, relying on the

use of speech and hearing organs; writing, a visual representation; and signing, a visual or tactile representation.

Speaking

The most common vehicle of linguistic communication is the voice, and speech is thus a primary mode of human language, with some advantages over other modes. Because it does not need to be viewed, speech can effectively accomplish its work in darkness or light, straight ahead or around corners. During the development of the human species, with hands and eyes occupied in hunting, fishing, and food gathering, speakers remained free to report, ask for and give directions, explain, promise, apologize, bargain, warn, and flirt.

Speaking has other advantages. For one thing, the human voice is complex. It has variable volume, pitch, rhythm, and speed; it is capable of wide-ranging modulation. Besides having a set of sounds, speech takes advantage of the organization of its sounds—their sequencing into words and sentences. Like writing and signing, speech can take advantage of word order and word choice. In its natural state, of course, speech evaporates and cannot span time, but modern technologies allow speech to be preserved indefinitely.

Writing

Long before the invention of writing, people painted stories on cave walls and exploited other visual signs to record events. Such *pictograms* were independent of language—a kind of cartoon world in which anyone with knowledge of the lives of people but without specific linguistic knowledge could reconstruct the depicted story. When shown to adult speakers, depicted stories can be told in any language. Pictograms (☠ ☀ ☺) can be understood by speakers of any language

because they are a direct, nonlinguistic symbolization, like a silent film or road signs used internationally to indicate a curved roadway or the availability of food and lodging. Of course some things are more difficult than others to convey using pictograms, as the photograph to the left indicates. Among icons common in email correspondence and able to be made with a standard keyboard are the emoticons :) and :(. But they are increasingly replaced by ☺ and ☹ , which can't be made directly with a standard keyboard.

If icons come to be associated not with the *entities* they represent but with the *words* referring to the entities, we have a more sophisticated system.



© Edward Finegan

Signs on a wall at the seaside, outside Dublin, Ireland, where fishing and nude bathing are forbidden. The fishing prohibition is communicated iconically. For obvious reasons, the prohibition of nude swimming relies on arbitrary linguistic signs.

Written representation becomes *linguistic* when it relies on language for its organization and communicative success. For example, while it is difficult to use pictograms to express a message about abstractions (such as hunger or fairness), the task becomes manageable if the graphic signs represent existing words. The moment some imaginative person first recognized that the written sign ☀ could represent not only the sun itself but also the word for 'sun' in his or her spoken language, the initial step was taken toward the development of writing. Basically, the same step is taken every time someone composes a license plate like "GR8D8B8" for an expensive sporty car, where 8 does not mean the number 8 but the sound of the number in English. Letters, too, can be used the same way and often are in texting, as with "4U QT" or "1CE" or "2EZ." Writing was invented about 5,000 years ago by ingenious people who chanced upon an occasion to use pictograms to represent spoken words instead of the objects they usually represented. Recent scholarship suggests that writing may have had multiple origins, and the social, economic, cultural, and historical circumstances of the invention in each case are important to an overall understanding of its origins.

Speech and writing are not related to the world they symbolize in the same way. Speech directly represents entities in the world—things such as the sun, the moon, her car, that book, fresh fish, and today's weather. Writing represents the physical world only indirectly. A written sentence such as *Ashley hates rainy weather* is a secondary symbolization: the written words represent the spoken words, not the entities and activities themselves.

We saw above that speech has many advantages over language expressed in other channels. Writing, however, does have certain advantages over speech. Although it generally takes longer to produce than speech, it can be read and understood much more quickly than speech. Writing (in correspondence or books or on cave walls or on touch screens) endures longer than nonrecorded speech, and publications can have a far greater reach. A message on a blackboard can be read after its author has left the room; not so for an unrecorded spoken utterance.

Signing

The third mode of linguistic communication is signing, the use of visible gestures to communicate. Speakers often use gestures and facial expressions to convey meaning in support of oral communication, but signing can be used as the sole means of accomplishing the work of language. American Sign Language, the principal sign language of North America, is used by somewhere between half a million and 2 million signers, not all of them deaf.

ASL, as it's called, and similar sign languages use manual signs and facial and bodily gestures, combining them under a system of grammatical rules to create an infinite number of sentences. The linguistic character of sign languages and the linguistic character of spoken languages such as Japanese and English are alike, except for the channel of expression, but sign languages like ASL are fundamentally independent of any spoken language.



Figure 1.2 ASL signs for SUMMER, UGLY, and DRY

Source: Lucas and Valli, 2004. "American Sign Language." In Edward Finegan & John R. Rickfort, eds. *Language in the USA* (Cambridge, UK: Cambridge University Press).

Manual signs in a language like ASL have three main components:

- ◆ *hand shape and orientation* (which fingers are open or closed and how the palm is facing)
- ◆ *hand location* with respect to the signer's body
- ◆ *hand movement*

Signs differ from one another in one or more of these components. As shown in Figure 1.2, the ASL signs SUMMER, UGLY, and DRY are identical in hand movement and in hand shape and orientation. For all three, the right hand, with index finger extended but the others closed, is drawn across and in front of the signer, the index finger closing in the process. The three signs are identical except in hand location. For SUMMER, the sign is drawn across the front of the signer's forehead; for UGLY, across the nose; for DRY, across the chin. Note that you cannot guess the meanings of these signs, and that would be true for most signs. But some other signs do show a resemblance to what they represent. For example, in ASL the sign TREE can be viewed as resembling a tree, and the sign CAT suggests a cat's whiskers, as you can see in Figure 1.3. As with spoken



Figure 1.3 ASL signs for TREE and CAT

Source: Lucas and Valli, 2004. "American Sign Language." In Edward Finegan & John R. Rickfort, eds. *Language in the USA* (Cambridge, UK: Cambridge University Press).

languages, then, sign languages like ASL are basically arbitrary, although some signs are representational, with meanings that can be guessed.

We've focused our discussion of sign language on vocabulary, but sign languages, like spoken languages, have grammars for combining vocabulary into phrases and sentences. Spoken languages don't construct sentences by stringing words together like beads on a string, and neither do sign languages. Both signed and spoken languages follow complex systems of grammatical rules, and the rules for ASL differ from the rules for American English in ways akin to those that differentiate the rules for German, say, from those for Spanish. Knowing even a long list of ASL signs and knowing English or French grammar, but not ASL grammar, would not allow a person to communicate effectively in ASL or understand a conversation between signers using ASL.

Given the way languages and dialects develop historically, it shouldn't be surprising that ASL has regional and ethnic varieties. What is not so obvious is that the dialects of sign language need not parallel spoken dialects in the same locations. Spoken American English, for example, derives chiefly from spoken British English, but American Sign Language does not derive from British Sign Language, and ASL and BSL aren't mutually intelligible. Thomas Hopkins Gallaudet, the founder of the first American school for the deaf, had studied French Sign Language in Paris and returned to Connecticut with a teacher of French Sign Language in 1816. As a consequence, ASL developed from a combination particularly of French Sign Language and a sign language used on Martha's Vineyard, in Massachusetts, at the time Gallaudet started his school.

The kind of signing used by Helen Keller differs fundamentally from ASL because Keller was blind as well as deaf. Her signing was a kind of manual spelling system and, unlike ASL, relied on an existing spoken language. Her signing consisted of spelling out words by "drawing" *with* the hands and *in* the hands the shape of written signs that are used to represent sounds (as with letters of the alphabet). Helen Keller's sign language depended not only on the prior existence of a spoken language but also on a written form of that language. Signing systems that rely on the modeling of letters are two steps removed from the linguistic system acquired by hearing and seeing children. To a small degree ASL also relies on finger spelling—for example, to represent people's names and other words that lack signs of their own. Unlike Helen Keller's manual system, the finger spelling of names and some other words in ASL is carried out to be seen rather than by drawing in the hand, which is designed to be felt.

In this book we focus on language as represented in spoken and written communication. Keep in mind that, both historically and developmentally, writing is a secondary mode of linguistic communication. Speaking is the primary mode. This priority can be a challenge to students, whose principal focus and context for discussing language in academic settings has been reading and writing.

Do Only Humans Have Language?

When you observe animals in groups, it's obvious that they interact—that they communicate with one another. Dogs display fangs to communicate displeasure or aggression; male frogs attract female frogs by croaking. It's only natural to

wonder how the forms of communication used by animals differ from human language.

We sometimes speak of porpoises, chimpanzees, gorillas, dolphins, whales, bees, and other animals as though they had communication systems similar to human language. Doubtless, all species of animals have developed systems of communication with which they can signal such things as danger and fear, and we know a good deal about how and what bees communicate. In recent decades, chimpanzees have been taught to use signs. But researchers have gradually lowered their expectations of animal language as we've better understood animal communication systems.

How Animals Communicate in Their Natural Environment

For a long time people wondered how bees were able to tell one another the exact location of a nectar source and speculated about a "language" of bees. After years of observation and hypothesizing, Karl von Frisch claimed that honeybees have an elaborate system of dancing by which they communicate the whereabouts of a honey supply. Various aspects of the dance of a bee returning to a hive indicate the distance and direction of a nectar source. The quality of the source can be gauged by sniffing the discoverer bee. Although some of von Frisch's interpretations have been questioned, his careful analysis demonstrated that the kind of creativity characteristic of a child's speech is lacking in the bee's dance. Bees do not use their communicative system to convey anything beyond a limited range of meanings (such as 'There's a pretty good source of nectar in this direction'), so analogies between bee dancing and a child's language are fundamentally misleading.

The same lack of creativity characterizes communication between other animals. Beyond a very limited repertoire of meanings, even intelligent mammals such as dogs lack the mental capacity to be communicatively creative. For one thing, much of the communication between animals relies on *nonarbitrary* signs. When gazelles sense potential danger, their fleeing sends a signal to nearby gazelles that danger is lurking, and the communicative function of the act is incidental to its survival function. Similarly, a dog signals the possibility that it might bite by displaying its fangs. These acts are nonarbitrary signs.

Whatever animals express through sounds seems to reflect not a logical sequence of thoughts but a sequence accompanying a series of emotional states. The communicative activities of most animals thus differ from human language in that they do not consist essentially of arbitrary signs. But what about chimpanzees, the closest living relative of human beings?

Can Chimpanzees Learn a Human Language?

In the wild, chimpanzees use a limited nonlinguistic communicative system similar to but generally more sophisticated than that of other mammals. Relying on the fact that the intelligence of chimps comes closest among other mammals to the intelligence of humans, researchers have attempted to teach them human language in ways that do not depend on vocal articulation.

Vicki The earliest chimp to gain notoriety for her communicative prowess was Vicki. After being raised for about seven years by psychologists Keith and Catherine Hayes, Vicki could utter only four words—*mama*, *papa*, *up*, and *cup*—and she managed those words only with considerable physical strain. Chimps are simply not equipped with suitable mouth and throat organs to speak.

Washoe Granted, then, that chimps lacked the *physiological* capacity to speak, the question remained whether they had the *mental* capacity to learn language. After viewing a film of Vicki trying to vocalize human language, psychologists Allen and Beatrix Gardner gave a home to a 10-month-old chimp named Washoe, whom they raised as a human child in as many ways as possible. Eventually Washoe ate with a fork and spoon, sat at a table and drank from a cup, and even washed dishes after a fashion. She wore diapers and became toilet trained; she played with dolls and showed them affection. Like human children her age, she was fond of picture books and enjoyed having her human friends tell her stories about the pictures in them.

Ingeniously, the Gardners arranged to conduct all communication with Washoe in American Sign Language (ASL), which they also used to communicate between themselves and with members of their research team whenever Washoe was present. Noting the kinds of simplified communication that human parents provide for children in some societies, the Gardners used repetition and simplified signing with Washoe, who in her first 7 months in this very human environment learned four signs and in the next 14 months an additional 30 signs. After 51 months, she had acquired 132 signs describing objects and notions, and she understood three times that many. She used the signs to designate particular objects and classes of objects. She used the sign for 'shoe' to mean shoes in general; she used the 'flower' sign for flowers in general and also for aromas like the smell of tobacco. She signed even to dogs and trees. She asked questions about the world of objects and events around her. After mastering only eight signs, she started combining them to make complex utterances: YOU ME HIDE; YOU ME GO OUT HURRY LISTEN DOG (when a dog barked); BABY MINE (referring to her doll); and so on. After just 10 months in her foster home, she made scores of combinations of three or more signs, such as ROGER WASHOE TICKLE and YOU TICKLE ME WASHOE.

Loulis In subsequent work with four other chimps (Moja, Pili, Tatu, and Dar) who arrived at their laboratory within days of birth, the Gardners demonstrated that chimps who are cross-fostered by human adults replicate some of the basic aspects of language acquisition characteristic of human children, including the use of signs to refer to natural language categories such as DOG, FLOWER, and SHOE. Remarkably, when these chimps later took up residence in another laboratory, an infant chimp named Loulis acquired at least 47 signs that had no source other than the signing of his fellow chimps.

Sarah and Lana In cross-fostering Washoe and her chimpanzee playmates, the Gardners made the pivotal assumptions that human language is acquired by children in a rich social and intellectual environment and that such richness contributes to the child's cognitive and linguistic life. In other laboratories, the language activities of other celebrity chimps were not vocal like Vicki's

nor gestural like Washoe's but visual. Sarah used plastic chips as symbols for words and showed considerable ability putting them in sequence. Lana used an appropriately marked computer terminal to create series of symbols similar to the plastic ones used by Sarah. As a result of the kinds of success enjoyed by the Gardners and other researchers working with chimps, some observers, especially among psychologists, came to believe there might be a continuum between human and nonhuman communication. But by no means was everyone persuaded.

Project Nim

Indeed, some psychologists have voiced skepticism about the various projects to teach chimps human-style language. They believe the individual words selected by the chimps in the various modes could have been triggered in some instances by inadvertent clues from the researchers. As a result, they claim, the sequences of strings produced by chimps are not productive sentences that parallel those created by human children. Other critics doubt that chimps have the ability to use language to make comments, ask questions, and express feelings as humans do.

In an attempt to provide more control on the effort to teach language to a chimp, a rigorous experiment sought to avoid many of the objections to previous research (though, inevitably, it introduced problems of its own). This chimp was named Nim Chimpsky, and in the course of his education Nim had several linguistic accomplishments, in part repeating the achievements of his predecessors. But after five years of work with Nim, psychologist Herbert Terrace concluded that chimps are incapable of learning language as children do. Even with elaborate training, Nim produced very few longer utterances and displayed little creativity and spontaneity in his use of signs. Unlike Washoe, Nim signed only when researchers prompted him, and he never initiated interactions. These characteristics, Terrace contends, clearly distinguish what Nim could learn from what children can do with language.

Critics of Project Nim note that Terrace employed more than 60 research assistants over five years and suspect that that fact may have contributed significantly to the limitations in Nim's linguistic achievements. Moreover, the assistants were instructed to treat Nim in a detached fashion and were forbidden to comfort him even when he cried during the night. Critics questioned how similar Nim's learning environment was to that in which a normal human child acquires language, and they maintained that the research conditions of Project Nim had a crippling impact on the chimp's emotional and linguistic education.

We see, then, that some researchers claim that human and animal language fall along a continuum, while others conclude that chimps cannot learn language as children do. To help reconcile these views, it is useful to consider the assessment of the distinguished primatologist Jane Goodall, who identified as the biggest difference between humans and chimps "Talking. Because we can discuss ideas, we can teach about things that aren't present. We can draw from the distant past and teach each other from it and we can plan the distant future." The consensus among psychologists and linguists now is that animal language, including that

of chimpanzees, does not exhibit all the hallmark features of human language. Despite some similarities between human language and the sign and other visual languages taught to chimps, several criterial features of human language, notably displacement and recursion and productivity, appear to be lacking.

The Origin of Human Languages: Babel to Babble

Many people are persuaded that language originated in a paradise where its form was logical and perfectly grammatical. Even beyond that, the belief is widespread that, with the passage of time, pristine languages have become contaminated with impurities, illogicalities, and ungrammaticalities.

As well as having different views on the origins of languages, people have different ways of explaining why languages differ from one another and why they change. The Old Testament relates that before the Tower of Babel all men and women spoke the same language and could understand one another. Eventually human pride provoked God into confounding their communication with mutually unintelligible tongues. In this view, language differences among people can be seen as a penalty for pride or misbehavior. Similarly, Muslims believe that Allah spoke to Mohammed in pure and perfect Arabic, as represented in the Koran. By contrast, the varieties of present-day Arabic spoken in the Persian Gulf, North Africa, and elsewhere are seen as deriving later from human weakness and culpability.

Professional linguists take a different approach. They see the multiplicity of languages as resulting from natural change over time, the inevitable product of reshaping speech to meet changing social and intellectual needs and reflecting contact with people speaking other languages. When people move to new places and mix with speakers of different tongues or settle with unfamiliar plants and animals, their language must adapt to new circumstances. Encountering unfamiliar aspects of nature and meeting others who engage in different cultural practices and hold different views prompts accommodation in one's language. As a result, languages evolve differently around the globe.

Still, despite obvious linguistic differences, the extent of similarity across languages around the globe is also remarkable. But because every language must conform to the character and abilities of the human vocal apparatus and to the character and abilities of the human brain, as well as to the acoustic limitations of the human ear, similarities in human language should not be surprising at all.

Still, the question persists: When and how did human language arise? In different forms, the question has been of interest for millennia. You may know the story of the Egyptian king Psammetichus. He entrusted two children to mute shepherds so the children would hear no language and whose first words, he reasoned, would therefore indicate the original language. When the first spoken word, recorded as *bekos*, was judged to be Phrygian for 'bread,' Psammetichus decided Phrygian must be the original language.

Around the time Darwin published his *Origin of Species* (1859), experimentation and speculation took a vigorous turn. Speculation grew so fantastical that in

1866 the Linguistic Society of Paris banned the offering of papers on the origins of language, and the spirit of that prohibition—coupled with a lack of sufficient knowledge in the necessary fields—kept the question on the back burner until late in the twentieth century. In the past 25 years, though, language origin has become a hot topic and engaged researchers from biology, psychology, neuroscience, anthropology, archaeology, modeling, and computer science, as well as linguistics. Since 1996, “Evolang” conferences have helped focus attention on the complexity of the question and enabled researchers across disciplines to understand one another’s insights. Unfortunately, while some issues have been clarified, answers remain elusive.

One of the principal challenges in determining the origin of language is the fact that three vastly different timelines are involved, and the computer modeling needed to help determine when and how language evolved is hampered by the greatly different scope of these timelines and by our incomplete knowledge about them. For example, while a human being acquires a language over a period of a decade or so, languages themselves develop over centuries or millennia, and the human brain and vocal apparatus developed over hundreds of thousands of years. Thus anything we know about language evolution along each of these dimensions is difficult to combine in a single model. But there also remain other fundamental questions—for example, about how and how much human language differs from animal communication; about exactly what human language is—which aspects of human language are characteristic of language per se and which belong to more general cognitive, sensory-motor, or sociocultural functions. Fascinating as they are, even detailing these questions would take us well beyond our scope. Some researchers, especially those with a more Eurocentric view of cultural evolution, are inclined to think that human language as we understand it today emerged only 40,000 to 50,000 years ago. Others think that it cannot be any younger than 400,000 or 500,000 years old. That span of time leaves lots of room for research, and references at the end of the chapter will point you to further information about the origins of human language.

What Is Linguistics?

Linguistics can be defined as the systematic inquiry into human language—into its structures and uses and the relationship between them, as well as into its development through history and its acquisition by children and adults. The scope of linguistics includes both language structure (and its underlying *grammatical competence*) and language use (and its underlying *communicative competence*).

Language is often defined as an arbitrary vocal system used by human beings to communicate with one another. This definition is useful as far as it goes although it downplays writing and signing. It also downplays an important fact that philosophers have emphasized about language, namely, that it is more than communication. Language is social action, with work to perform. It is a system that speakers, writers, and signers exploit purposefully. It is used to *do* things, not just report them or describe them or discuss them. “That shirt looks cool on you!” is more likely a compliment than a mere report, whereas

“Halloween falls on a Tuesday this year” might well be a report. “Out!” is a mere opinion or conjecture when a fan behind home plate shouts it during a baseball game, but when said by the umpire, “Out!” is a *call* and can end an inning—or a game.

As we said previously, people have been interested in analyzing language for millennia. Plato and Aristotle discussed language in the fourth and third centuries B.C., and we have inherited several categories of grammatical analysis from them. More than a century earlier, the Indian scholar Pāṇini wrote a description of Sanskrit that is one of the finest grammars ever produced for any language. Today the empirical study of language has taken on additional importance in an age in which communication is critical to social, intellectual, political, economic, and ethical concerns. Now augmented by insights from cognitive science and neuroscience, from computer science, psychology, sociology, anthropology, philosophy, and rhetoric, as well as from communications engineering and other sciences, linguistics has become a prominent and vital academic discipline in universities and research centers throughout the world.

What Are the Branches of Linguistics?

Historically, the central focus of language study has been *grammar*—the structural patterns of speech sounds, words, sentence formation, and meaning. More recently, in a field called *pragmatics*, attention has also focused on the relationship between expression and meaning, on the one hand, and context and interpretation, on the other. Some linguists describe particular languages; others examine universal patterns across languages and strive to explain them in cognitive or social terms.

Some linguists focus on *language variation* across speech communities or, within a single community, across time or across situations of use, such as conversation and sports announcer talk. Linguists studying variation seek two kinds of explanation—cognitive ones having to do with constraints on the human language-processing capacities and social ones having to do with human interaction and the organization of societies.

A third group of linguists applies the findings of the discipline to real-world problems in *educational* matters, including the acquisition of literacy (reading and writing) and of second languages and foreign languages; in *clinical* matters, to understanding aspects of Alzheimer’s disease and aphasia; in *forensic* settings, to analysis of conversation for evidence of conspiracy, threats, defamation, and other matters of legal concern, to interpretation of contracts (from rental agreements and insurance policies to agreements for manufacturing airplanes), to clarification of public safety instructions (such as medical labels and dosage directions), and to voice identification, disputed authorship, and plagiarism. Some applied linguists address problems in *language policy* at national and local levels: what languages to designate for use in schools, courts, voting booths, and so on; what kind of writing system to employ in a culturally diverse modern nation; what resources and regulation of existing languages are needed, as with dictionaries and grammars. As the world shrinks and cultures mix, linguists are also applying their skills to the challenges of cross-cultural communication.

COMPUTERS AND LINGUISTICS

At the end of each chapter in this book, you'll find a section that discusses aspects of computers and language related to the topics in the chapter. You don't have to be a techie or sophisticated computer user to understand these sections and appreciate them. Seeing parts of a chapter from a different perspective will help you grasp them. The section below serves as an introduction to the parallel sections in later chapters.

Computers and Corpora

In the eighteenth century, Samuel Johnson's dictionary provided illustrative citations taken from books to exemplify how words were used. During his own reading, Dr. Johnson marked sentences whose context made a word's meaning or use especially clear. Assistants then transcribed the marked passages onto individual sheets of paper, and Johnson organized them in the entries of the dictionary. In the nineteenth century, Noah Webster used essentially the same process in preparing his two-volume *American Dictionary of the English Language*. The same practice was followed for the multi-volume *Oxford English Dictionary*, a project so grand that it required thousands of readers and half a century to complete. In the twentieth century, the makers of *Webster's Third New International Dictionary* (1961) also mined a collection of several million paper citations to uncover different word senses.

Today, dictionary making is identical in principle but could hardly be more different in practice. Machine-readable collections of texts—known as *corpora*—and the inexpensive availability of computing power has revolutionized dictionary making and begun to revolutionize how courts, including the United States Supreme Court, address language matters.



Corpus linguistics is the term used for compiling collections of texts and using them to probe language use.

Corpus is the Latin word for 'body,' and a **corpus** is a body of texts. You're familiar with the kinds of machine-readable texts created by word processors and by scanning: in a machine-readable text you can search for particular words or phrases. The most obvious examples, of course, are the various engines used today to search the web such as Google, Yahoo!, and Bing. The first computerized corpus—the Brown University Corpus—included a mere 500 texts of 2,000 words each from U.S. books, newspapers, and magazines that were published in the year 1961. The texts represented 15 genres, including science fiction, romance fiction, press reportage, and scholarly and scientific writing. Brown contained what at the time was a very impressive one million words. Today, one million words is peanuts, chump change, nada, (almost) nothing! A hundred million words is readily available even in a single genre—soap operas! And instead of being limited to a single year, corpora extend back in time for centuries and right up to the present day—and not only for English but for dozens of languages. It is no exaggeration to say that large corpora have revolutionized dictionaries and the way they are made; they have revolutionized our understanding of language in some ways and begun affecting many professions, including law and engineering, and many innovations, including speech recognition on smartphones and the software that powers it, as with Andy and Siri. In this section of each chapter, we'll explore how the intersection of computers and linguistics has come to play vital roles in everyday life. ■

Summary

- The total number of spoken and signed languages in the world is about 7,000.
- According to U.S. Census data, 55 million U.S. residents aged 5 or over speak a home language other than English.
- The United States does *not* have an official language and has *never* had one.
- Human language is an enormously complex system that is nevertheless easily mastered by children in a remarkably short time.
- Natural processes of linguistic change affect all languages over time, and linguistic change is not linguistic decay.
- A human language is primarily a system of arbitrary signs, but some linguistic signs are representational.
- Grammar is a system of elements and patterns that organizes linguistic expression.
- Five hallmarks of human language systems are arbitrariness, discreteness, duality, displacement, and recursion/productivity.
- Rather than being a two-sided coin, a language system is better viewed as a triangle whose faces are meaning and expression and whose base is context.
- Linguistic communication can operate in three modes: speaking, writing, and signing.
- Everyone speaks a dialect, and languages do not exist independently of their dialects.
- Chimpanzees do not have a suitable vocal apparatus for speaking but in very limited ways are capable of joining several signs in meaningful strings.
- The degree to which the language of chimps and that of very young children are alike remains under investigation, but they appear fundamentally different at least in displacement and recursion/productivity.
- Large bodies of machine-readable texts called corpora are used to explore natural language use in all its contexts.



What Do You Think? REVISITED

1. *Rob and Rod*. The question can't be answered exactly, but there are in the world about 7,000 languages, including sign languages.
2. *Multilingual utility bill*. The United States has never had an official language (although some states have official languages). Customers need to understand their utility bills, and that means the bills must be written in a familiar language. When confronted with an important piece of text in an unfamiliar language, who isn't relieved to then find a translation in a familiar language? In the United States, with large numbers of immigrants who often cluster in urban or suburban areas, it makes good sense for commercial and governmental entities to communicate with them in a language they can understand.
3. *Sixth-grade Sydney*. *Note* carries different senses in *musical note*, *bank note*, and the metaphorical *discordant note*. And besides its senses as a noun, *note* may be a verb. So knowing the context of use may be essential in recognizing

the appropriate sense of a word in an utterance, and a sentence such as “The U.S. Treasury issued new \$20 notes in 2003 to help stem the tide of counterfeiting” makes clear that “notes” is a synonym for “bills.”

4. *Seth's birthday.* Seth understands that utterances must be interpreted in context. Felix pretends that utterances have only a literal meaning, irrespective of context. Because everyone knows his or her own birthday, Felix's question (“Do you know when your birthday is?”) would appear silly if taken at face value, so Seth figures out an interpretation that makes sense of it. “May ninth” is information Seth presumes Felix doesn't have: Why else would he ask the question? ■

Exercises

- 1.1. Here's a set of questions that could constitute a basis for your linguistic autobiography. Jot down your answers in bulleted form. (1) When did you first become aware that people judge certain linguistic expressions to be naughty or nice, and what do you think the basis for those judgments must have been? (2) When did you first become aware that some people judge certain linguistic expressions to be grammatically correct or incorrect, and what do you think the basis for their judgments would have been? (3) Do you think of speech as more fundamental than writing—and, if so, why do you think that? (4) Was there ever a time when you judged writing to be the basis for speech, and, if so, what gave you that impression? For example, have you ever thought that *palm*, *salmon*, and *almond* should be pronounced with an *l* because they're spelled with an *l*? What about *herb*: would it be more correct to pronounce it with an *h*? (5) Can you identify two aspects of your current views about language that place writing in a superior position to speaking?
- 1-2. Over the course of a single day, write down every instance you hear (on radio or television programs, in class lectures, or in talk among your acquaintances) of representational expressions (representing length, loudness, speed, repetition, emphasis, ordering, etc.). (You may find it easier to gather examples from a sitcom or children's program.) Compare your examples with those collected by your classmates to see how many kinds have been uncovered.
- 1-3. Below is a list of characteristics that describe linguistic communication through speaking, writing, and signing. Decide which modes of linguistic communication the characteristic applies to, and provide an example to illustrate your claim. Pay particular attention to the different types of spoken, written, and signed communication because certain of these characteristics might apply to some but not other types of communication. Also note the impact of modern communication technology on these characteristics.
- A linguistic message is ephemeral—that is, it cannot be made to endure.
 - A linguistic message can be revised once it has been produced.

- c. A linguistic message has the potential of reaching large audiences.
- d. A linguistic message can be transmitted over great distances.
- e. A linguistic message can rely on the context in which it is produced; the producer can refer to the time and place in which the message is produced without fearing misunderstanding.
- f. A linguistic message relies on the senses of hearing, touching, and seeing.
- g. The ability to produce linguistic messages is innate; it does not have to be learned consciously.
- h. A linguistic message must be planned carefully before it is produced.
- i. The production of a linguistic message can be accomplished simultaneously with another activity.

- 1-4. Writing and gesture are visual modes of linguistic communication. What is the relationship between writing and Braille (the writing system used for blind readers)? Is Braille a mode of linguistic communication? How many modes of linguistic communication are there?
- 1-5. When there is a choice between linguistic modes, as in telephoning a friend or sending a letter or email note, what are the advantages and disadvantages of each mode? What about instant messaging and texting? List some of the circumstances in which each mode of linguistic communication would be preferred over the others.
- 1-6. Using the bracketing that appears on page 10 as a model (but ignoring the font differences displayed there), bracket the sentence below in a way that accurately represents all of the relative clauses with respect to their relationships to one another.

It's a weight-loss program that people who can't easily stick to a diet that requires giving up chocolate that's made in Belgium can succeed at.

Especially for Educators and Future Teachers

- 1-7. For students whose home language matches the language of instruction in school, do you regard the primary focus of teaching language arts to be reading and writing or speaking and listening? Explain your position.
- 1-8. For the same group of students, do you think the emphasis of the curriculum is actually on reading and writing, or is it on speaking and listening? Explain the basis for your view and answer whether the emphasis you've identified is the better one.
- 1-9. For students whose home language differs from that of school instruction (for example, for students who speak Spanish at home but attend an English-language school), would your answers to the previous two questions be different? If so, how?

Other Resources



Internet

Information and the results of considerable laboratory research are available on the Internet. In this section of each chapter you'll find Internet addresses that will help you understand the contents of the chapter and provide a laboratory unlike any previously available to students at even the best-equipped universities. Because Internet addresses can change unexpectedly, the ones given here may no longer be operational by the time you try them. If an address has changed, you may find yourself automatically linked to a new address or you may be able to use hints from the description below to find the new site or a comparable one.

- **LISU Website:** <http://www.CengageBrain.com> For users of this textbook. Provides supplemental material for students and instructors. You will find chapter-specific interactive learning tools in your English CourseMate, accessed through this URL.
- **The Domain of Linguistics:** <http://www.lsadc.org/info/ling-fields.cfm> For general information, the website of the Linguistic Society of America. You'll find brief treatments on a wide variety of topics, including language and the brain, language and thought, computers and language, endangered languages, writing, slips of the tongue, linguistics and literature.
- **An Animated ASL Dictionary:** <http://commtechlab.msu.edu/sites/aslweb/browser.htm> You can see animated representations of ASL signs.
- **James Crawford's Language Policy Website & Emporium:** <http://www.Languagepolicy.net/> Bilingual education and bilingual policy are important matters in the United States and around the world. In the United States, the subject of bilingual education is a tender one. This website is a rich source of information and interpretation of language policy in the United States. Much of what is of value at the site can be reached through the "Archives" tab or at <http://www.Languagepolicy.net/archives/home.htm> directly.
- **Ethnologue:** <http://www.ethnologue.com/> This reference database offers a wealth of information about the distribution of languages, numbers of speakers, dialects, and so on. Organized by country and by language name.
- **Census 2000 Gateway:** <http://www.census.gov/main/www/cen2000.html> At the official site of the U.S. Census Bureau, you can see which languages are spoken by how many people in every state and the entire United States, with easy-to-read tables. At <http://www.census.gov/2010census> you'll find information gathered in the 2010 U.S. Census.
- **EvoLang:** <http://evolang.org> Website of the Evolution of Language International Conferences. These biennial conferences began in 1996 and have attracted many of the world's leading researchers on the evolution of language from every discipline with an interest in the question. You'll find links to independent sites for earlier and upcoming Evolang conferences, and they sometimes contain very useful information and links. For example, video presentations of the plenary addresses given at

the 2012 Evolang conference can be viewed by clicking on the Kyoto 2012 button. You'll find presentations on "Outgroup: The Study of Chimpanzees to Know the Human Mind," on "Why Language Has Structure: New Evidence from Studying Cultural Evolution in the Lab and What It Means for Biological Evolution," and on the topic "Neither Nature nor Nurture: Coevolution, Devolution, and Universality of Language."

- **Endangered Language Alliance:** <http://elalliance.org> Established by Daniel Kaufman, who calls New York City "an endangered language hot spot," the alliance fosters recording of endangered languages that sometimes have more speakers in NYC than in their original homelands.

Films and Videos

- **UNESCO Endangered Languages:** http://www.unesco.org/culture/en/endangered_languages A rich site, with maps documenting the state of endangered languages around the globe.
- **The Linguists:** <http://thelinguists.com/> *The Linguists* is a riveting story of two linguists' efforts to record endangered languages around the globe. Supported by the National Science Foundation, the film had its world premier at the Sundance Film Festival in 2008. At the website you'll find a trailer and ordering information.
- **Student Productions:** Inspired in part by *The Linguists* (see preceding bullet), two undergraduate students at the University of Southern California produced videos about remote and endangered languages. Focusing on the Central Alaskan Yup'ik language, Lydia Green's project at <http://www.lydiajewelgreen.com/> shows how involved a college junior can become and how much firsthand information and insight can be gained. At <http://archive.org/details/SouthernCaliforniaIndigenousLanguagesPilotFilm> you can join college senior Joseph Henderer on his journey to find the last fluent speakers of Southern California's indigenous languages.
- **The Human Languages Project:** An award-winning set of informative and entertaining videos, originally broadcast on PBS in 1995: *Discovering the Human Language: "Colorless Green Ideas"; Acquiring the Human Language: "Playing the Language Game";* and *The Human Language Evolves: "With and without Words."* 55-minutes each. Produced by Equinox Films, Inc. (Visit www.equinoxfilms.net/page1.html.)
- **Sound and Fury:** Nominated in 2001 for an Academy Award as a documentary feature, *Sound and Fury* tells the story of two families grappling with the complicated and controversial decision about whether to have a deaf child receive a cochlear implant. At <http://www.pbs.org/wnet/soundandfury/> you'll find links to information about Deaf culture, cochlear implants, American Sign Language, and related topics. Information about two follow-up films, *Sound and Fury: Six Years Later* and *Twins*, made by the same director can be found at <http://soundandfuryfilm.com> or <http://www.aronsonfilms.com/filmsReleased.html>.

Suggestions for Further Reading

- Jean Aitchison. 2000. *The Seeds of Speech: Language Origin and Evolution* (Cambridge, UK: Cambridge University Press). A basic treatment of language beginnings.
- Douglas Biber, Susan Conrad & Randi Reppen. 1998. *Corpus Linguistics: Investigating Language Structure and Use* (Cambridge, UK: Cambridge University Press). An accessible introduction to corpus linguistics.
- David Crystal. 2010. *Cambridge Encyclopedia of Language*, 3rd ed. (Cambridge, UK: Cambridge University Press). Treats topics in a few pages each, with illustrations.
- Edward Finegan & John R. Rickford, eds. 2004. *Language in the USA: Themes for the Twenty-first Century* (Cambridge, UK: Cambridge University Press). These essays treat multilingualism, Spanish in the Southwest and the Northeast, African-American English,

- Asian-American voices, language and education, the language of cyberspace, rap and hip-hop, slang, and 17 other current topics.
- **Ray Jackendoff.** 1994. *Patterns in the Mind: Language and Human Nature* (New York: Basic Books). Accessible, fascinating discussion of the cognitive aspects of language structure and language acquisition.
 - **Donna Jo Napoli.** 2003. *Language Matters: A Guide to Everyday Thinking about Language* (New York: Oxford University Press). An enjoyable introduction to most topics treated in this chapter.
 - **Edward Sapir.** 1921. *Language: An Introduction to the Study of Speech* (New York: Harvest). Because this accessible classic continues to yield insight, it remains in print and is worth reading.

Advanced Reading

For discussion of the relationship between arbitrary and nonarbitrary signs, consult de Saussure (1959). The papers in Haiman (1985) touch on iconic elements in syntax and intonation. For speaking and writing, see "Suggestions for Further Reading" in Chapter 15 (available at the LISU website). The linguistic aspects of the Ripper case are discussed in Ellis (1994), Windsor Lewis (1994), and French et al. (2006). Because of the extraordinary nature of the manhunt in England and the attention it garnered, you can readily locate YouTube recordings of Humble's original hoax tape on the Internet (try "Yorkshire Ripper Hoaxer"). The quote from Jane Goodall comes from Middleton and Else (2005). For information on American Sign Language, see Lucas and Valli (2004), upon which our exposition here relies, and for a survey of sign languages among Native Americans and Australian Aborigines, see Umiker-Sebeok and Sebeok (1978). On the origins of language, see Lieberman (2006). Articles reviewing the state of knowledge about language evolution include the excellent and accessible Christiansen and Kirby (2003) and the even more accessible Croom (2003). For corpus linguistics, see McEnery et al. (2006). Crystal's (2008) *Dictionary of Linguistics and Phonetics* is a useful reference work for a wide set of terms and concepts.

References

- Christiansen, Morten H. & Simon Kirby. 2003. "Language Evolution: Consensus and Controversies." *Trends in Cognitive Sciences* 7 (7): 300–307.
- Croom, Christopher. 2003. "Language Origins: Did Language Evolve Like the Vertebrate Eye, or Was It More Like Bird Feathers?" Available at <http://www.csa.com/discoveryguides/lang/overview.php>.
- Crystal, David. 2008. *A Dictionary of Linguistics and Phonetics*, 6th ed. (Oxford: Wiley-Blackwell).
- De Saussure, Ferdinand. 1959. *Course in General Linguistics*, trans. from French by Wade Baskin (New York: Philosophic Library).
- Ellis, Stanley. 1994. "The Yorkshire Ripper Enquiry: Part I." *Forensic Linguistics* 1: 197–206.
- French, Peter, Philip Harrison & Jack Windsor Lewis. 2006. "R v John Samuel Humble: The Yorkshire Ripper Hoaxer Trial." *The International Journal of Speech, Language and the Law* 13: 255–73.
- Haiman, John, ed. 1985. *Iconicity in Syntax* (Amsterdam: Benjamins).
- Lieberman, Philip. 2006. *Toward an Evolutionary Biology of Language* (Cambridge, MA: Harvard University Press).
- Lucas, Ceil & Clayton Valli. 2004. "American Sign Language." In Edward Finegan & John R. Rickford, eds. *Language in the USA* (Cambridge, UK: Cambridge University Press), pp. 230–44.
- McEnery, Tony, Richard Xiao & Yukio Tono. 2006. *Corpus-Based Language Studies: An Advanced Resource Book* (London: Routledge).
- Middleton, Lucy & Liz Else. 2005. "Close Encounters." *New Scientist* 186.2502: 46–47.
- Umiker-Sebeok, Jean D. & Thomas A. Sebeok, eds. 1978. *Aboriginal Sign Languages of the Americas and Australia*, 2 vols. (New York: Plenum).
- Windsor Lewis, Jack. 1994. "The Yorkshire Ripper Enquiry: Part II." *Forensic Linguistics* 1: 207–16.

Part One

Language Structures

When you think about languages—about a child acquiring its first language or you yourself learning a foreign language—you are likely to think about *words*. Children are said to have spoken their first “word.” Students studying a foreign language ask, “What’s the word for X?” Dictionaries, the most familiar and frequently consulted source of information about language, are organized by word entries. In this book, we begin our investigation of language structure by looking at words, and we look at them from several perspectives:

- their meaningful parts
- the sounds that make them up
- the principles that organize them into phrases and sentences
- the semantic relationships that link them in sets.

In the first part of this book, you’ll see how just a few basic elements combine into speech sounds, how just a few speech sounds combine to form a larger number of syllables, how syllables combine to produce a far larger number of word parts that carry meaning, and how languages package these word parts into a huge but finite vocabulary and an infinite number of phrases and sentences. You’ll also see how the systematic principles underlying language structure enable you to understand countless utterances you’ve never before encountered. Finally, you’ll examine various kinds of semantic relationships among sets of words.

2

Words and Their Parts: Lexicon and Morphology



What Do You Think?

- If you were the parent of a 3-year-old daughter who asked if you “maked” a cake and “telled” your friends about it, how would you describe her pattern for marking past time on these verbs?
- You’ve agreed to make a list of drinks and dishes that volunteers can contribute to a fundraiser for a college choir undertaking an international tour. All the items must bear a name that English borrowed from another language. You think of *sushi* from Japanese, *chop suey* from Chinese, *tortilla* from Spanish, *pâté* and *champagne* from French, and *curry* from the Tamil language of India. Can you identify other food names from these languages? What about from other languages?
- Your pal Walter Wordley challenges you to identify the ten most frequent words in printed English. Which ones would you name? Are there any nouns or adjectives on your list?
- Nerdy Nan natters away about *Time* magazine calling the director of the Johns Hopkins Cancer Center “a successful rock looker underer.” She wonders why English tolerates such nonsense. What do you tell her?
- Grumpy Greg grouses that his history prof corrected the word *snuck* to *sneaked* on his term paper about World War II. “Everyone I know says *snuck*,” Greg protests. “What planet is *she* living on?” Who’s right—Greg or his prof?

Introduction: Words Seem Tangible

The most tangible elements of a language are its words. You've heard people say "There's no such word!" or ask "What does the word *lollapalooza* mean?" Someone doing a crossword puzzle may ask you, "What's a three-letter word for *excessively*?" We say one person likes to use "two-bit" words and another has a preference for "four-letter" words. In these instances people have clear notions of what a word is. When it comes to meaningful parts smaller than a word, we readily understand that *bookstore*, *laptop*, and *headset* have two each, but our intuitions may be less confident about *bookkeeper*, *sneakers*, *women's*, *sang*, and *impracticality*. This chapter examines words and their meaningful parts, as well as the principles that govern the composition of words and their functions in sentences. You'll learn what it means to know a word and how languages increase their store of words.

What Does It Mean to Know a Word?

Consider what a child must know in order to use a word. The child who asks "Can you take off my shoes?" knows a good deal more about the word *shoes* than what it refers to. She knows the sounds in *shoes* and the sequence in which they occur. She knows that the word can be used in the plural (unlike, say, *milk* or *sugar*) and that the plural is formed regularly (it isn't irregular like *teeth* or *children*). She also knows how to use the word in a sentence.

Using a word requires four kinds of information:

- ◆ its sounds and their sequencing (this is called *phonological* information, the topic of Chapters 3 and 4)
- ◆ its meanings (*semantic* information, discussed in Chapter 6)
- ◆ its category (e.g., noun, verb, or adjective) and how to use it in a sentence (*syntactic* information, discussed here and in Chapter 5)
- ◆ how related words (such as plural for nouns, past tense for verbs) are formed (*morphological* information, treated in this chapter).

For children and adults, using any word requires information about sounds, meanings, related words, and use in sentences, and that information is stored in the brain's dictionary. That dictionary is called the *mental lexicon* or simply *lexicon*.

There are parallels between the kinds of information stored in the mental lexicon and those found in a dictionary. Both contain information about pronunciation, meaning, related words, and sentence use. But a dictionary also contains information that is not needed for speaking—for example, a word's spelling and its historical development (called its *etymology*). Dictionary entries also provide illustrative phrases or sentences, showing how a word has been used by writers or speakers. A mental lexicon does not contain etymological, illustrative, or spelling information.

Lexical Categories (Parts of Speech)

The ability to use any word in a sentence requires knowledge of its **lexical category**. Even young children must know the category of every word they use—whether it is a verb, noun, adjective, or other part of speech. To be able to say, “I need some help” and “Can you help me?” requires knowing that *help* can be a noun (first sentence) or a verb (second sentence). Of course, a child’s knowledge about lexical categories is unconscious, and even a grammarian’s child wouldn’t ordinarily know the category names.

How to Identify Lexical Categories

There are several ways to help identify the lexical category of a word, and to some extent they rely on principles similar to those children must use in figuring out that same information. One way focuses on a word’s closely related forms. *Spoon* and *spoons*, *book* and *books*, *game* and *games* show parallel patterns of related forms, and words with parallel forms belong to the same category—in this case, noun. *Old*, *tall*, and *bright* follow a different pattern. Their related forms have *-er* and *-est* endings: *older/oldest*, *taller/tallest*, *brighter/brightest*. *Old*, *tall*, and *bright* belong to the adjective category. Words such as *wait*, *kick*, and *laugh* appear with a different set of parallel endings, including *-ed* (*waited*, *kicked*, *laughed*), *-ing* (*waiting*, *kicking*, *laughing*), and *-s* (*waits*, *kicks*, *laughs*); they belong to the verb category.

Another way to identify lexical categories focuses on which words and categories can occur together in phrases. (We’ll analyze phrases in Chapter 5. Until then, you can regard a phrase simply as a group of words functioning as a unit. A noun phrase is a word or group of words functioning like a noun and having a noun as the head: *smartphones* and *Kyle* are nouns and also noun phrases; *those smartphones* and *Kyle’s old smartphones* are noun phrases whose head is *smartphones*. A verb phrase is a word or group of words functioning like a verb and having a verb as the head; *study* and *visit* are verbs and also verb phrases; *won’t study* is a verb phrase with *study* as its head; *will visit later today* is a verb phrase with *visit* as its head.) For example, the nouns above can be preceded by *the* and *a* (or *an*): *the spoon/a game*, and the plural forms ending in *-s* can be preceded by *the*. Basic adjectives can be preceded by *very* or *too*, as in *very smart* and *too bright*. Basic verbs can be preceded by *can*, *must*, or *will*: *will laugh*. Below are examples of these patterns for the three lexical categories discussed so far.

Nouns

bike	bikes	a bike	the bike(s)
aunt	aunts	an aunt	the aunt(s)
exam	exams	an exam	the exam(s)

Adjectives

old	older	oldest	very old	too old
new	newer	newest	very new	too new
strong	stronger	strongest	very strong	too strong

Verbs

look	looks	looked	looking	can look	will look
play	plays	played	playing	can play	will play
march	marches	marched	marching	can march	will march

Knowing the typical related forms in each category enables you to gauge that a word like *sharper* is related to the adjective *sharp* (note *too sharp* and *very sharp*), whereas a word like *shaver* is related to the noun *shavers* (“shavest” isn’t a word, and you can’t say *very shave* or *shaver*). A word like *missed* can be related to the verb *miss*, with its related forms *missing/misses*, *can miss*, and *will miss*. From an early age, children recognize that words belonging to different categories have characteristic endings or forms and characteristic distributions in phrases.

Relying on meaning is a third way to identify a word’s lexical category. While it is less reliable than the other two methods, it can be useful in forming an initial hypothesis about a word’s category. From the perspective of meaning, nouns refer to (or name) entities: persons, places, things. Thus, *swimmer*, *sugar*, *city*, and *trees* are nouns, as in the phrases *that swimmer*, *some sugar*, *the city*, and *these trees*. Adjectives identify, refer to, or name qualities or properties of nouns, as with *tall* and *impressive* in the phrases *tall trees* and *an impressive city*. Verbs identify or name actions, as with *jump* and *sings*, or states of being, as with *remain* and *seem*, but some nouns refer to actions, as with *walking* (*walking is good exercise*) and *singing* (*her singing was surprisingly good*). But these nouns don’t have related verb forms like “walkings” or “walkinged.”

Nouns

Nouns constitute the most familiar category. You have already seen that English nouns share a set of endings, technically called inflections. The inflection at the end of *spoons* represents information about *number*. **Number** is the term used to cover *singular* and *plural* (and sometimes other features), and nearly all English nouns have distinct singular and plural forms. Some nouns are “regular” (*cat/cats*, *dish/dishes*), while others are “irregular” (*tooth/teeth*, *child/children*). A few exceptions like *deer* and *sheep* have the same singular and plural form.

Adjectives

Many **adjectives** can be recognized by the pattern of their related forms: the endings *-er* and *-est*, as in *hotter* and *hottest*. But others, especially those whose basic form contains more than two syllables, do not permit *-er* and *-est* endings; **beautifuler* and **intuitivest* are not well-formed English words (so we mark them with asterisks). But *beautiful* does share what are called co-occurrence patterns with other adjectives. In particular, like other adjectives it can be preceded by *very* or *too* (*very beautiful*) and can precede nouns (*intuitive powers*). Not all words preceding nouns and seeming to modify them are adjectives. In fact, nouns are often preceded by other nouns, as with *summer fun*, *graduation party*, *egg industry*, and *student government*. The prenominal position of such nouns does not make them adjectives. As a third frame, the only single words that fit into “it seems —” or “he/she seems —”

would be adjectives: *odd, sad, certain, funny, beautiful, intuitive*. Later in this chapter we'll discuss suffixes that transform words of one lexical category into words of another lexical category. As a consequence of their ability to form adjectives from words belonging to certain other lexical categories, some suffixes may also be helpful in identifying adjectives, as with *-able* (*breakable, changeable, debatable*), *-ful* (*peaceful, colorful, thoughtful*), *-ish* (*Danish, childish, thirtyish*), *-ous* (*dangerous, joyous, odorous*), *-al* (*fictional, parental, global, cultural*), *-ic* (*allergic, scientific, academic, gastronomic*), *-less* (*harmless, priceless, odorless, timeless*), and *-y* (*slimey, creepy, moody*).

Verbs

English-speaking children implicitly know that each **verb** has a set of related forms (*talk, talks, talked, talking*) and that the base form—the one without an ending—can be preceded by *can* or *will*.

Subcategories of Verbs Beyond knowing its category, a child must also implicitly understand the kinds of sentence structures a verb allows. Like other knowledge about each verb, knowledge about permitted sentence structures is also stored in the mental lexicon. Consider items 1 through 6, where the asterisks preceding 2, 3, and 5 mark those sentences as ill formed.

1. Sarah told a joke.
2. *Sarah laughed a joke.
3. *Sarah told at a joke.
4. Sarah laughed at a joke.
5. *Sarah told.
6. Sarah laughed.

You can see that the verbs *told* and *laughed* permit different structures after them. Sentences with *tell/told* require a noun after the verb; if they didn't require a noun, sentence 5 would not be marked with an asterisk. Sentence 6 demonstrates that not all verbs require a noun after them, and *laugh/laughed* is one that does not permit a noun to follow directly, as 2 shows. *Laugh* can be followed by the phrase *at a joke*, but "laugh a joke" is ungrammatical. Still other verbs permit—but don't require—a following noun, as 7 and 8 show:

7. The diva then sang.
8. The diva then sang an aria.

Sentences 1 through 8 illustrate that words such as *tell, laugh, and play* belong to the *category* verb but do not permit the same sentence structures. They belong to the same category but to different *subcategories*, and if children did not possess implicit information about verb **subcategorization**, they couldn't avoid uttering sentences like 2, 3, and 5.

Verbs that take a noun (or noun phrase) directly after them (*told a joke, caught the train*) are called **transitive** verbs. Those that do not require a noun are called **intransitive** (*lie, pray, shower*). In a child's mental lexicon, every verb is subcategorized as transitive or intransitive. Again, of course, it is the structural knowledge—not the terms—that a child understands.

Adverbs

Adverbs have been called “the most nebulous and puzzling of the traditional word classes.” For one thing, adverbs can’t be identified solely by their form and don’t generally have related forms. While many adverbs are derived from adjectives by adding *-ly*, as with *swiftly*, *oddly*, *brightly*, and *impossibly*, other adverbs, including the most common ones (for example, *then*, *now*, *here*, *soon*, *very*, *always*) carry no distinctive marker. Furthermore, not all words ending in *-ly* are adverbs; *manly*, *lovely*, *timely*, *homely*, and *lonely* are adjectives. The most reliable way of identifying adverbs is by their patterns of distribution—where they occur in sentences and with which other categories they co-occur. Meaning can also help identify adverbs as they often indicate something about the action expressed by the verb phrase—for example, when (*now*, *then*, *often*, *soon*), where (*here*, *there*), and how (*quickly*, *suddenly*, *fiercely*). Some grammarians count the word *not* as an adverb. Grammatically, adverbs display a wide range of functions: besides modifying verbs (and verb phrases), they modify adjectives, other adverbs, and even whole sentences.

Adverbs Modify Verbs (related adjectives in parentheses)

He talked *loudly*. (He was a *loud* talker.)

She slept *soundly*. (She was a *sound* sleeper.)

She thought *quickly*. (She was a *quick* thinker.)

They studied *diligently*. (They were *diligent* students.)

Adverbs Modify Verbs

She spoke *often*.

She studied *here*.

It *suddenly* stopped.

He *fiercely* denied it.

Below, the modifying adverbs are underlined, and the modified *adjectives* or *adverbs* are italicized.

Adverbs Modify Adjectives

a very *tall* tree

a bitterly *cold* winter

a truly *splendid* evening

Adverbs Modify Adverbs

very *soon*

unbelievably *quickly*

truly *unbelievably* fast

Adverbs Modifying Sentences

Actually, it was Danielle who said it.

I don’t know why I’m here, *frankly*.

Unfortunately, the remark wasn’t funny.

Pronouns

Pronouns constitute a category of relatively few words (it’s a closed category, as we’ll see below), and they fall into several subcategories. Generally speaking,

we can say that pronouns substitute for nouns and noun phrases in specific kinds of contexts. They function independently, taking the place of nouns and noun phrases and not as modifiers of other words. Below are listed some of the sub-categories of pronouns.

Personal Pronouns The most familiar pronouns are **personal pronouns**, such as *I, me, she, him, they, and theirs*. Personal pronouns are distinguished from one another by representing different parties to a conversation or other social interaction. This aspect of pronouns is called *person*, and the first person is the speaker or speakers; the second person is the person or persons spoken to (the addressee); and any persons or things spoken about are third persons.

First person—speaker: *I, me, mine, we, us, ours*

Second person—addressee: *you, yours*

Third person—spoken about: *she, her, hers, he, him, his, it, its, they, them, theirs*

Interrogative Pronouns *Who* in *Who played the role of Batman?* and *what* in *You told Jessica what?* and *What did you tell Ashley?* are **interrogative pronouns**, which are used in asking questions. The sentence *Whose and what are those?* contains two interrogative pronouns.

Relative Pronouns In form, **relative pronouns** resemble other kinds of pronouns, but they're used differently. Examples include *who* (in 1 below), *that* (in 2), and *which* (in 3). Other relative pronouns include *whose* and *whom*. Notice that relative pronouns are *related* to some preceding noun phrase. In the examples, the relative pronoun and the related noun phrase are italicized, with the relative pronoun underlined.

1. Amanda's a doctor who specializes in gerontology.
2. The show that won the most awards was *60 Minutes*.
3. She's a licensed masseuse, which I am not.

Indefinite Pronouns Pronouns whose referents can't be specifically identified are called indefinite, as in *Somebody ate my pizza!* Other **indefinite pronouns** are *some, someone, anyone, everyone, one, no one, somebody, anybody, everybody, nobody, something, anything, everything, nothing*.

Demonstrative Pronouns To refer to things relatively near (*this, these*) or, by contrast, relatively far away (*that, those*) when the referent can be identified by pointing or from the context of a discussion, **demonstrative pronouns** are used. Examples include *that* in *That really bothers Kyle* and *those* in *Those are Kyle's*. The forms *this* and *that* and *these* and *those* are pronouns when they are used independently of a noun, but in other contexts they are not pronouns, as we discuss just below.

Determiners

Determiners precede nouns (*an orchestra, the players, this smartphone, those guys, which film, whose iPad*), although words in some other categories can intervene (*an acclaimed orchestra, the most talented players*). Determiners do not

have endings. Despite being a category with few members, determiners fall into several subcategories:

- ◆ definite and indefinite articles: *the book, an iPad*
- ◆ demonstratives: *this book, that iPad, these shoes, those kids*
- ◆ possessives: *my, our, your, her, his, its, their*
- ◆ interrogatives: *which kind, what time, whose pizza*

Unlike nouns, adjectives, and verbs, whose members cannot be fully enumerated, determiners can be fully enumerated, as we have just done in the subcategories above. Categories whose members can be fully enumerated are called closed categories or closed classes. Note that some determiners are identical in form to pronouns, but determiners introduce nouns or noun phrases, whereas pronouns substitute for them. In *Whose is this?* *whose* and *this* are pronouns. In *Whose book is that red one?* *whose* and *that* are determiners.

Prepositions and Postpositions

Prepositions are invariant in form and constitute a category with few enough members that they can be enumerated. They typically precede a noun phrase, as in *on Tuesday, at a concert, or under the pool table*, and they indicate a semantic relationship between other entities. In *The book is on (or under or near) the table*, the preposition indicates the location of the book with respect to the table. Notice the underlined prepositions in *Tina rode to (or from) Athens* (indicating direction with respect to Athens) *with (or without) Daniel* (indicating accompaniment) *at (or near or by) her side* (indicating location of Daniel with respect to Tina).

Instead of prepositions, some languages have **postpositions**. They function like prepositions in that they indicate a semantic relationship between other entities, but instead of preceding the noun or noun phrase they follow it. Compare the Japanese postpositions with the English prepositions below:

Japanese Postpositions	English Prepositions
Taroo <i>no</i>	<i>of</i> Taro
hasi <i>de</i>	<i>with</i> chopsticks
Tookyoo <i>e</i>	<i>to</i> Tokyo

The placement of prepositions before a noun, which seems natural to speakers of English (and French, Spanish, Russian, and many other languages), would seem unnatural to speakers of Japanese, Turkish, Hindi, and many other languages that **postpose** rather than **prepose** words in this lexical category.

Conjunctions

Conjunctions are of two principal kinds. **Coordinating conjunctions** conjoin expressions of the same category or status—for example, noun phrases (*fish and wildlife, tea or coffee*), verbs (*sing and dance, trip and fall*), adjectives (*slow and painful, tough but fair*), and clauses (*she sang and he danced*). **Subordinating conjunctions** serve to link one clause to one another in a noncoordinate role,

as in *She visited Montreal twice while she attended Colby College* or *They whispered quietly because he was sleeping*. (Clauses are discussed in Chapter 5.) Subordinating conjunctions may be referred to simply as *subordinators* and coordinating conjunctions as *conjunctions* or *coordinators*.

Morphemes: Word Parts with Meaning or Function

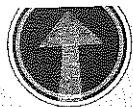
You understand that words such as *girl*, *ask*, *tall*, *uncle*, and *orange* cannot be divided into smaller meaningful units. *Orange* isn't made up of *o* + *range* or *or* + *ange*, and *uncle* isn't composed of *un* and *cle*. But most words do have more than one meaningful part. You can readily note two elements each in *grandmother*, *bookshelf*, *home-made*, *asked*, *taller*, *oranges*, and *uncles*. *Beautiful*, *supermarkets*, and *decomposed* also have more than one meaningful element. A set of words can be built up by adding elements to a core element. All of the following have *true* as a core, or root, element:

truer	untrue	truthfully
truest	truth	untruthfully
truly	truthful	untruthfulness

They share a root whose meaning or lexical category has been altered by the addition of elements. The meaningful elements in a word are called **morphemes**. Thus, *true* is a word with a single morpheme, while *untrue* and *truly* contain two morphemes each, and *untruthfulness* contains five (UN- + TRUE + -TH + -FUL + -NESS). *Truer* has the elements TRUE and -ER (meaning 'more'). The morphemes in *truly* are TRUE and -LY; in *untrue*, TRUE and UN-; in *truthful*, TRUE + -TH + -FUL.

Most morphemes have what is called lexical meaning, as with *look*, *kite*, and *tall*. Other morphemes represent a grammatical category such as past tense (-ed in *looked*), plural (-s in *kites*), or comparative degree (-er in *taller*).

Morphemes cannot be equated with syllables: they are not the same. (Until we analyze syllables in Chapter 4, you can rely on your intuitive understanding of them.) The words *harvest*, *grammar*, and *river* contain two syllables each but only a single morpheme. *Gorilla* contains three syllables, *Connecticut* four, and *hippopotamus* five, but each of these words contains only a single morpheme: they are **monomorphemic**. The other way around, a single syllable can represent more than one morpheme: *kissed* is a monosyllabic word that contains two morphemes (KISS + 'PAST TENSE'). *Dogs* (DOG + 'PLURAL') and *feet* (FOOT + 'PLURAL') are also monosyllabic and contain two morphemes each. In a single syllable, *men's* contains three morphemes (MAN + 'PLURAL' + 'POSSESSIVE').



Try It Yourself Identify all the morphemes in these words and whether they're free or bound: *bakery*, *baseball*, *borderlands*, *cider*, *dusty*, *fried*, *outlaw*, *prayer*, *prefabs*.

Morphemes Can Be Free or Bound

Morphemes like TRUE, MOTHER, and ORANGE can stand alone as words and are called free morphemes. Others,

UN-, TELE-, -NESS, and -ER, can function only as part of a word and are called **bound morphemes**.

Morphemes That Derive Other Words

Certain bound morphemes change the category of the word to which they are attached, as with the underlined parts of *doubtful*, *sweetly*, *establishment*, *teacher*, and *darken*. When added to the noun *doubt*, -FUL derives the adjective *doubtful*, as represented in Figure 2.1. When added to the adjective *sweet*, -LY derives the adverb *sweetly*; when added to the verb *establish*, -MENT derives the noun *establishment*. *Teach* is a verb, *teacher* a noun; *dark* is an adjective, *darken* a verb. English has additional derivational affixes, and derivational morphemes in English (but not all languages) tend to be added to the ends of words (they're suffixes). We can represent these relationships in the following derivational rules:

Noun + -FUL	→	Adjective (<i>doubtful</i> , <i>beautiful</i> , <i>bountiful</i> , <i>careful</i>)
Adjective + -LY	→	Adverb (<i>sweetly</i> , <i>really</i> , <i>slowly</i> , <i>quickly</i> , <i>responsibly</i>)
Verb + -MENT	→	Noun (<i>enforcement</i> , <i>advancement</i> , <i>amazement</i> , <i>commencement</i>)
Verb + -ER	→	Noun (<i>teacher</i> , <i>rider</i> , <i>glider</i> , <i>driver</i> , <i>tweeter</i> , <i>viewer</i>)
Verb + -ABLE	→	Adjective (<i>controllable</i> , <i>forgivable</i> , <i>treatable</i> , <i>enjoyable</i>)
Adjective + -ITY	→	Noun (<i>sanity</i> , <i>abnormality</i> , <i>reality</i> , <i>frugality</i> , <i>stupidity</i>)
Adjective + -EN	→	Verb (<i>darken</i> , <i>sweeten</i> , <i>brighten</i> , <i>harden</i>)
Noun + -ISH	→	Adjective (<i>childish</i> , <i>baboonish</i> , <i>sheepish</i> , <i>nannyish</i>)

A similar process adds a morpheme at the beginning of a word—as a prefix. English prefixes typically change the meaning of a word but *not* its lexical category. Thus, *align* and *realign* are both verbs with related but different meanings; *kind* and *unkind* are both adjectives with related but contrasting meanings; *spell* and *misspell* are both verbs. The structure of *realign* is illustrated graphically in Figure 2.2; the structure of *realignment*, which contains a prefix and a suffix, is shown in Figure 2.3.

RE- + Verb	→	Verb (<i>realign</i> , <i>repaint</i> , <i>rephrase</i> , <i>rewrite</i> , <i>reassess</i> , <i>retake</i>)
DIS- + Verb	→	Verb (<i>disappear</i> , <i>disconnect</i> , <i>discontinue</i> , <i>disengage</i> , <i>disinfect</i>)
MIS- + Verb	→	Verb (<i>misspell</i> , <i>misstep</i> , <i>miscalculate</i> , <i>misfire</i> , <i>misclassify</i>)
IN- + Adjective	→	Adjective (<i>incompetent</i> , <i>intolerable</i> , <i>inflexible</i> , <i>indecent</i>)
IM- + Adjective	→	Adjective (<i>imprecise</i> , <i>impatient</i> , <i>imbalanced</i> , <i>immature</i>)

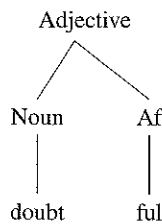


Figure 2.1 DOUBTFUL

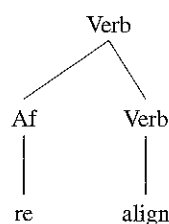


Figure 2.2 REALIGN

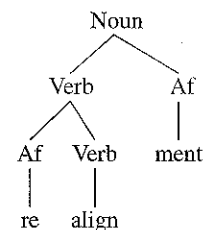


Figure 2.3 REALIGNMENT

- UN- + Adjective → Adjective (*unkind, unsafe, unfair, unreal, unimportant*)
 UN- + Verb → Verb (*undo, untie, uncover, undress, uninstall*)
 UNDER- + Verb → Verb (*underbid, undercount, underrate, underappreciate*)
 OVER- + Verb → Verb (*overeat, overestimate, overprescribe, overstimulate*)

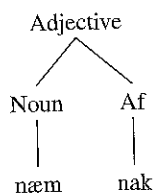


Figure 2.4
NÆMNAK

Processes of **derivation** that transform a word into another word with a related meaning but belonging to a different lexical category are common in the languages of the world. Here's an example from Persian, and the structure of *næmnaK* is illustrated in Figure 2.4. (Note: æ is pronounced like the *a* in English *hat*, *x* like the *ch* in German *Bach*.)

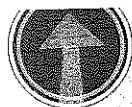
næm 'dampness'	næmnaK 'damp'
dærd 'pain'	dærdnaK 'painful'
xætær 'danger'	xætærnaK 'dangerous'

The suffix *-naK*, when added to certain nouns, derives adjectives. Thus Persian has the following derivational rule:

Noun X + -NAK → Adjective 'the quality of being or having X'

Another derivational suffix of Persian creates abstract nouns from adjectives, as illustrated in these word pairs:

gærm 'warm'	gærma 'heat'
pæhn 'wide'	pæhna 'width'



Try It Yourself Using as models the derivational rules given above, write the rule by which *rusty* is derived from *rust*, and cite two additional examples of words formed by that rule. Write the rule by which *bookish* is derived from *book*, and cite two additional examples of words formed by that rule.

This process of derivational morphology can be expressed by this rule:

Adjective + -A → Noun

Not every word belonging to the lexical category specified in a rule can undergo a given derivational process. In English, the nouns *doubt* and *beauty* become adjectives by addition of the suffix *-FUL*, but the adjective related to

rust is formed by addition not of *-FUL* but of *-Y*, and to make *book* an adjective requires the suffix *-ISH* (not *-Y*). Thus, words are marked in the mental lexicon for particular derivational processes, yielding the grammatical *rusty* and *bookish* rather than the ungrammatical forms **rustful* and **booky*.

In Fijian, *vaka-*, meaning 'in the manner of,' is a derivational morpheme that can be prefixed to adjectives and nouns to derive adverbs according to these two rules:

- VAKA- + Adjective → Adverb
 VAKA- + Noun → Adverb

Exhibiting the morpheme VAKA- are the words *vaka-Viti* 'in the Fijian fashion' (from *Viti* 'Fijian') and *vakatotolo* 'in a rapid manner, rapidly' (from *totolo* 'fast, rapid'), whose structure is shown in Figure 2.5. To illustrate derivation from a noun, consider *vakamaarama* 'lady-like' (formed by prefixing *vaka-* to *maarama* 'lady').

Adding DIS- or RE- to a verb or UN- to an adjective (*disappear*, *repaint*, *unfavorable*) changes the word's meaning but not its category. For example, *appear* and *disappear* are both verbs, as are *paint* and *repaint*; *favorable* and *unfavorable* are both adjectives. English shows a notable tendency for morphemes that change meaning without altering category to be added to the front of a word as a prefix, but this is not true of all languages, and some languages—Turkish is an example—lack prefixes altogether.

The two morpheme types we have just examined are **derivational morphemes**. They derive (or produce) new words from existing words by changing the meaning of a word (*true* versus *untrue*; *align* versus *realign*) or its lexical category (*doubt* is a noun, *doubtful* an adjective).

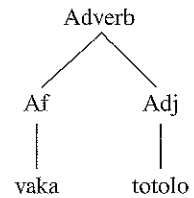


Figure 2.5 VAKATOTOLO

Inflectional Morphemes

The underlined parts of the words *cats*, *collected*, *sleeps*, and *louder* represent another type of bound morpheme. These **inflectional morphemes** change the form of a word but not its lexical category, and they don't alter a word's central meaning. Inflectional morphemes create variant forms of a word to conform to different roles in a sentence or a discourse. On nouns and pronouns, inflectional morphemes serve to mark semantic notions such as *number* or grammatical categories such as *case*. On verbs, they may mark categories such as *tense* or *number*. On adjectives they may indicate *degree*. They shape the "related forms" we talked about earlier in the chapter to help identify lexical categories, and we'll return to inflectional morphology later in this chapter.

How Are Morphemes Organized Within Words?

Morphemes Are Ordered in Sequence

Within a word, morphemes have a strict and systematic linear sequence: they are not arranged randomly.

Affixes Some morphemes, called **suffixes**, follow the stems they attach to, such as 'PLURAL' in *girls* and -MENT in *commitment*: both **sgirl* and **mentcommit* would be ill-formed. **Prefixes** attach to the front of a stem, as in *untrue*, *disappear*, and *repaint* (not the end, as with the ill-formed **trueun*, **appeardis*, and **paintre*.)

Derivational morphemes can be prefixes (*unhappy*, *disappear*) or suffixes (*happiness*, *appearance*), and some words have both, as with *unhappiness* and *disappearance*. Generally, inflectional morphemes are added to the outermost parts of words. The term **affix** is used to cover both prefixes and suffixes.

Infixes Besides affixes, some languages have infixes. An **infix** is a morpheme inserted within another morpheme. In Tagalog (a language spoken in the Philippines), you can see infixing by comparing the word *gulay*, meaning 'greenish vegetables,' with the word *ginulay*, meaning 'greenish blue,' which contains the infix -IN-.

Morphemes Can Be Discontinuous

Not all morphological processes can be viewed as joining or concatenating morphemes to one another by adding a continuous sequence of sounds (or letters) to a stem. In other words, not all morphological processes involve prefixes, suffixes, or infixes. The technical term for discontinuous morphology is *nonconcatenative*.

Circumfixes Some languages combine a prefix and a suffix into a **circumfix**—a morpheme that occurs in two parts, one on each side of a stem. Samoan has a morpheme FE-/AʔI, meaning 'reciprocal': the verb 'to quarrel' is *finau*, and the verb 'to quarrel with each other' is *fefinauaʔi*—FE + FINAU + AʔI. Indonesian has a circumfix KE-/AN that derives an abstract noun from a concrete one; thus, *baik* 'good, kind' yields *kebaikan*—KE-BAIK-AN—'kindness.' As another example, the circumfix SE-/NYA derives an adverb from an adjective, as in *benar* 'true' and *sebenarnya*—SE-BENAR-NYA—'actually.'

Interweaving Morphemes Semitic languages, including Arabic and Hebrew, can have **interweaving morphemes**. Arabic nouns and verbs generally have a root consisting of three consonants, such as KTB. The Arabic word for 'book' is *kitaab*. By interweaving K-T-B and various other morphemes, Arabic creates a great many nouns, verbs, and adjectives from this single root. All the nouns and verbs in Table 2.1 contain the same KTB root, with other morphemes interwoven.

Incidentally, the English words *Muslim*, *Islam*, and *salaam* all contain the Arabic root SLM, with a core meaning of 'peace, submission.'

Portmanteau Words Contain Merged Morphemes

Another phenomenon combines several morphemes in such a way that the sounds in the word cannot be assigned tidily to each of its morphemes. A classic example is the French word *du*, which combines the two morphemes DE 'of' and LE 'the.'

Table 2.1 Derivational Morphology in Arabic

kitaaba	'writing'	kataba	'he wrote'
kaatib	'writer'	kaataba	'he corresponded with'
maktab	'office'	?aktaba	'he dictated'
maktaba	'library'	?iktataba	'he was registered'
maktuub	'letter'	takaataba	'he exchanged letters with'
miktaab	'typewriter'	inkataba	'he subscribed'
kutubii	'bookseller'	iktataba	'he had a copy made'

You can see the difficulty of assigning the sounds to one morpheme or the other. Some analysts call blends like *smog* (from *smoke* and *fog*) **portmanteau words**.

Morphemes Are Layered Within Words

Morphemes are organized within words in highly patterned ways. Typically they have an obvious linear order (what comes first and second and so on), but they also have a hidden hierarchical or **layered structure**. Take the word *untrue*: it has the obvious linear order UN- followed by TRUE. But in terms of layers TRUE is the **root**, to which UN- is prefixed; *untrue* is not UN- with TRUE suffixed; it is TRUE with UN- prefixed. *Truthful* comprises a **stem** *truth* with -FUL suffixed to it (*truth* itself is TRUE with -TH suffixed). The morphemes in *untruthful* have the manifest linear order UN- + TRUE + -TH + -FUL. But in terms of its layers, TRUE is the root to which -TH is added; *truth* is the stem to which -FUL is added; *truthful* is the stem to which UN- is added. In other words, the layering for *untruthful* proceeds this way: *true* > *truth* > *truthful* > *untruthful*.

Now consider the word *uncontrollably*. It's helpful to picture the sequence of layering from the root morpheme CONTROL, built up by a set of derivational rules (which also apply widely to other words):

control (Verb)

Verb + -ABLE → Adjective (See Figure 2.6)

controllable (Adjective)

UN- + Adjective → Adjective (See Figure 2.7)

uncontrollable (Adjective)

Adjective + -LY → Adverb (See Figure 2.8)

uncontrollably (Adverb)

The root of *uncontrollably* is *control*, which functions as the stem for -ABLE; in turn, *controllable* functions as the stem for *uncontrollable*; finally, *uncontrollable* functions as the stem to which is added the derivational suffix -LY, which derives the adverb *uncontrollably*.

The structure of *uncontrollably* can be represented graphically in a tree diagram that shows both the linear and hierarchical structure. In Figure 2.8, look at the lowest category node, labeled Verb. Because the lowest category node in a tree diagram like this identifies the root, the verb *control* is the root of the whole word, which is built up layer by layer. To the root verb is added the affix -ABLE, yielding the adjective *controllable* (as in Figure 2.6). If instead you were to add UN- before adding -ABLE, you'd produce "uncontrol," which is not an English word. With -ABLE, the adjective *controllable* then serves as a stem to which can be added the affix UN-, producing the adjective *uncontrollable* (see Figure 2.7). The final step in the derivation adds the affix -LY to the adjective *uncontrollable*, yielding the adverb *uncontrollably*. Looking at Figure 2.8 from the top down, you can see that the Adverb node, encompassing the entire word *uncontrollably*,

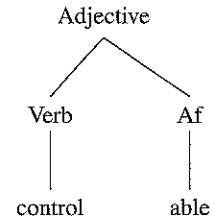


Figure 2.6 CONTROLLABLE

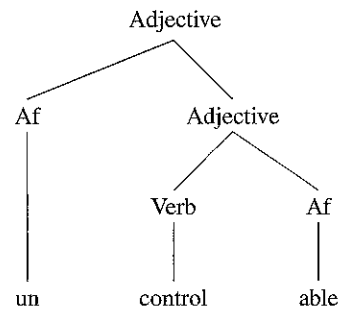


Figure 2.7 UNCONTROLLABLE

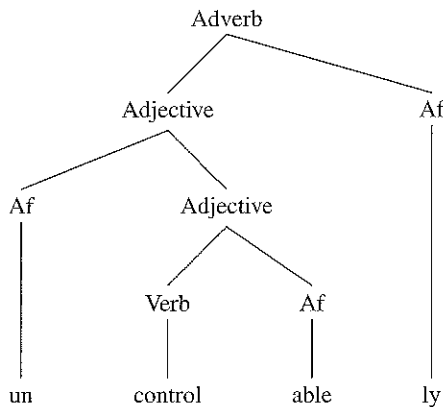


Figure 2.8 UNCONTROLLABLY

has two branches. One is labeled Adjective and encompasses *un-control-able*, and the other is labeled Af (for affix) and leads to the suffix *-LY*. In turn, the Adjective node has two branches—one leading to the affix *UN-* and the other to the Adjective node for *controllable*. Finally, the Adjective node that encompasses *control-able* has two branches—one labeled Verb and leading to *control* and the other leading to the affix *-ABLE*. The lowermost node for a lexical category (Verb) leads to the base form *CONTROL*, the root of *uncontrollably*.



Try It Yourself In the verb *misfired*, what is the linear order of morphemes?

What is its root morpheme and how many affixes does it have? Which affix is first added to the root, thus creating the stem for subsequent affixation? Is the final morpheme that is added derivational or inflectional?

How Does a Language Increase Its Vocabulary?

Languages have three principal ways of extending their vocabulary:

- ◆ New words can be formed from existing words and word parts.
- ◆ Words can be “borrowed” from another language.
- ◆ New words can be made up, created from scratch.

Some Word Classes Are Open, Others Closed

In some societies, the need for new nouns, adjectives, and verbs may arise frequently, and additions to these lexical categories occur freely. For this reason, nouns, adjectives, and verbs are called *open classes*. By contrast, prepositions, pronouns, and determiners are *closed classes*, and new words are seldom added to these categories. Century after century, English speakers have added thousands of new words, borrowing some from other languages and constructing others from already available elements.

How to Derive New Words

Affixes Adding morphemes to a word is a common way of creating new words. English has added the agentive suffix *-ER* to the prepositions *up* and *down* to create the nouns *upper* and *downer* to refer to phenomena that lift or dampen your spirits. More commonly, *-ER* is added to a verb (V) to create a word with the sense ‘one who Vs’: *campaigner* ‘one who campaigns’; *designer* ‘one who



AT THE BAR McSleep Inn or Out?

Quality Inns once wanted to launch a chain of inexpensive, standardized inns under the name *McSleep Inn*. When McDonald's Corporation objected, the case went to court (*Quality Inns Int'l. v. McDonald's Corp.*). Linguist Roger W. Shuy helped Quality Inns determine whether the *Mc* prefix had become part of everyday English, with a meaning independent of its association with McDonald's. He cited journalistic sources that, without making any reference to McDonald's, appeared to use *Mc* in a general way to mean "basic, convenient, inexpensive, and standardized." Among the examples he found were *McFood*, *McFashion*, *McFuneral*, *McLaw*, *McMedicine*, *McMovies*, *McLube*, *McMail*, *McArt*, and even *McPaper* (referring to *USA Today*, "Fast News for the Fast Food Generation"). The judge in the case quipped that "A news report . . . referred to the trial . . . as taking place in the McCourt,

which . . . would make [him] the McJudge," and in his "McPinion" he quoted an article that captured the situation: "This is the era of instant gratification, of pop tops, quick wash, fast fix, frozen foods, McEverything." On the other side, McDonald's retained a linguist named David W. Lightfoot, and he disputed the claim that *Mc* had become a generic prefix, used without reference to McDonald's and therefore available for anyone to use, including Quality Inns. He testified that the meanings Dr. Shuy uncovered were characteristics of McDonald's and its reputation, and he concluded, according to the judge, that "in every case the allusion [in Shuy's examples] was to McDonald's and its family of marks [and] intended to be cute and playful." In the end, the McJudge decided that *McSleep* would infringe on McDonald's legitimate family of trademarks, and he did not allow Quality Inns to name its new chain *McSleep Inn*. ■

designs'; *blogger* 'one who blogs'; *twitterer* 'one who twitters or tweets (on Twitter.com).'

English adds new words principally by prefixing or suffixing. Prefixes like UN-, PRE-, and DIS- alter the meaning of words but not usually their lexical category, as we saw earlier. Added to the adjectives *true*, *popular*, *successful*, and *favorable*, the prefix UN- creates new adjectives with the opposite meanings: *untrue*, *unpopular*, *unsuccessful*, *unfavorable*. Prefixed to verbs, UN- creates new verbs with the opposite meaning: *undo*, *untie*, *unplug*, *unbutton*, *uninstall*. DIS- prefixed to a verb creates a new verb with an opposite meaning: *disobey*, *disapprove*, *disappear*, *displease*, *dishonor*. PRE- serves as a prefix to several categories: verbs (*preregister*, *premix*, *preboard*, *preview*), adjectives (*pre-Copernican*, *precollegiate*, *preconscious*, *prenatal*, *prenuptial*, *presurgical*), and nouns (*precaution*, *precalculus*, *preadult*, *precancer*). PRE- has roughly the same sense in each case, and from an existing word it creates a new word belonging to the same lexical category. Recently productive prefixes include CYBER- (*cyberspace*, *cyberpal*, *cybercafé*, *cybersex*), BIO- (*biochip*, *bioethics*, *bioterrorism*, *biotechnology*, *bioweapons*), NANO- (*nanotube*, *nanosecond*, *nanotechnology*, *nanoworld*, *nanogram*), and E- (*email*, *ebook*, *ecard*, *ebuddy*, *eshop*), none of which changes the category of the stem it attaches to.

English derivational suffixes are added to the tail end of a stem. Unlike prefixes, derivational suffixes usually change the lexical category of the stem—from, say, a verb to a noun. For example, adding -MENT to a verb makes it a

noun: *agreement, assignment, establishment*. Likewise, the suffix *-ATION* derives a noun from a verb: *resignation, organization, implementation, observation, reformation* (from *resign, organize, implement, observe, reform*).

Suffixes are widely exploited in the languages of the world. The Indonesian suffix *-KAN* changes a noun to a verb, and among the various meanings it can produce are: 'to cause to become X' (*rajukan* 'to crown' from *raja* 'king') and 'to put in X' (as in *penjarakan* 'imprison' from *penjara* 'prison').

Reduplication The process by which a morpheme or part of a morpheme is repeated to create a new word with a different meaning or different category is called **reduplication**. The Mandarin Chinese word *sànsàn* 'to take a leisurely walk' is formed by reduplicating the first syllable of *sàn* 'to walk'; *hónghóng* 'bright red' is formed by reduplicating *hóng* 'red.' Partial reduplication repeats only part of the morpheme, while full reduplication reduplicates the entire morpheme. In the Papua New Guinea language Motu, the verb *mahuta* 'to sleep' reduplicates fully as *mahutamahuta* 'to sleep constantly' and reduplicates partially as *mamahuta* 'to sleep' (when agreeing with a plural subject). In Turkish, adjectives like *açık* 'open,' *ayrı* 'separate,' and *uzun* 'long' are reduplicated (by prefixing the initial vowel followed by a consonant) as *apaçık* 'wide open,' *apayrı* 'entirely separate,' and *upuzun* 'very long.' Reduplication is *not* repetition: repetition does not create a new word but simply reiterates the same word, as in English *very, very (tired), night-night, and long, long, long (shower)*. English does not have a productive process like the reduplication of Chinese, Motu, or Turkish.

Reduplication can have various functions. It can intensify the meaning of a word or moderate it, as illustrated by the Chinese, Motu, and Turkish examples just given. Alternatively, it can mark grammatical categories, as in Indonesian, where certain kinds of noun plurals are formed by reduplication: *babibabi* 'an assortment of pigs' is a reduplicated form of *babi* 'pig.'

Compounds

English speakers show a vigorous disposition to put words together to create new words in a process called **compounding**. Relatively recent compounds include *app store, date rape, upskirt, anchor baby, computerlike, dust bunny, gut-buster, electropop, shout-out, debt crisis, speed dating, euro-zone, touch screen, fat finger, download, ringtone, and Facebook*. (Notice that compounds have heavier stress or emphasis on the first element than on the second element.) To gauge the popularity of compounding, consider that a single two-page piece in an issue of *The Atlantic* contained these (and many more) examples, not all of which are brand new.

Nouns	Nouns	Nouns	Adjectives
cellphones	airtime	businesswomen	life-threatening
camera phone	shapewear	plastic surgeon	gum-chewing
nip slip	skin-care	death-rattle	beach-tawny
sideboob	lunch parties	dumbbell	hardheaded
law firm	fitness concept	nutrition bars	god-awful
lawsuit	nut job	cookbooks	sun-dazed
housewife	air kisses	ghostwriter	gossip-industrial

Bear Encounters

Bears may try to share your lunch. Your food can lure them dangerously close.

If a bear approaches, scare it away by yelling, clapping your hands, and acting aggressive toward it.

Protect your food from all park animals. Pack food and trash in closed containers.

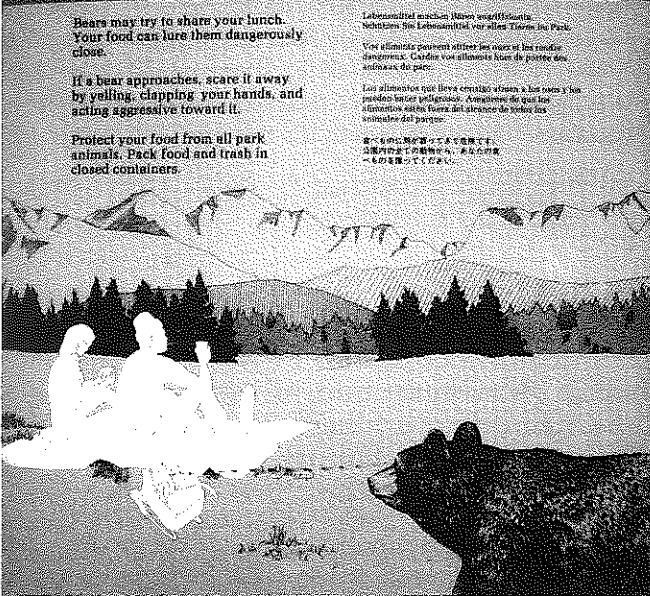
Lebensmittel machen Bären angriffslustig. Schützen Sie Lebensmittel vor allen Tieren im Park.

Vos aliments peuvent attirer les ours et les rendre dangereux. Cachez vos aliments bien de portée des animaux du parc.

Los alimentos que lleva consigo atraen a los osos y los pueden hacer peligrosos. Evitemos de que los alimentos estén fuera del alcance de todos los animales del parque.

熊は人の食べ物に興味を持ち、危険な行動をとることがあります。熊が近づいたら、大声を叫び、手を叩き、威嚇する行動をとってください。

公園の動物から食べ物を守るために、食品とゴミは必ず密閉容器に入れてください。



Bear Encounters at Yosemite National Park. A warning in five languages. Can you spot any likely compounds in the German version (top right)? *Lebensmittel machen Bären angriffslustig. Schützen Sie Lebensmittel vor allen Tieren im Park.*

Compounding is widespread in the languages of the world. Mandarin, for example, has numerous compounds, such as *fàn-wǎn* ‘rice bowl,’ *diàn-nǎo* (‘electric’ + ‘brain’) ‘computer,’ *tái-bù* ‘tablecloth,’ *fēi-jī* (‘fly’ + ‘machine’) ‘airplane,’ and *hēi-bǎn* (‘black’ + ‘board’) ‘blackboard.’ German is famous for its compounding. The word *Fernsprecher* (literally ‘far speaker’) was for a long time the preferred word for what is now usually called *Telefon*. A ballpoint pen is called *Kugelschreiber* (‘ball’ + ‘writer’); a glove *Handschuh* (‘hand’ + ‘shoe’); mayor is *Bürgermeister* (‘citizen’ + ‘master’). Indonesian has exploited compounding in a word made familiar to Westerners from its use as the assumed name of a well-known World War I socialite and spy: *matahari*, meaning ‘sun,’ comes from *mata* ‘eye’ and *hari* ‘day.’ The word for ‘eyeglasses’ is *kacamata*, a compound of *kaca* ‘glass’ and *mata* ‘eye’ (similar to the English compound *eyeglasses* but with a different order of elements).

Shortening

Shortenings are a popular means of multiplying the words of a language. Besides ordinary shortenings like *plane* for airplane and *jet* for jet airplane, we’ll examine acronyms, initialisms, and blends.

Ordinary shortenings are common. In shortenings like *radials* for radial tires, *feds* for federal agents, *indie* for independent film, and *narc* for narcotics agent, the first element of a compound or a shortened version of it remains and the second element is omitted. We can see a variant of that in *app* for computer

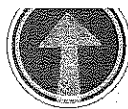
application program, where one of the words is shortened and the others omitted. In shortenings like *info* for information, *obits* for obituaries, *dorm* for dormitory, and *poli-sci* for political science, a word or parts of words in a compound are shortened. In shortenings like *rec room* ‘recreation room’ and *comp time* ‘compensatory time,’ the first (multisyllabic) word is shortened and the second (often monosyllabic) word retained. Note that the shortened forms need not be morphemes in the full words from which they come: while *jet* is a morpheme, *narc*, *fed*, *indie*, *obit*, *rec*, and *comp* are not morphemes in the full-form originals.

Acronyms Shortenings in which the initial letters of the words in an expression are joined and pronounced as a word are **acronyms**:

UNESCO	NATO	radar (radio detecting and ranging)
WI-FI	NASA	laser (light amplification by stimulated emission of radiation)

Initialisms Some shortenings resemble acronyms but are pronounced as a sequence of letters. At *U-S-C* (University of Southern California) and *N-Y-U* (New

York University), a student’s grade-point average may be called a *G-P-A*. *PC* carries two meanings—‘politically correct’ and ‘personal computer.’ Given their pronunciation as a series of letters, these are called *initialisms*. Many initialisms (*AI*, *CD*, *DNA*, *DVD*, *fMRI*, *FTC*, *MTV*, *NHL*, *PDA*, *IED*, *GPS*, *LCD*) could not easily be pronounced as



Try It Yourself Among these, which are acronyms? EU, USA, FBI, CIA, NYPD, ER, DARPA, FWB, LOL, MBA, POTUS, VIP, DJ, ATM, IT, IM, iPod, SNAFU, TGIF, TMI.

ordinary words, while others could but aren’t: *COO* for chief operating officer, *ADD* for attention deficit disorder, *SUV* for sport utility vehicle, and *MO* for modus operandi. Perhaps the most popular initialism in the world is *OK*, often spelled *okay*.

Blends Blends are words created by combining usually the first sounds of one word with the final sounds of another. *Smog* (from *smoke* and *fog*), *glob* (*gob* and *blob*), and *motel* (*motor* and *hotel*) are well-established examples. Newer ones include *fanzine* (*fan* and *magazine*), *punkumentary* (*punk* and *documentary*), *infomercial* (*information* and *commercial*), *agribot* (*agriculture* and *robot*), *glitterati* (*glitter* and *literati*), and *celanthropist* (*celebrity* and *philanthropist*).



Try It Yourself Identify the words whose parts have been combined into these

blends: *affluent* (*affluent* and *influential*), *andropause*, *automagically*, *beefalo*, *Bollywood*, *botel*, *brainiac*, *Chunnel*, *cremains*, *cybrarian*, *digicam*, *emoticon*, *exploitainment*, *gasohol*, *gaydar*, *gimmes*, *guesstimate*, *himbo*, *murderabilia*, *prebuttal*, *sexpert*, *sexting*, *wannabes*.

netizens and *netiquette* blend *net* (a shortened form of *Internet*) with *citizens* and *etiquette*. Combining the existing blend *smog* with the tail end of *metropolis* forms *smogopolis*. Blends like *Spanglish*, *Franglais*, *Chinglish*, and *Yinglish* suggest how heavily some language varieties have borrowed words and other elements from other languages. Some blends combine the beginning sounds

or syllables of two words, as with *modem*, which is well known although its elements are not (*modulator* and *demodulator*), and *biotech* (*biology* and *technology*). Blends frequently serve as trade names and names of related products: *Amtrak*, *Eurailpass*, *Eurorail*, *Eurotrip*, and *Flexipass*. Blends often combine two nouns, but *Flexipass* combines the adjective *flexible* with the noun *pass*, and other blends combine other categories, as in *whatchamacallit*.

Back Formation

Another type of word formation is exemplified by “pronunciate,” which an occasional college student can be heard to say when searching for the verb corresponding to the noun *pronunciation* (and likely being influenced by the actual verb *enunciate*). The new verb they have “back formed” from the noun may not be sufficiently established to be entered in dictionaries, but other back formations are well established, including the verbs *typewrite*, *baby-sit*, *spectate*, and *edit*, back formed from the nouns *typewriter*, *baby-sitter*, *spectator*, and *editor*.

Conversion or Functional Shift

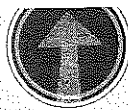
In some languages, a word belonging to one category can be converted to another category without any changes to the word's form. This is called **conversion** or functional shift. We request someone to *update* (verb) a report and call the revised report an *update* (noun). We ask a fellow worker to *email* or *fax* the report, both of which are verbs converted from shortened forms of nouns (*electronic mail*, *facsimile*). Companies *hire* (verb) employees and call them new *hires* (noun). To promote a product in the *market*, we *market* it. Conversion of this type commonly leads to noun/verb or noun/adjective pairs, as in Table 2.2, which also illustrates that occasionally the same form can serve as noun, verb, and adjective. Once a form has been shifted to a new lexical category, it conforms to the inflectional morphology of that category: an *update*, several *updates*, she's *updating* the report now, he *updated* it yesterday.

Semantic Shift

Words can take on new meanings by extending or shrinking the scope of their reference. Two well-known examples of semantic shift have remained popular since the Vietnam War, when *hawk* came to be used frequently for supporters of the war and *dove* for its opponents,

Table 2.2 Some English Forms Belonging to More Than One Lexical Category

Noun	Verb	Adjective
email	email	
bookmark	bookmark	
party	party	
outrage	outrage	
friend	friend	
delay	delay	
plot	plot	
blog	blog	
text	text	
tweet	tweet	
local		local
inaugural		inaugural
illegal		illegal
average	average	average



Try It Yourself The Oxford English

Dictionary recently added extended meanings for the nouns *coyote*, *crack*, and *thumbnail* and for the verb *zone* (*out*). Compared to the traditional meanings 'wolflike animal,' 'fissure or break,' 'nail of the thumb,' and 'divide into zones,' can you provide glosses for newer meanings of these terms?

extending the meaning of these words from the combative nature of hawks and the symbolically peaceful role of doves. Today, computer users utilize a *mouse* and *bookmark* Internet addresses. These new meanings did not replace earlier ones but extended the range of application for the words *mouse* and *bookmark*. Called *semantic shift* or *metaphorical extension*, this phenomenon creates *metaphors*. Over time the metaphorical origins of words can fade, as

in the meanings of the underlined parts of these phrases: to derail congressional legislation, a buoyant spokesman, an abrasive chief of staff, to sweeten the farm bill with several billion dollars to skirt a veto fight.

Borrowed Words

"Neither a borrower nor a lender be," Shakespeare advised, but speakers pay little heed when it comes to language. English has been extraordinarily receptive to *borrowed words*, accepting words from about a hundred languages in the last hundred years. As in most of its history, English borrowed more from French during the twentieth century than from any other language. Following French at some distance were Japanese and Spanish, Italian and Latin and Greek, German, and Yiddish. In smaller numbers, English is now host to words borrowed from Russian, Chinese, Arabic, Portuguese, and Hindi, as well as from numerous languages of Africa and some Native American languages.

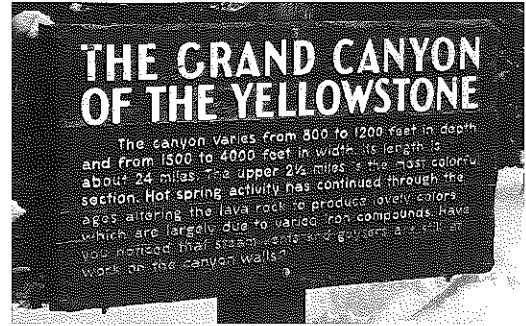
In turn, many languages have welcomed English words into their stock, although some cultures resist borrowings. The Japanese have drafted the words *beesbooru* 'baseball,' *futtobooru* 'football,' and *booringu* 'bowling' along with the sports they name, trading them for *judo*, *jujitsu*, and *karate*, which have joined the English-language team. Officially at least, the French are not open to borrowings, especially from English, and have banned the use of words like *weekend*, *drugstore*, *brainstorming*, and *countdown*. For using the borrowed term *jumbo jet*, Air France was given a stiff fine by the French government, which had insisted that *gros porteur* was the proper French name for, well, for the jumbo jet. The Americanism *OK* is now in use virtually everywhere, as are terms such as *jeans* and *discos*, which accompanied the named items as they spread around the globe. More recently, the terms *Internet*, *web*, and *tweet* have likewise circled the globe.

You can readily recognize as borrowings such popular words as *paparazzi*, *karaoke*, and *résumé*. Among borrowed nouns having to do with food and drink are *hummus* (from Arabic), *aioli* (from Provençal), *mai tai* (from Tahitian), *calzone*, *focaccia*, and *pizzetta* (from Italian), and *burrito*, *enchilada*, *fajita*, and *taco* (from Spanish). Yiddish has given us the more general term *nosh*. Other popular borrowings include the Cantonese *wok*, German *glitch*, Italian *ciao*, Spanish *macho*, *pronto*, and *mañana*, and Yiddish *chutzpah*, *klutz*, *nebbish*, *schlep*, and *schlepper*. As is true of other languages, most borrowings into English have been nouns,

but some adjectives and a few verbs, adverbs, and interjections have been borrowed.

Borrowed words sooner or later conform to the pronunciation patterns and grammatical rules of the borrowing language. They undergo the same processes that affect native words. *Nosh* was borrowed as a verb that could not take an object (*I feel like noshing*) but has since taken on new use as a transitive verb—one that can take an object (*Let's nosh some hot dogs*). The verb *nosh* with the suffix *-ER* produces the noun *nosh-er* 'one who noshes,' and *nosh* itself can be used as a noun

meaning 'a snack.' In Britain, *nosh* has been compounded into the noun *nosh-up*, meaning 'a large or elaborate meal.' In Los Angeles, a sign draped across a restaurant undergoing a change of cuisine announced "Burritofication in Progress," signaling the opening of a Mexican restaurant by punning on *beautificalion* and mimicking the morphological processes that created it.



© Edward Fiegan

Grand Canyon of the Yellowstone (See Ex. 2-13)

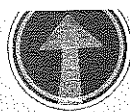
Inventing Words

The advantages of using familiar elements in forming new words and the ease of borrowing from other languages are obvious. As a result, inventing words from scratch is not common. Invention has contributed such words as *granola*, *zap*, and *quark* to the English word stock. *Nerd* appears to have been invented by Dr. Seuss and *bandersnatch*, *galumph*, and *snark* by Lewis Carroll. *Gizmo*, *pizzazz*, and *lol-lapalooza*, whose origins are unknown, may also be invented words. Some names like *Pyrex*, *Kodak*, *Teflon*, and *Xerox* are invented precisely to serve as trademarks for particular products or companies. But speakers and writers sometimes

Jeans and Discos

The *jean* in your favorite blue jeans is a form of the word *Genoa* borrowed into Middle English (around the time of Chaucer). *Jeans* is a shortening of *jean fustian* 'Genoa fustian,' referring to a coarse cloth once produced in Genoa, Italy. The word *denim*, for the cloth from which jeans are made, evolved from *serge de Nîmes*, a cloth product from the French city of Nîmes. You might wear your favorite *jeans* to a *disco*. The French word *discothèque* 'record library' is a compound of two French morphemes: *DISQUE* meaning 'disk' or 'record' and the suffix *-THÈQUE*, as in *bibliothèque* 'library.'

Discothèque appears to have been first used in English in 1954 and in its shortened form *disco* ten years later. The noun *disco* then underwent a functional shift and was used as a verb, meaning 'dance to disco music,' a use first noted in 1979. Both noun and verb can be heard around the world in cities whose residents speak neither English nor French.



Try It Yourself Britain's Prince

Charles once fretted about a threat to “proper English” (by which he meant *English* English) from the spread of *American* English around the globe. Referring to American English, he said, “People tend to invent all sorts of nouns and verbs, and make words that shouldn’t be.” What do you think of his claim that some invented nouns and verbs “shouldn’t be”?

extend the use of brand names beyond reference to the branded products. As a result, *xerox* is sometimes used to mean ‘photocopy,’ not necessarily using a machine made by Xerox. Likewise, when students *google* something on the Internet, they are not necessarily using Google’s search engine to do it. By 2010 the verb *google*, meaning ‘to search the Internet,’ had become so popular that members of the American Dialect Society voted it the word of the decade. As their word of the year, that same group

chose *tweet* (noun ‘a short message sent via the Twitter.com service’ and verb ‘the act of sending such a message’).

What Types of Morphological Systems Do Languages Have?

You have now seen examples of derivational morphology and inflectional morphology in several languages. But not all languages have inflectional morphology, and some have little or no morphology at all. Still others have complex words with distinct parts, each part representing a morpheme. Some languages are mixed in the kinds of morphology they use.

Isolating Morphology

Chinese is a language with isolating morphology—in which each word tends to be a single isolated morpheme. An isolating language lacks both derivational and inflectional morphology. Using separate words, Chinese expresses certain content that an inflecting language might express with inflectional affixes. For example, whereas English has an inflectional possessive (*the boy’s hat*) and a so-called analytical possessive (*hat of the boy*), Chinese permits only *hat of the boy* possessives. Chinese also does not have tense markers, and on pronouns it does not mark distinctions of gender (*he/she*), number (*she/they*), or case (*they/them*). Where English has six words—*he, she, him, her, they, and them*—Chinese uses only a single word, though it can indicate plurality with a separate word. The sentence below illustrates the one-morpheme-per-word pattern typical of Chinese:

wǒ gāng yào gěi nǐ nà yì bēi chá
 I just will give you that one cup tea
 ‘I am about to bring you a cup of tea.’

Even more than Chinese, Vietnamese approximates the one-morpheme-per-word model that characterizes isolating languages. Each word in the sentence below has only one form. You can see that the word *tôi* is translated as *I, my, and we*. Note that to say ‘we,’ Vietnamese pairs *chúng* and *tôi* (the words

for 'PLURAL' and 'I'). Like Chinese, Vietnamese lacks tense markers on verbs and case markers on nouns and pronouns, as well as number distinctions (though, like Chinese, it can indicate plurality with a separate word).

khi tôi đến nhà bạn tôi, chúng tôi bắt đầu làm bài
 when I come house friend I plural I begin do lesson
 'When I came to my friend's house, we began to do lessons.'

Some languages that tend to minimize inflectional morphology nevertheless exploit derivational morphology to extend their word stocks in economical ways. Indonesian, for example, has only two inflectional affixes, but it utilizes about two dozen derivational morphemes, some of which we've seen in this chapter.

Agglutinating Morphology

In agglutinating languages, words can have several prefixes and suffixes, but they are characteristically distinct and readily segmented into their parts—like English *announce-ment-s* or *pre-affirm-ed* but unlike *sang* (SING + 'PAST') or *men* (MAN + 'PLURAL'). Turkish (an Altaic language) and Swahili and Gikuyu (both Bantu languages) illustrate agglutinating morphology, as in the following examples. (Hyphens represent morpheme boundaries *within* a word.)

Turkish (an Altaic language)

herkes ben üniversite-ye başla-yacağ-ım san-ıyor
 everyone I university-to start-FUTURE-I believe-PRESENT
 'Everyone believes that I will start university.'

Swahili (a Bantu language of eastern Africa)

h-a-fany-I	a-si-fany-e
NEG-he-do-PRES	he-NEG-do-SUBJUNC
's/he isn't doing (it)'	's/he shouldn't do (it)'
a-me-fanya	a-ta-fanya
he-PERF-do	he-FUT-do
's/he has done (it)'	's/he will do (it)'
a-fany-e	h-a-ta-fanya
he-do-SUBJUNC	NEG-he-FUT-do
's/he should do (it)'	's/he won't do (it)'

Gikuyu (a Bantu language of Kenya)

ĩ-ngĩ-ka-na-endia
 I-HYPOTHETICAL-FUTURE-INDEFINITE TIME-sell
 'if I should ever sell (it)'

Inflectional Morphology

Many languages have large inventories of inflectional morphemes. Finnish, Russian, and German maintain elaborate inflectional systems. By contrast, over

Table 2.3 Inflectional Morphemes of English

Lexical Category	Grammatical Category	Examples
Noun	Plural	cars, churches
Noun	Possessive	car's, children's
Verb	Third person	(she) swims, (it) seems
Verb	Past tense	wanted, showed
Verb	Past participle	wanted, shown (or showed)
Verb	Present participle	wanting, showing
Adjective	Comparative	taller, sweeter
Adjective	Superlative	tallest, sweetest

the centuries English has shed most of its inflections, until today it has only eight remaining ones—two on nouns, four on verbs, and two on adjectives, as shown in Table 2.3. When new nouns, verbs, and adjectives are added to English or when a child learns new words, the words are extremely likely to be inflected like the examples listed, and the eight inflectional morphemes of English are thus said to be *productive*.

Compare this inflectional system of English with the examples from the Russian noun *žena* 'wife' and verb *pisat* 'to write' in Tables 2.4 and Table 2.5.

Grammatical Functions of Inflections Consider the sentences below. They contain exactly the same words, but they express different meanings.

1. The farmer saw the wolf.
2. The wolf saw the farmer.

These sentences illustrate how English exploits word order to express meaning: different orders communicate different scenarios about *who* did what to *whom*. When semantic facts such as *who* did what to *whom* are expressed by word order rather than by inflection, it is not a morphological matter but a syntactic one, and syntax is the subject of Chapter 5.

Table 2.4 Russian Noun Inflections: žena 'wife'

Case	Singular	Plural
Nominative	žena	ženy
Accusative	ženu	žen
Genitive	ženy	žen
Dative	žene	ženam
Instrumental	ženoy	ženami
After some prepositions	žene	ženax

Table 2.5 Russian Present-Tense Verb Inflections: pisat 'write'

Person	Singular	Plural
First person	pišu	pišem
Second person	pišeš	pišete
Third person	pišet	pišut

A comparison of word order. German could arrange others):

Agricola v
Lupum vidit

Latin does not have inflections of word orders. I saw the wolf

Agricola
FARMER

Lupum
WOLF

Agricola
FARMER

To say,

Agricola
FARMER

Lupus
WOLF

Agrico
FARME

The infle
The infle

A loo
noun us
one an

She

He p

Engl
nouns

case (s

mon c

except

prepos
Lat

nomir
rect o

A comparison with Latin is enlightening because Latin has relatively free word order. Given that *agricola* means 'farmer' and *lupum* 'wolf,' a Latin speaker could arrange sentence 1 ('The farmer saw the wolf') in these two ways (among others):

Agricola vidit lupum.

Lupum vidit agricola.

Latin does not rely on word order to signal *who* is seeing *whom*. Instead, inflections on the nouns signal that information. Irrespective of their different word orders, the following three sentences all mean the same thing: 'The farmer saw the wolf.'

<u>Agricola</u>	vidit	<u>lupum.</u>	}	'The farmer saw the wolf.'
FARMER	SAW	WOLF		
<u>Lupum</u>	vidit	<u>agricola.</u>		
WOLF	SAW	FARMER		
<u>Agricola</u>	<u>lupum</u>	vidit.		
FARMER	WOLF	SAW		

To say, instead, 'The wolf saw the farmer' would require different inflections:

<u>Agricolam</u>	vidit	<u>lupus.</u>	}	'The wolf saw the farmer.'
FARMER	SAW	WOLF		
<u>Lupus</u>	vidit	<u>agricolam.</u>		
WOLF	SAW	FARMER		
<u>Agricolam</u>	<u>lupus</u>	vidit.		
FARMER	WOLF	SAW		

The inflectional suffixes *-a* on *agricola* and *-us* on *lupus* identify them as subjects. The inflections *-am* and *-um* on *agricolam* and *lupum* make them direct objects.

A loose English parallel to Latin noun inflections can be seen in certain pronoun uses, where the form of the pronouns and the order of the words reinforce one another:

She praised him. (*She* is the subject, *him* the object.)

He praised her. (*He* is the subject, *her* the object.)

English and Latin nouns have inflections for number and case. English nouns exhibit only two cases, called possessive and common. The possessive case (sometimes called genitive) is marked by a suffix (*cat's*, *robot's*). The common case is unmarked (*cat*, *robot*) and is used for all grammatical functions except possession: subject, direct object, indirect object, and object of a preposition.

Latin, too, had a genitive case, and inflections for several other cases, notably nominative (principally for subjects), dative (indirect objects), accusative (for direct objects and objects of some prepositions), and ablative (for objects of other

Table 2.6 *Paradigms for Two Latin Nouns*

Singular	'Farmer'	'Garden'
Nominative	agricola	hortus
Accusative	agricolam	hortum
Genitive	agricolae	horti
Dative	agricolae	hortō
Ablative/instrumental	agricolā	hortō
Plural	'Farmer'	'Garden'
Nominative	agricolae	horti
Accusative	agricolās	hortōs
Genitive	agricolārum	hortōrum
Dative	agricolis	hortis
Ablative/instrumental	agricolis	hortis

Table 2.7 *Paradigms for Two English Nouns*

Singular		
Common	farmer	garden
Possessive	farmer's	garden's
Plural		
Common	farmers	gardens
Possessive	farmers'	gardens'

the plural. Third-person pronouns have distinct masculine, feminine, and neuter forms in the singular but make no distinction for gender in the plural.

Table 2.8 *Paradigms for First- and Third-Person Pronouns in English*

Singular	FIRST PERSON	THIRD PERSON		
		Masculine	Feminine	Neuter
Common	I	he	she	it
Possessive	mine	his	hers	its
Objective	me	him	her	it
Plural				
Common	we	they		
Possessive	ours	theirs		
Objective	us	them		

prepositions). Latin generally had five or six case inflections in the singular and in the plural, although some inflectional forms were pronounced alike, as can be seen in Table 2.6.

The set of forms constituting the inflectional variants of a word is known as a **paradigm**, and paradigms for nouns are called **declensions**. Latin had several declensions, such as the two given for *agricola* and *hortus* in Table 2.6. The paradigms for the equivalent English words *farmer* and *garden* appear in Table 2.7.

Notice that the four written forms in the English paradigms represent only two distinct pronunciations because *farmers*, *farmer's*, and *farmers'* are pronounced alike, and so are *gardens*, *garden's*, and *gardens'*. Spoken English usually has only two forms of a regular noun, but irregularly formed plurals may have four spoken and four written forms: *man*, *man's*, *men*, *men's*; *child*, *child's*, *children*, *children's*.

Some English pronouns have a third form for the objective case. In Table 2.8, you can compare the paradigms for first-person and third-person pronouns. First-person pronouns show three distinct case forms in the singular and three in

Table 2.9 *Second-Person and Third-Person Pronouns Compared to Nouns in English*

Singular	PRONOUNS		NOUNS
	Second	Third	
Common	you	it	farmer
Possessive	yours	its	farmer's
Plural			
Common	you		farmers
Possessive	yours		farmers'

The second-person pronoun (*you*) and third-person singular neuter pronoun (*it*) do not have distinct objective forms, as can be seen in Table 2.9. Instead, they have only two forms, just like regular nouns.

Gender and Agreement In English, gender distinctions in pronouns are based on biological sex: reference to males requires the masculine pronouns *he*, *his*, or *him*, while reference to females requires the feminine pronouns *she*, *hers*, or *her*. To refer to something neither male nor female, English speakers say *it*. In German, French, Spanish, Russian, Old English, and many other languages, nouns do not have biological gender but grammatical gender. In these languages, certain other word categories such as determiners and adjectives that occur within a noun phrase carry inflections that *agree* with the noun in gender, number, and case.

In contrast to the English definite article *the* (with a single written form representing the two pronunciations “*thuh*” and “*thee*”), the German definite article has forms for three genders and four cases in the singular, though there are no distinct gender markers in the plural, as Table 2.10 illustrates.

French and Spanish also exhibit variant forms of the definite article, though neither is as varied as German. French distinguishes only two noun genders; it marks masculine nouns with the definite article *le* or the indefinite *un* and feminine nouns with the definite article *la* or indefinite *une*; the plural form of the definite article for both genders is *les*. Gender in Spanish is marked in both the singular and the plural. Table 2.11 gives examples in French and Spanish.

Table 2.10 *Paradigm for German Definite Article*

	SINGULAR			PLURAL
	Masculine	Feminine	Neuter	All Genders
Nominative	der	die	das	die
Accusative	den	die	das	die
Genitive	des	der	des	der
Dative	dem	der	dem	den

Table 2.11 *French and Spanish Definite Articles with Nouns*

	French	Spanish	
Masculine	le chat	el gato	'the cat'
	les chats	los gatos	'the cats'
Feminine	la maison	la casa	'the house'
	les maisons	las casas	'the houses'

There is not always a strict demarcation between agglutinating and inflectional languages, and some languages are difficult to classify. Still, the distinction among inflectional, isolating, and agglutinating is useful in characterizing languages with respect to their morphological systems.

Variant Pronunciations of a Morpheme: Allomorphy

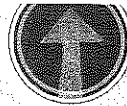
A given morpheme may not be pronounced the same in all its occurrences. For example, the morpheme METAL is pronounced one way in "a *metallic* silver" and another way in "the clang of *metal*." The pronunciation of the vowel represented by <a> differs in *metallic* and *metal*, and in American English so does the sound represented by <t>. Consider that the final sound in *house* is [s] and in *wife* is [f], but in *houses* the HOUSE morpheme ends in [z], and in *wives* the WIFE morpheme ends in [v]. Variant pronunciations of a morpheme are called **allomorphs**, and allomorphic variation is widespread in English and some other languages. (Although English spelling captures the alternate pronunciations of *wife/wives*, it doesn't acknowledge the different allomorphs of HOUSE in *house/houses* or of METAL in *metal/metallic*. In fact, while some people complain about it, others see the absence of spelling variation among allomorphs of the same morpheme as a big advantage for English spelling and for speedy reading.)

Variation in the pronunciation of morphemes isn't limited to free morphemes like METAL, WIFE, and HOUSE. It can affect any morpheme, including affixes. The English past-tense morpheme is pronounced [t] in *pick-ed*, [d] in *play-ed*, and a third way in *twist-ed*. Which pronunciation goes with which verb depends on the final sound of the verb stem (a matter we explore in Chapter 4). Children unconsciously learn the rules of allomorphic variation at an early age. If you taught English-speaking children a set of fictitious verbs they'd never heard before such as "plick" and "tevin," they'd pronounce the past-tense forms by adding [t] to "plick" and [d] to "tevin" in compliance with those unconscious rules.

The English plural morpheme on nouns also varies. The plural of *cups* and *cats* is pronounced with a final [s], but *pads* and *rags* end with [z], and *churches*, *dishes*, and *bridges* end differently again. These rule-governed pronunciations of the plural morpheme depend systematically on the final sound of the noun stem, and children who were taught imaginary nouns (like "wuck," "lutt," "mub," and "wug") would add the correct plural allomorph to the invented words: [s] to "wuck" and "lutt" and [z] to "mub" and "wug." Of course, not all English

nouns form their plural in these “regular” ways. Besides the rule-governed allomorphs, some nouns form plurals irregularly. The plural allomorphs of *deer* and *fish* and of *man*, *ox*, and *tooth* are irregular and must be learned individually. (Such irregular forms typically have an explanation in a word’s history, as discussed in Chapter 13.)

Most English verbs form the past tense by rule, and new verbs tend to form the past tense in the “regular” way. But some of the most common verbs don’t follow regular patterns and must be learned one by one, as with *teach* and *taught*, *make* and *made*, *speak* and *spoke*, *sleep* and *slept*, *tell* and *told*, and *go* and *went*. Children too young to have mastered the irregular past-tense forms are likely to say “teached,” “maked,” “speaked,” “sleped,” “telled,” and “goed,” treating them as though they followed the regular rules.



Try It Yourself

Identify which verbs have an irregular past-tense allomorph: *swim*, *head up*, *read*, *reach*, *break*, *bake*, *melt*, *say*, *see*, *meet*, *repeat*, *lead*, *eat*, *try*, *shave*, *bleed*. Identify which nouns have an irregular plural allomorph: *mouse*, *house*, *hen*, *egg*, *boy*, *girl*, *woman*. Identify which nouns have more than one allomorph: *moth*, *calf*, *child*, *woman*, *music*.

USING COMPUTERS TO EXPLORE THE WORLD IN WORDS: AMAZING CULTUROMICS!

In a corpus like the Brown Corpus (described in the “Computers and Linguistics” section of Chapter 1), the most frequent and least frequent word forms can be identified in the whole corpus or in any genre. Not surprisingly, three of Brown’s four most frequent words are the function words *the*, *of*, and *and*, and they occur in all genres, in fact in all 500 texts. By contrast, *oblong*, *obstinate*, *narcosis*, *mystification*, and many other lexical words occur only once in Brown’s million words.

Individual words in a corpus are often tagged with grammatical and other information, for example about their lexical category (noun, verb, adjective). Tagging makes it possible to analyze characteristics of words carrying any shared tag. Automated tagging has been applied to increasingly larger corpora with rates of accuracy for lexical tags around 95 percent. Direct automated tagging of lexical category simply transfers tags from a computerized reference dictionary to as many words as possible in the corpus. *Lexical* and *frequent* would receive the tag adjective, *identify* and *write* the tag verb, and so on.

Some word forms cannot be directly tagged because they belong to more than one category (as illustrated in Table 2.2). Of course, in context every form belongs to only a single category, and that means

tagging of such ambiguous forms can be aided by recognizing the category of surrounding words. Take the expression *a big deal*. Automatic pairing with dictionary entries would tag the first word (*a*) as a determiner and the second word (*big*) as an adjective but would be unable immediately to assign a tag of noun or verb to *deal* because it could be either, depending on context. Given the determiner and adjective sequence immediately preceding it, however, *deal* is very likely a noun, and it would be so labeled in that context. In other words, once word forms with unambiguous tags are tagged, those tags can help decide the correct tag for forms that belong to more than one category.

When linguistic corpora (the plural of *corpus*) began to be compiled in the 1960s and 1970s, well before the age of the 128- or 250-gigabyte USB flash drive, a one-million-word corpus like the Brown required so much computer storage space that, even as late as the 1980s, research centers kept them only on reels of magnetic tape that had to be mounted temporarily on mainframe computers for the time during which research was carried out. By the 1990s a corpus of 100 million words was thought to be big, and matters have escalated since then. We’ll discuss



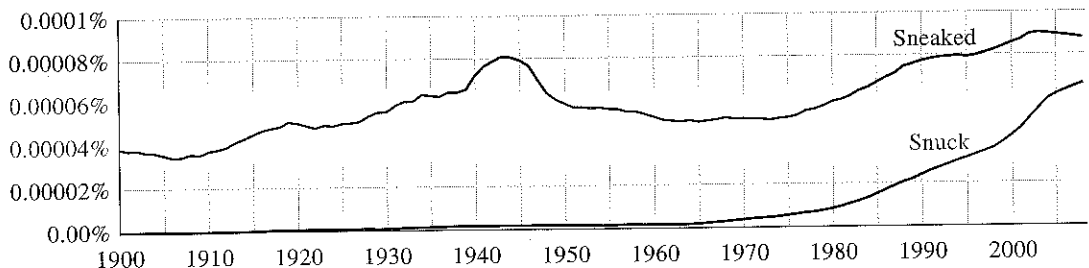
particular corpora in later chapters. Meanwhile, here it's sufficient to mention several corpora and their staggering sizes compared to the earliest million-word Brown Corpus. The Corpus of Contemporary American English contains 450 million words; the Corpus of Historical American English 400 million words; the Corpus of American Soap Operas 100 million words. Corpora of other languages are increasingly available, including a 100-million-word Corpus del Español and a 45-million-word Corpus do Português. And all of these—and more—are freely available to you for exploration on your desktop, laptop, or notebook computer. (You'll find information about these corpora in the "Internet and Other Resources" section below.) Even these mammoth corpora, however, are dwarfed by the Google Books project. Despite being older than the so-called touch-screen generation, college students are familiar with the extraordinary versatility of computers, smartphones, tablets, and the like. Still, Google has digitized millions of books—so many *billions* of words that the letters in those words, if laid end to end, would reach to the moon and back many, many times. No two ways about it: that's awesome!

In a remarkable development—one whose benefits you can experience yourself—Google was persuaded by a group of young researchers to make available not electronic versions of the books themselves (that threatened copyright privileges) but sets of n-grams of all those books. N-grams are nothing more than word sets, and Google provided complete sets of unigrams, bigrams, and so on, up to 5-grams for millions of books. Because a unigram (1-gram) is any single word (*war*; *Wagner*; *linguistics*), a complete set of unigrams is the set of all the words in the books. Bigrams (2-grams) are pairs of adjacent words (*digitized books*; *Abraham Lincoln*;

would include; *neither of*); 3-grams are sets of three adjacent words; 4-grams would include *millions of digitized books* and *complete set of unigrams*. Since late 2010 researchers have been exploring these n-grams for insight into history, culture, society, and language. Because the words are tagged not only for lexical category but for date and place of publication, the Google Ngram Viewer (see below) will show the frequency in printed books of any n-gram over the past few centuries, and it can distinguish British English from American English and fiction from nonfiction or treat them as a single set of n-grams.

As of October 2012, Google had digitized 20 million books. Relying on the n-grams, the young scholars published an article in the journal *Science* in which they explored "culturomics," a word intended to suggest genomics and probably not yet in your dictionary. As every college student can appreciate, quantitative analysis and the humanities are generally not regarded as the best of buddies, but in that first article the researchers tackled several fascinating cultural and social issues.

As a single, simple example of what a Google Ngram Viewer output looks like, examine the one accompanying the *sneaked/snuck* discussion below in "What Do You Think? REVISITED." Time is represented along the horizontal axis, and the numbers along the vertical axis represent the percentage of all unigrams that *sneaked* and *snuck* represent at any point in time. Explore a few n-grams that interest you by entering them into the search box at <http://books.google.com/ngrams/>. In subsequent chapters, we'll return to the Ngram Viewer. Meanwhile, you can find out anything you need to know by clicking on "About Ngram Viewer" at the bottom right-hand side of any Ngram Viewer page or at <http://books.google.com/ngrams/info>. ■



Google Ngram: *Sneaked* and *Snuck* since 1900

Summary

- A morpheme is a minimal linguistic unit that has a meaning or grammatical function associated with it.
- Words can contain a single morpheme (*camel*, *swim*) or several (*bookshops*, *premeditation*).
- In the mental lexicon, each morpheme contains information about sounds, related words, phrasal co-occurrence patterns, and meaning.
- Free morphemes are those that can occur as independent words: CAR, HOUSE, FOR.
- Bound morphemes cannot occur as independent words but must be attached to another morpheme: CAR + -S, LOOK + -ED, ESTABLISH + -MENT.
- Bound morphemes can mark nouns for number (e.g., 'PLURAL') and case (e.g., 'POSSESSIVE') and verbs for tense (e.g., 'PAST') or person (e.g., 'THIRD PERSON').
- Bound morphemes can derive different words from existing morphemes; for example, UN- (*untrue*), DIS- (*displease*), and -MENT (*commitment*).
- Bound morphemes can be affixes (prefixes or suffixes), infixes, or circumfixes.
- In words, morphemes have significant linear and hierarchical structures.
- The array of morphological processes for increasing a language's word stock may include compounding, reduplication, affixation, and shortening.
- Languages borrow words from other languages and sooner or later submit the borrowed words to their own pronunciation patterns and morphological processes.
- Among the types of morphological systems are inflectional, isolating, and agglutinating systems.
- Isolating systems (e.g., Vietnamese) tend to have one morpheme per word.
- Agglutinating systems (e.g., Turkish) tend to have distinct affixes.
- Corpus study is useful in showing the distribution of categories of words and morphemes as well as particular words and morphemes in different genres of text, information that has been crucial in designing automatic speech recognition systems.
- *Collocation* is the term used to refer to co-occurrences of a word with other words.
- Words in a corpus can be automatically tagged for lexical category, although several rounds of tagging that rely successively on unambiguous tags may be needed to tag all words.



What Do You Think? REVISITED

- *Maked a cake.* Most English verbs form the past tense by rule, but some of the most common ones have irregular past-tense forms. Relatively early, children learn the rule for forming past tenses of regular verbs (see Chapter 14) and tend to form all verbs the same way. Instead of using the irregular past-tense forms *made* and *told*, the 3-year-old child forms past tenses by the general rule, as though *make* and *tell* formed past tenses like *baked* and *spelled*.

- *Food terms.* From Japanese come *sashimi*, *wasabi*, *miso*, *sake*, *sukiyaki*, *teriyaki*, *ramen*, and *tofu*; from Chinese, *dim sum*, *bok choy*, and *wonton*; from Spanish, *alfalfa*, *salsa*, *guacamole*, *tostada*, and *anchovy*; from the languages of India, *mulligatawny*, *chutney*, and *samosa*; from French, *beef*, *bouillon*, *chowder*, *cutlet*, *mackerel*, *mutton*, *mustard*, *quiche*, and the more obvious *crepe* (or *crêpe*) and *croque monsieur* (as well as *du jour*, *au jus*, and *a la mode*). English has borrowed food names from many other languages, among them Arabic (*falafel*, *hummus*), Greek (*gyro*, *pita*), Turkish (*dolma*, *yogurt*), German (*bratwurst*, *pretzel*).
- *Top ten words.* The top ten words in printed American English are *the*, *be*, *of*, *and*, *a*, *in*, *he*, *to* (the infinitive marker, as in *to study*), *have*, and *to* (the preposition, as in *to Cancun*) and the next three (for the record) are *it*, *for*, and *I* (according to the Brown Corpus). No nouns, adjectives, or adverbs appear on the list and the only verbs are *be* and *have*. With few exceptions, that's not the kind of information you (or Walter Wordley) could have guessed. It's an empirical fact, dependent in part on the source of data you consult.
- *Rock looker underer.* Nerdy Nan may claim that's nonsense, but it's perfectly understandable. People talk about looking under rocks in hope of finding the unexpected. So you can bet that *Time's* "successful rock looker underer" has uncovered leads for curing cancer in unexpected places. "Hey, Nan, versatility keeps English alive and thriving. What's not to like?"
- *Sneak and snuck.* Turns out both Greg and his prof are right—or both wrong. An Ngram Viewer for *sneaked* and *snuck* (see page 62) shows that book writers have favored *sneaked* for most of the twentieth century but for the past forty years have increasingly written *snuck*. It's unusual for an irregular past-tense verb form like *snuck* to gain ground over a regular form like *sneaked*, but that's what's happening in this case. The two are now nearly neck and neck in books—and books are more conservative than magazines and newspapers, which are more conservative than conversation. So Greg and his friends are in the forefront of a linguistic change, while his prof's preferred form is losing ground. (Note: Graphed lines in a Google Ngram Viewer appear in different colors to distinguish the n-grams from one another. In this book, limited to black-and-white figures, we've adapted the Ngram Viewer by identifying the particular n-gram represented on each line. Here, as indicated, the upper line represents *sneaked* and the lower one represents *snuck*.)

Exercises

Practice Exercises

- A. For the words with more than one morpheme, identify (sometimes approximately) where the morpheme boundaries exist.

Example: governments: govern - ment - s

- | | | |
|--------------|----------------|------------------|
| 1. human | 4. undisclosed | 7. dispassionate |
| 2. tasteless | 5. impolitely | 8. respectfully |
| 3. watching | 6. troublesome | 9. mudslinging |

- | | | |
|------------------|-----------------|-----------------------|
| 10. indelicate | 17. automakers | 24. illegality |
| 11. token | 18. indirectly | 25. supermarkets |
| 12. breadwinners | 19. higher | 26. globalization |
| 13. turmoil | 20. outpaces | 27. antiterrorism |
| 14. heartstrings | 21. remodeling | 28. interdependence |
| 15. underdog | 22. inexpensive | 29. antiunemployment |
| 16. biggest | 23. child's | 30. bouncebackability |

- B. Identify by number three words that contain a prefix.
- C. Identify by number three with a suffix (three different suffixes).
- D. By number, identify three that contain both a prefix and a suffix.
- E. By number, identify three compound words.
- F. By number, identify three words containing an inflectional suffix.
- G. By number, identify three words containing a derivational suffix.

Based on English

- 2-1. Identify the category of the underscored words in the following sentences. Use the abbreviations N for noun, V for verb, Adj for adjective, Adv for adverb, Prep for preposition, Pro for pronoun.
- People who rarely read in bedrooms can feel abnormal.
 - Nobody really knows what normal reading is.
 - The market for audio books is very large.
- 2-2. For three of the five words that belong to three lexical categories in Table 2.2 provide a sentence illustrating their use in each category. Examples for *average* might be: Is there a difference between an *average* and a median? (noun); A guide can *average* \$75 a day in tips. (verb); He earned only *average* grades. (adjective)
- 2-3. a. Identify *the lexical category* of each word in the following list.
- b. List all the morphemes (each word contains more than one) and indicate whether they are free or bound.
- c. Indicate for each affix whether it is derivational or inflectional.

heard	tinier	unproductive
toys	saw	bookshops
listened	reassessment	children's
fixer-upper	fatherly	improbable
improbability	repayment	unamusing
tiest	realignments	calculating
disarms	unremarkable	forewarned
untidiness	realigned	unpretentiousness

- 2-4. a. The three sentences below contain capitalized DEMONSTRATIVE PRONOUNS and italicized *demonstrative determiners*. Characterize the difference in how they are used. (*Hint*: Consider the lexical categories of any words they precede.)

- 1) THIS is the last time I'm doing THAT.
- 2) *This* time I'm not going to make one of *those* fancy pizzas.
- 3) I've had enough of THESE; give me one of *those* red ones.

- b. List each pronoun in the passage below and identify its kind (personal, demonstrative, interrogative, relative). For personal pronouns, also indicate the person (first, second, third).

What about those books? Whose are they? They look like they come from the library, so they should be returned. If you want, you can put them into a shopping bag and I'll return them for you if I can get Ashley to take me in her car. It's been in the shop for a few days. I hope it's ready now.

- 2-5. Consider two popular compounds. *Convenience food* 'food that is convenient to buy, cook, or eat' comprises a noun and a noun. *Natural food* 'food made with natural ingredients, free of chemical preservatives and pesticides' is an adjective + noun compound. As a whole, each compound functions as a noun. List six compound nouns that combine two nouns and six that are unmistakably a combination of adjective + noun. (Note that not all adjectives preceding nouns are compounds, so it is helpful to pay attention to the stress pattern. In the following sentences, the compounds are italicized; say them aloud to see the pattern. Not every white house is the *White House*! Not every black bird is a *blackbird*.)

- 2-6. From a passage of about 500 words in a weekly newsmagazine like *Time* or *The Economist* make a list of 20 compounds, marking the lexical category of each constituent word of the compound and of the compound as a whole. Thus, given *telephone tag* you would identify *telephone* as noun (or N), *tag* as noun, and the compound *telephone tag* as noun.

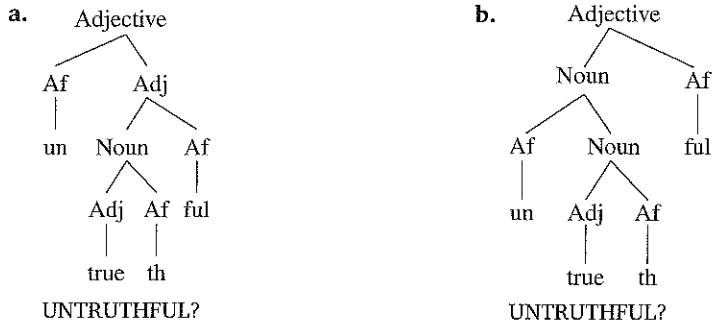
Category of	1st Element	2nd Element	Compound
telephone tag	N	N	N
software	Adj	N	N
bozo filter	N	N	N

- 2-7. *Cyber-* became a popular prefix in the 1990s. It was attached principally to nouns to form new nouns, as in *cyberlove*, *cyberland*, *cyberfiction*, and *cybercowboy*. List eight words that use the prefix *cyber-*, identifying any examples of *cyber-* prefixed to a lexical category other than noun. What about the increasingly popular *eco-*? List eight words that use the prefix *eco-*, identifying any examples of *eco-* prefixed to a lexical category other than noun.
- 2-8. a. Write rules like those given in the section on "Morphemes That Derive Other Words" for the underscored affixes in the words below. Example: disqualify: DIS- + Verb → Verb

- b. Using the derivational rules you've written for Ex. 2-8.a immediately above and those given on pages 41-43, draw trees (as in figures 2.1 to 2.8) that illustrate the derivational structure for these words:

appearance uproot unworthy preheat
 enforcement brightness disgraceful recapitalize

- 2-9. Consider the differing analyses of *untruthful* as represented in the tree structures below. Relying on the rules for deriving other words by affixation, give arguments for preferring one of the analyses over the other.



- 2-10. The following terms are associated with instant messaging, chat rooms, Internet use, and so on. For each one, identify the kind of formation (compound, shortening, acronym, conversion, and so on) and its lexical category.

Example: *chatgroup*—compound, noun

share	YouTube	FAQ	source code
mouse	cyberspace	PC	smartphone
a flame	to flame out	IMHO	to download
apps	to surf	to email	mobile
emojicons	blog	to lurk	newbee (newbie)
a permaban	netiquette	netter	chat room
webcam	to friend	smileys	Facebook
spamming	a sysop	a thread	cyberenthusiast
share	YOLO	to tweet	PayPal

- 2-11. a. From the list that follows, identify four acronyms and four initialisms: DNA, STD, AIDS, SIDS, FIFA, NBA, HIV, HDTV, NHL, UNESCO, UN, UK, BBC, NATO, MOOC.
 b. In terms of initialisms and acronyms, what's interesting about the words *CD-ROM* and *JPEG*?
- 2-12. a. As determined by their frequency in a million-word corpus of texts (the Brown Corpus), the 26 most common words in printed American English are listed on the next page. The category of a few of these words is given. For each of the others, specify its category and then answer

the questions that follow. Choose categories from this list: N (noun), V (verb), Adj (adjective), Prep (preposition), Det (determiners, including articles), Pro (pronoun).

the	_____	they	_____
be	_____	with	_____
of	_____	not	adverb
and	_____	that	conjunction
a	_____	on	_____
in	_____	she	_____
he	_____	as	conjunction
to	infinitive marker	at	_____
have	_____	by	_____
to	_____	this	_____
it	_____	we	_____
for	_____	you	_____
I	_____	from	_____

- 1) List the pronouns that fall among the 26 most frequent words of written English: _____
 - 2) List the prepositions: _____
 - 3) List the determiners: _____
 - 4) List the verbs: _____
 - 5) List the adjectives: _____
 - 6) List the nouns: _____
- b. The words listed in the two columns are found so frequently in print that one of every four words in the Brown Corpus ranks among the first eight words on the list (*the* through the infinitive marker *to*). To put it another way, over 250,000 of the million words in the Brown Corpus are the same eight words used repeatedly. With that in mind, answer the following questions.
- 1) Which two lexical categories are strikingly absent from the list? What explanation can you offer for their infrequency?
 - 2) What explanation can you offer for the frequency of prepositions in the Brown Corpus? (*Hint*: It may help to think about what prepositions do.)
 - 3) What explanation can you offer for the frequency of pronouns as compared to nouns?
 - 4) The verbs *be* and *have* appear on the list. If you knew that the 27th word on the list was a verb, which verb would you guess it to be? Why?
 - 5) Of the 21 words whose lexical category you were asked to identify in part a, how many belong to closed classes of words and how many to open classes?

- 2-13. Examine the photo of “The Grand Canyon of the Yellowstone” sign on page 53. It contains examples of several phenomena discussed in this chapter. Answer the following questions about the words in that sign.
- The words *lovely* and *largely* seem to have the same derivational suffix, but are they the same? Explain.
 - You recognize that *depth*, *width*, and *length* are nouns and that they derived by affixation from *deep*, *wide*, and *long*, but you might expect the derived forms to be pronounced and probably spelled differently, at least if *-TH* were a fully productive derivational morpheme. Use a dictionary to help figure out what accounts for their irregular pronunciations.
- 2-14. The expressions below come from an article discussing new telecommuting policies for county employees in San Luis Obispo, California (*New Times*, April 22–29, 2010, p. 7). Next to each expression write the name of the process by which the italicized word has come to have its use in this discussion, drawing the terms from this list: compounding, affixation, invention, shortening, conversion, derivation, semantic shift, borrowing, blend.
- SLO* County ‘San Luis Obispo’ _____
 - won’t* _____
 - a *telecommuting* policy _____
 - the *groundwork* _____
 - infrastructure* _____
 - greenhouse* gas emissions _____
 - to *curb* the main source of pollution _____
 - staffers *crafted* the schedule _____
 - municipalities* _____
 - pollution* _____
 - allow employees to *patch* into the network _____
 - PJs* ‘pajamas’ _____

Based on Languages Other Than English

- 2-15. Consider the following pairs of singular and plural nouns for human beings in Persian. How does Persian form these noun plurals? (*Note*: æ represents a vowel sound like the one in English *hat*, and x represents a sound like the final consonant of the German *Bach*.)

zæn	‘woman’	zænan	‘women’
mærd	‘man’	mærdan	‘men’
bæradær	‘brother’	bæradæran	‘brothers’
pesær	‘boy’	pesæran	‘boys’
xahær	‘sister’	xahæran	‘sisters’
doxtær	‘daughter’	doxtæran	‘daughters’

2-16. Spanish nouns have grammatical gender; for example, *coche* 'car' and *viento* 'wind' are masculine; *manzana* 'apple' and *sopa* 'soup' are feminine. Examine the expressions below to uncover patterns of agreement among noun, article, and adjective.

- a. el coche rojo 'the red car'
- b. el viento frío 'the cold wind'
- c. el camino corto 'the short road'
- d. la manzana roja 'the red apple'
- e. la sopa fría 'the cold soup'
- f. la falda corta 'the short skirt'

In those expressions, what is the masculine form of the Spanish definite article 'the' and what is the feminine form? Given the masculine form of an adjective, state the rule for forming the feminine singular for these adjectives and then provide the missing adjective in items i and j.

- g. el libro amarillo 'the yellow book'
- h. el hombre alto 'the tall man'
- i. la flor _____ 'the yellow flower'
- j. la mujer _____ 'the tall woman'

Identify how the adjectives in the following examples differ from the preceding ones. Given the masculine form of the adjectives below, state the rule for forming the feminine singular forms, and apply the rule to provide the missing adjectives in items s and t.

- k. el niño feliz 'the happy boy'
- l. el gato grande 'the big cat'
- m. el pan caliente 'the hot bread'
- n. el caballo fuerte 'the strong horse'
- o. el examen fácil 'the easy exam'
- p. la niña feliz 'the happy girl'
- q. la casa grande 'the big house'
- r. la estufa caliente 'the hot stove'
- s. la persona _____ 'the strong person'
- t. la tarea _____ 'the easy homework'

2-17. Consider the following Persian word pairs with their English glosses. Note the lexical category of the words in column A, and give the complete rule for forming the words of column B from those in column A. (Note: x represents a sound like the final consonant of the German *Bach*.)

A		B	
dana	'wise'	danai	'wisdom'
xub	'good'	xubi	'goodness'
darosht	'thick'	daroshti	'thickness'
bozorg	'big'	bozorgi	'size'
shirin	'sweet'	shirini	'sweetness'

- 2-18. a. Analyze the Turkish nouns below and provide a list of their constituent morphemes, along with a gloss for each. (Note: *i* represents a vowel similar to *u*.)

kitap	'book'	elmalar	'apples'	saplar	'stalks'
at	'horse'	masa	'table'	adamlar	'men'
oda	'room'	odalar	'rooms'	masalar	'tables'
sap	'stalk'	atlar	'horses'	sonlar	'ends'
elma	'apple'	kiz	'girl'	meyvar	'fruit' (singular)

- b. On the basis of your analysis, provide the Turkish words for the following English ones: *books*, *man*, *girls*, *end*, *fruit* (plural).
- c. Given Turkish *odalarda* 'in the rooms' and *masalarda* 'on the tables,' provide the Turkish words that mean 'in the books' and 'on the horse.'
- 2-19. In the Niutao dialect of the Polynesian language Tuvaluan, some verbs and adjectives have different forms with singular and plural subjects, as in these examples:

Singular	Plural	
mafuli	mafufuli	'turned around'
fepaki	fepapaki	'collide'
apulu	apupulu	'capsize'
nofo	nonofo	'stay'
maasei	maasesei	'bad'
takato	takakato	'lie down'
valea	valelea	'stupid'
kai	kakai	'eat'

- a. Describe the rule of morphology that derives the plural forms of these verbs and adjectives from the singular forms.
- b. In the Funaafuti dialect of the same language the process is slightly different, as the following plural forms of the same verbs and adjectives show. (Double consonants indicate that the sound is held for a longer period of time.) How are plurals formed from singular forms in this dialect? How does that process differ from the process of plural formation in the Niutao dialect described in the first part of this exercise?

valea	nnofu
maffuli	maassei
feppaki	takkato
appulu	kkai

- 2-20. On the basis of the examples given below, determine whether the following languages have an isolating, inflectional, or agglutinating morphology, and justify your answer.

SAMOAN

?ua	maalamalama	a?u	i	le	mataa?upu
PRESENT	understand	I	OBJECT	the	topic

'I understand the topic.'

FINNISH

tyttö	silitti	paidat
girl-SUBJECT-SING.	iron-PAST-SING.	shirt-OBJECT-PLURAL

'The girl ironed the shirts.'

JAPANESE

akiko-ga	haruko-ni	mainiti	tegami-o	kaku
Akiko-SUBJECT	Haruko-to	everyday	letter-OBJECT	write

'Akiko writes a letter to Haruko every day.'

MOHAWK

t-en-s-hon-te-rist-a-wenrat-e?

DUAL-FUTURE-REPETITIVE-PLURAL-REFLEXIVE-metal-cross-PUNCTUAL

'They will cross over the railroad track.'

THAI

k ^h ruu	hây	sâmùt	nákrian	săam	lêm
teacher	give	notebook	student	three	ARTICLE

'The teacher gave the students three notebooks.'

- 2-21. Examine the following sentences of Tok Pisin (New Guinea Pidgin English) to identify the morphemes needed to translate the seven English sentences given at the end of this exercise.

- a. manmeri ol wokabaut long rot
 people they stroll on road
 'People are strolling on the road.'
- b. mi harim toktok bilong yupela
 I listen speech of you-PLURAL
 'I listen to your (PLURAL) speech.'
- c. mi harim toktok bilong yu
 I listen speech of you-SING.
 'I listen to your (SING.) speech.'
- d. em no brata bilong em ol harim toktok bilong mi
 he and brother of he they listen speech of me
 'He and his brother listen to my speech.'

- e. mi laikim dispela manmeri long rot
 I like these people on road
 'I like these people (who are) on the road.'
- f. dispela man no prend bilong mi ol laikim dispela toktok
 this man and friend of me they like this speech
 'This man and my friend like this speech.'

Now, relying on the meaning of the morphemes you can identify in those Tok Pisin sentences, translate the following sentences into Tok Pisin:

- 1) These people like my speech.
- 2) I am strolling on the road.
- 3) I like my friend's speech.
- 4) I like my brother and these people.
- 5) These people on the road and my friend like his speech.
- 6) You (SING.) and my brother like the speech of these people.
- 7) These people listen to my friend's and my brother's speech.

Especially for Educators and Future Teachers

- 2-22. Assume you are teaching young ESL students to change verbs into their "opposites"—for example, *appear* into *disappear*. How would you get them to provide as many different English-language prefixes to turn verbs into other verbs with an opposite meaning?
- 2-23. a. Suppose you are teaching a middle-school English class how to figure out the lexical category (part of speech) of the words *newer*, *books*, *about*, and *surprise*. Would it be better to present them in isolation or in sentences? Why?
- b. Now reconsider *books* and *surprise*, and try putting them into two sentences each, used as a verb in one and as a noun in the other. Any further observations about which way of presenting them is better? What else could you do to provide your students useful tools for deciding the part of speech of these words? How did you determine their part of speech for yourself?
- 2-24. Imagine you and your students are looking at a map of the United States, examining place names and river names. Concentrating on one side or the other of the Mississippi River, can you anticipate six names your students would rightly guess are borrowed from Native American languages? Examining the whole map, which states or state capitals could they identify as being named after some person? After another country? As using words borrowed from Spanish? French? Dutch?
- 2-25. How could you encourage your students to identify the names of food items that English has borrowed from the languages of the students' respective ethnic heritages?

- 2-26. How would you determine which words to include in the vocabulary of a textbook for international students learning basic English?
- 2-27. What would you tell students who thought English had infixes because in the film *My Fair Lady* they heard Eliza Doolittle use the word “abso-bloomin’-lutely” in these lyrics: “Oh wouldn’t it be loverly? Oh, so loverly sittin’ abso-bloomin’-lutely still! I would never budge ‘til spring crept over my window sill”? Does English have other examples similar to this one?

Resources



Internet

The Internet addresses listed in this section will be helpful in understanding this chapter and exploring related aspects of language. Because Internet addresses often change, the ones given here may go out of date; using your detective skills will lead you to others.

- **LISU Website:** <http://www.CengageBrain.com> For users of this textbook. Provides updated Internet links as well as supplemental material for students and instructors. Here you will find chapter-specific interactive learning tools in your English CourseMate, accessed through the URL above.
- **UCREL Links:** <http://ucrel.lancs.ac.uk/links.html> The University Centre for Computer Corpus Research on Language at Lancaster University in England supports a website that provides links to a wide range of other resources.
- **Merriam-Webster OnLine:** <http://m-w.com> Entry to the world of dictionaries produced by the Merriam-Webster Company. Well worth bookmarking for its online definitions.
- **American Dialect Society:** <http://www.americandialect.org/> Interesting especially for its lists of nominated and winning words of the year over the past decade and a half and for its links to other fascinating websites.
- **Corpus Linguistics:** http://athel.com/corpus_introduction.html Maintained by Michael Barlow, this website is a goldmine of information and guidance about corpora and corpus linguistics.
- **Bookmarks for Corpus-Based Linguistics:** <http://tiny.cc/corpora> This excellent source for beginners and others is maintained by David Lee and provides a wealth of information about corpora and corpus-based linguistics, chiefly for English but including other languages.
- **Corpus of Contemporary American English:** <http://corpus.byu.edu/coca/> A 450-million-word corpus of American English, comprising five genres in equal measure (spoken, fiction, popular magazines, newspapers, and academic texts) and representing several subgenres (movie scripts, sports magazines, scientific journals, and more). COCA contains 20 million words from each year since 1990. At the website, a five-minute hands-on introduction to

using COCA will greatly repay your effort. You'll learn to search for individual words and other expressions and to exploit wild cards and other tools in exploring authentic spoken and written texts lexically, semantically, and grammatically.

- **phpSyntaxTree:** <http://ironcreek.net/phpsyntaxtree/> At this site (at which we have produced the illustrative trees in this chapter) you can produce a labeled tree diagram from a bracketed representation of a word. Some experimenting and practice should enable you to produce trees akin to those in this chapter. Take care to ensure that you have the same number of open and closed brackets. As an example, the bracketing [Adj [Af un] [Adj [Noun [Adj true] [Af th]] [Af ful]]] produced the tree on the left in Exercise 2-9. Note that there must be the same number of open and closed brackets—here, seven.

Suggestions for Further Reading

- **Jean Aitchison.** 2003. *Words in the Mind: An Introduction to the Mental Lexicon*, 3rd ed. (Malden, MA: Blackwell). An entertaining and accessible treatment.
- **Andrew Carstairs-McCarthy.** 2002. *An Introduction to English Morphology* (Edinburgh: Edinburgh University Press). A brief, accessible treatment of English words and their structure, including a chapter on the historical sources of English word formation.
- **Rochelle Lieber.** 2010. *Introducing Morphology* (Cambridge, UK: Cambridge University Press). Appearing in the new Cambridge Introductions to Language and Linguistics series, this hands-on approach will appeal to students.
- **Ronald W. Langacker.** 1972. *Fundamentals of Linguistic Analysis* (New York: Harcourt). An excellent introduction to linguistic analysis and problem solving. Chapter 2 discusses morphological analysis, with illustrations from many languages, including Native American languages. Model solutions are provided for some problems.
- **Donka Minkova & Robert Stockwell.** 2009. *English Words: Structure and History*, 2nd ed. (Cambridge, UK: Cambridge University Press). Rich in examples and easy to use.
- **"Among the New Words."** A regular column in *American Speech* (the journal of the American Dialect Society) that defines the most recent additions to the English word stock. You'll be surprised at how many words in your everyday life are brand-new to English. Look up the most recent "Among the New Words" in the periodicals room of your library or with online access to *American Speech*.
- **The Merriam-Webster New Book of Word Histories.** 1991. (Springfield, MA: Merriam). This exciting book provides word histories for thousands of English words from *assassin* to *zombie* and all sorts of words in between, such as *jeep* and *OK*.
- **Jean-Baptiste Michel, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K. Gray, William Brockman, The Google Books Team, Joseph P. Pickett, Dale Hoiberg, Dan Clancy, Peter Norvig, Jon Orwant, Steven Pinker, Martin A. Nowak, and Erez Lieberman Aiden.** "Quantitative Analysis of Culture Using Millions of Digitized Books." 2011. *Science* 331:176–82. This is the article that prominently announced Google Ngram Viewer. It is accessible and highly recommended.

Advanced Reading

A good general treatment of morphological processes can be found in Katamba and Stonham (2006). Bauer (2003) is more advanced and contains a brief chapter on morphology in the mind and another on the historical development and disappearance of morphology in languages. Our examples of reduplication in Turkish come from Underhill (1976). Good treatments of morphology can be found in Shopen (2007), especially the chapters by Alexandra Y. Aikhenvald on "Typological Distinctions in Word-Formation," by Balthasar Bickel and Johanna Nichols

on “Inflectional Morphology,” and by Bernard Comrie and Sandra A. Thompson on “Lexical Nominalization.” Comrie (2009), from which several examples in this chapter are taken, provides valuable descriptions of more than 40 major languages, including treatment of morphology. The Vietnamese example comes from Comrie (1989).

References

- Bauer, Laurie. 2003. *Introducing Linguistic Morphology*, 2nd ed. (Washington, DC: Georgetown University Press).
- Comrie, Bernard. 1989. *Language Universals and Linguistic Typology*, 2nd ed. (Chicago: University of Chicago Press).
- Comrie, Bernard, ed. 2009. *The World's Major Languages*, 2nd ed. (London: Routledge).
- Katamba, Francis & John Stonham. 2006. *Morphology*, 2nd ed. (Basingstoke: Palgrave Macmillan).
- Shopen, Timothy, ed. 2007. *Grammatical Categories and the Lexicon*, vol. 3 of *Language Typology and Syntactic Description*, 2nd ed. (Cambridge, UK: Cambridge University Press).
- Underhill, Robert. 1976. *Turkish Grammar*. (Cambridge, MA: MIT Press).