1 Introduction: language, languages, and linguistics

CHAPTER PREVIEW

Language plays a crucial role in our lives as a functional system of human communication. It is central to our cultures and societies, and has played a significant role in western intellectual history of the study of philosophy, mind, ancient history, and culture. Linguistics is the scientific study of language. This chapter provides an orientation both to language and to the field of linguistics. It introduces the languages of the world, their distribution and demographics, the important issue of language endangerment and death, and the worldwide effort to document and conserve the world's languages. It then provides an orientation to the field of linguistics and an overview of the major subfields of the discipline.

LIST OF AIMS

At the end of this chapter, students should be able to:

- **articulate the importance of language to human lives and society;**
- **discuss the ways in which language is a functional system of human communication;**
- **take an objective, descriptive approach to discussion of language-related issues;**
- **begin to identify fine details of linguistic structure;**

- state basic demographic facts about the world's languages, including issues of language vitality and endangerment;
- state in what ways linguistics is scientific and objective;
- provide a brief overview of the major subfields of linguistics.

1.1 Language

1.1.1 Language and you; language and us

Language *is an essential and ubiquitous component of our lives.* To see that this statement is true for yourself, take a moment to think about your day. Cast your mind back to when you first awoke. What were your thoughts and how were they expressed? Trace the day in your mind and try to count how many people you spoke with, even if it was just a quick "hi" or "thank you." Did you listen to a lecture? Watch television? Talk on the phone? Make an appointment? Sing a song? All of these activities centrally involve language. Now think about what you read today. Perhaps a newspaper, pages on the Internet, e-mail, advertisements, labels, signs, homework assignments? Now move on to thought itself. What thoughts and ideas have passed through your mind? Have you made explicit plans, imagined conversations, debated with yourself? If you are like most people, this brief exercise has revealed that language is both within and around you, a constant part of your internal and external existence. Language is the primary medium which you use to interact with people and institutions in our society. Your particular use of language is also a reflection of who you are as an individual; all of us use language as a means to build and portray our identities in the world around us. We also use language to shape and interpret the great and small experiences of our lives.

Think about the broader world in which we live. Language is the principal means by which societies are constructed and cultures are developed. Think of the size of our society's great libraries, and how the majority of the volumes in those vast collections (14.6 million volumes in the Harvard University Library alone) are language in its written form. The intellectual achievements of humankind are essentially embodied in language. This is not only true of the written works that formally encapsulate our knowledge, but it is also true of the huge body of indigenous knowledge held by the speakers of thousands of languages across the globe, from the Brazilian Amazon to the Mongolian steppes. Some may argue that music and art are non-linguistic, but note that they often incorporate language, as with lyrics. Even works that do not contain language are interpreted and understood through verbal thought, discussion, and critical analysis. Similarly, mathematics could be argued to be non-linguistic, but again language is used to teach, understand, and interpret it.

Beyond the modern world, consider that language has been used by humans for at least 30,000 years, by thousands of groups across the globe, wherever humans have ventured. Speakers of each generation endow their language with their own unique mark, their own contribution, changing it in myriad subtle ways. As language passes

from generation to generation, it shifts and adapts to the ever-changing world in which it is embedded.

The preceding paragraphs emphasized that ***language is a pervasive and essential part both of your own life and of who we are as humankind***. The goal of this book is to begin to address the question: *How does language work?* It is a simple question, and one that most people never think to ask. Language is so automatic – almost like breathing – that most people don't realize the complexity that underlies it and the subtle and effortless skill with which they wield it.

The question *How does language work?* may itself be simple but the answer is highly complex. It can be broken down into many smaller questions. To begin with, one must ask: *How do individual languages work?* We really can't understand the nature of language in its broad sense if we don't understand the mechanisms underlying particular languages, preferably of many and diverse kinds. Other key questions include: What are all the pieces of a language? How do the pieces combine and work together to allow for communication to occur? How are languages learned and transmitted? How do languages influence each other? How do languages change over time? These are but a small number of the many questions that define the field of **linguistics**, the scientific study of language. But before discussing the field in more detail, it is important to continue with our exploration of the nature of language.

1.1.2 Language is human and all that that implies

Language is one of the defining traits of humankind. Language is tied up with our thought processes, our ability to reason, to self-reflect, and to develop advanced civilizations. Other animal species have developed communication systems, but they pale in comparison to human language. A simple illustration of this is the fact that no system of animal communication appears to be able to communicate events that occurred in the past or events that are imaginary. Neither are there animal communication systems that have adverbs or other devices that allow for detailed descriptions of actions. Animals have nothing comparable in scale, complexity, subtlety, or adaptability to human language.

The fact that language is human has a number of important implications for the nature of language. ***Language is embedded into our physiology, our cognition, and our thought processes.*** Many of the details of linguistic structure are directly dependent on this. For example, the fact that no language makes sounds by curling the tip of the tongue back to touch the uvula (the small appendage hanging down in the middle of the back of the mouth) is directly explainable by the details of human anatomy. Less trivially, anatomical facts are also responsible for a number of features of sound systems, such as the common trend to pronounce a sequence of *t* and *y* as "ch" (e.g., *gotcha* from *got you*). More importantly, language processes are largely resident in the brain and so language shares characteristics with other cognitive functions; for example, language is both learnable and adaptable.

Humans use language for a wide variety of purposes. We communicate everything from urgent warnings to random thoughts, proposals of marriage to complaints. We use it to cajole, threaten, placate, inform, entertain, and command. In other words, ***language is functional***; ***it is a tool of human communication***. The fact that language is used for a wide variety of tasks has direct implications for how it is structured. Linguistic structures are flexible and adaptable, able to express all that humans convey to each other in the course of a conversation, a day, a lifetime, a civilization.

Language is also human in that ***language is a form of human social behavior***. It can be used to build or break social bonds. It serves as a social cue to the formality or informality of a situation, and to the degree of social intimacy or distance among the people speaking. When children acquire language, they do so by using it as a tool of social interaction within particular social settings. The social component of human language is also reflected in how language is used and structured.

Humans use language to interact, and ***using language is an inherently interactional task***. Not only are we listening to our conversational partner and picking up on the many subtleties of word choice, sentence structure, rate of speech, and intonation, we are also constantly assessing when and how to take a turn, and how to communicate our message so that the person to whom we are speaking (the **addressee**) will correctly interpret what we are saying. To take a simple example, I wouldn't say *He is coming for dinner tonight* if I didn't think that the addressee had in mind the person I refer to as *he*. Otherwise, I could use a proper name like *Mike* or a more elaborate phrase like *the guy from across the hall*. I could also start off with an introduction, such as *You know that guy I was telling you about, that owns the cocker spaniels?* All three of these strategies accomplish a similar end of introducing the idea of the person I wish to discuss into the mind of the addressee. Once I am confident that the addressee can identify the correct individual, I can communicate the primary message *He's coming to dinner tonight*. Thus, we see that the interactional component of language is both deep and subtle. The structures of human language reflect our interactional needs.

Humans are creative and ***language is structured to take advantage of human creativity***. All languages are constructed in a way that allows for the creation of novel utterances; any language can produce an infinite number of sentences. Therefore we cannot describe a language by simply making a list of all the possible sentences it contains. Instead, our task is to describe the design principles underlying language that make that infinite number of sentences possible. Obvious instances of human creativity with language include word games, puns, and puzzles. Humans also use language creatively when they innovate new expressions, or use one or more words in a new way. For example, the English word *way* has been used for some time to intensify the meaning of certain types of quantifiers (*way too much, way more than necessary*) or prepositions (*way up, way over*). Younger speakers of some English dialects can now use this intensifier with adjectives; e.g., *way cool*. The use of *way* with adjectives can have specific affective (emotional) implications, e.g., *way unfair*. We don't know who

first used *way* to intensify an adjective, but in doing so that person was performing a creative act, using the word in a new grammatical environment. People do this every day. Most of the time grammatical innovations are not repeated, but sometimes particular innovations catch on. Other speakers hear the innovation and use it themselves, spreading it wave-like across a significant portion of the **speech community**, a group of people who share a common language or dialect and cultural practices. If an innovation continues to spread, it could become a regular feature of the language and constitute a **language change**. Many instances of language change are direct reflections of human creativity.

To summarize, just as language is deeply a part of humankind, the human element is deeply a part of language. The structures of language take the form they do because language is instantiated by the human body, as a tool of human communication, and is embedded in human interaction within societies and cultures. Language is at the core of what it is to be human, and humanity is at the core of language.

1.1.3   Language is dynamic and adaptable

Language is in a constant process of change. The language you speak with your friends today is somewhat different from the way your grandparents spoke to their friends when they were your age. Chances are good that your own grandchildren will probably think that your speech sounds a little old-fashioned. While the difference between grandparents and grandchildren may not be dramatic, over a longer time span, for example, that between oneself and one's grandchildren's grandchildren's grandchildren, the cumulative effect of those generations becomes more noticeable. We can see this in the history of English. Consider the following passage, written by William Shakespeare just over three hundred years ago, and taken from the play *King Henry V*:

> Now, fie upon my false French! By mine honour in true English, I love thee, Kate: by which honour I dare not swear thou lovest me; yet my blood begins to flatter me that thou dost, notwithstanding the poor and untempering effect of my visage.

While educated English speakers will be able to understand this passage, children and adults with less formal education will find it difficult. It is easy to identify the linguistic features that mark this as archaic: the use of the old second-person familiar pronouns, *thee* and *thou*; the inflected verb forms *lovest* and *dost*; and the use of now antiquated words and expressions, such as *fie upon* and *visage*. When we look further back, for example at *The Canterbury Tales*, written by Geoffrey Chaucer more than six hundred years ago, the language becomes even harder to decipher. Consider these lines from "The Wife of Bath's Tale":

> And if thou kanst nat tellen it anon
> Yet shal I yeve thee leve for to gon
> A twelf-month and a day to seche and leere

An answere suffisant in this mateere;
And suretee wol I han, er that thou pace,
Thy body for to yelden in this place.

While some of it seems familiar and suggestive of meaning, much is unclear to the eye of the untrained modern English speaker. The passage is easier to decipher if one learns that *yeve* means 'give,' *seche and leere* means 'search and learn,' *suretee* means 'certainty,' and *yelden* means 'surrender.' Try providing a modern English translation and compare it with that given in Textbox 1.1.

TEXTBOX 1.1 **MODERN ENGLISH TRANSLATIONS OF *THE CANTERBURY TALES***

Here is one translation of the excerpt from "The Wife of Bath's Tale," provided by Librarius at the following URL: www.librarius.com/canttran/wftltrfs.htm:

> And if you cannot tell it me anon, then will I give you license to be gone a twelvemonth and a day, to search and learn sufficient answer in this grave concern. And your knight's word I'll have, before forth you pace, to yield your body to me in this place.

Of course, you would never speak this way to someone in a conversation today. A more colloquial current translation might be "And if you can't tell me soon, then I'll give you permission to be gone for a year and a day, to find the right answer to this important question. I'll have you promise as a knight, before you leave, that you will give me your life in this place."

All aspects of language can undergo change. Sounds can enter a language or fall out of use. Sentence structures can shift in interesting ways. Words can develop into prefixes, suffixes, or other small linguistic units. Word meanings can be broadened, narrowed, or otherwise shifted. The social implications of using particular words and phrases can change over time, as can larger patterns, such as how we structure and present information.

Language adapts to the world around it. Think of all the vocabulary you use in daily life that your grandparents did not use when they were your age. The words *e-mail*, *nanotechnology*, *cell phone*, and *Internet* are just a few of the terms that reflect the technological changes that swept over us in the late twentieth century. In the meantime, words like *hogshead* (a large cask or barrel) and *demijohn* (a narrow-necked bottle enclosed in wicker) are not part of the vocabulary of most people living today (although they might persist in certain subgroups of the population). Changes in vocabulary can reflect social changes as well. The English word *spinster*, meaning an unmarried woman past the age of marrying, has vanished from everyday vocabulary in most of modern society, together with the idea that there is an age of marrying and that marriage and family are the primary goals of a woman's life.

While changes in vocabulary reflecting innovations or social change are probably the most obvious examples of the adaptability of language, languages also undergo adaptations under the influence of **language contact**. When speakers of two distinct languages interact with each other in large numbers over a period of time,

one or both languages generally undergo change. An example of a language affected by language contact is English, which adopted huge numbers of words from French after the Norman invasion. Indeed, in the sentence I just wrote, the words *example*, *adopt*, *huge*, *number*, *French*, *Norman*, and *invasion* all came into English from French!

Language contact can have a much greater effect than simply adding new vocabulary. Sounds, word structures, and sentence structures can also take on qualities of adjacent languages. For example, in the Tibeto-Burman language family (comprising over three hundred related languages distributed over Southeast Asia, Tibet, and the Himalayan region), the majority of languages place the verb at the end of the sentence. A simplified and translated version of a sentence with this word order might be, for example, *John apple ate*. However, there is one group of Tibeto-Burman languages, the Karenic group, which places the verb in the middle of the sentence. Thus, they would say *John ate apple*. Interestingly, speakers of the Karenic languages have been interacting for centuries with the Thai and the Chinese, and both groups speak languages that put the verb in the middle. It is clear that over the centuries, **bilingual** Karenic speakers matched their sentence structures to those of their neighboring languages. Thus, a significant change to Karenic grammar resulted from language contact through the medium of bilingualism. We see that ***languages adapt not only to the changing technological world, but also to their broader social environment***.

Language is structured and systematic

When one begins to look closely at language, one is immediately struck by the fact that ***regular and recurring patterns form the basis of linguistic structure***. To begin to explore this aspect of language, take a moment to work through the following small exercise on English grammar:

Regular patterning of the English past-tense suffix

In English most verbs have a predictable past-tense form. It is written as *-ed* but has different pronunciations. You can discover this in your own speech very easily. Pronounce the following lists of words and listen closely to the sound at the end of each word:

List A: *baked, blessed, heaped, puffed, crashed*
List B: *rubbed, waved, lagged, billed, hummed*
List C: *waited, faded, booted, coded, righted*

If you are a native English speaker and have a sensitive ear, you will have noticed that the words in List A end in <t>, the words in List B end in <d>, and the words in List C end in <ed>. We can now refer to these as the T-List, the D-List, and the ED-List.

Now try pronouncing the following three nonsense words, again listening carefully to how the suffix is pronounced in each word:

Word 1: *smipped*
Word 2: *croomed*
Word 3: *pluted*

Notice that you don't have to think for an instant which sound to put at the end, but that you automatically end Word 1 with <t>, Word 2 with <d>, and Word 3 with <ed>, even though these are nonsense words which you are unlikely to have ever heard or pronounced before.

Take a minute to examine the consonants that directly precede the suffix (i.e., the "pre-suffixal" consonants) in the T-List words. Now compare the pre-suffixal consonants in the D- and ED-List words. Notice that the lists are distinct; you don't find any of the T-List pre-suffixal consonants in D-List words, etc. Now determine which lists Words 1–3 fall into, based on their pre-suffixal consonants.

You will see that Word 1 has a T-List consonant (p) and the suffix is pronounced as <t>, Word 2 has a D-List consonant (m) and the suffix is pronounced as <d>, and Word 3 has an ED-List consonant (t) and the suffix is pronounced as <ed>. You have discovered a systematic fact of English: the pronunciation of the past-tense suffix depends upon the pre-suffixal consonant. ***Even though Words 1–3 are nonsense words, they still follow the systematic patterns of pronunciation*** that form a significant part of the English language. We can state this pattern as follows:

(1) In English, the past tense *-ed* will be pronounced: as <t> following the consonants <k, s, p, f, sh>, as <d> following <b, v, d, l, m>, and as <ed> following <t> or <d>.

SIDEBAR 1.1

The statement in (1) is only part of the pattern, as not all possible consonants are exemplified. The lists for two of the groups are actually much larger than shown here. Can you determine which two groups these are?

This is a statement of a pattern or systematic fact of English (sometimes referred to as a rule). One can predict how the past tense *-ed* will be pronounced on any English verb as long as one knows the pre-suffixal consonant (see Sidebar 1.1).

Once we have observed a regular pattern in language, ***we are led to the question of why this pattern should occur***. This question is critical, because it takes us from recognition and description of a pattern to a search for an explanation of the observed facts. In this case, the explanation is physiological, based on how we produce sounds in our vocal tracts. Since this is a topic covered in the next chapter, we will not go into detail here. The important point is that ***patterns in language can be explained by the role of language as a functional system of human communication***. In this case, the explanation comes from the embedding of language in our human physiologies; in other cases, other aspects of the functional nature of language explain linguistic patterns.

Regular patterns such as this occur in every language many times and at many levels. Some patterns are concerned exclusively with sounds, other patterns are found at other levels, such as word structure or sentence structure. One of the fascinating

aspects of language is the interaction of these patterns, which at times can be quite complex. All the patterns in a language that explicitly involve sounds make up the "sound system" or **phonology** of a language; the patterns which involve word structure make up the **morphology**, while the patterns which involve sentence structure make up the **syntax**. Each of these subsystems of language is independent, but each is also interwoven with the others. In the example above, both the phonology (in this case, which sound is pronounced where) and the morphology (the past tense suffix -*ed*) are involved. The morphology and syntax of a language are together referred to as the language's **grammar**. For further discussion of the sub-areas examined in linguistic analysis, see Textbox 1.2.

TEXTBOX 1.2 **LINGUISTIC ANALYSIS**

Many examples of systematicity in language will be presented throughout the following chapters. One of the goals of this book is to teach you how to recognize and analyze systematic patterns in a wide variety of languages, that is, how to perform **linguistic analysis**. This requires learning the common – and sometimes the rare – linguistic categories that are found in the languages of the world, the terminology that accompanies those categories, and the theories underlying them. Linguistic analysis requires logical thought, a clear understanding of linguistic concepts, and concise description and argumentation.

Once linguistic structures are accurately described, the analysis is completed by explanation. ***The critical question is: what motivates the linguistic structures to be formed in precisely that way?*** This question goes to the very core of linguistic theory. The answer will depend crucially upon the particular structure being explained. There are a number of distinct domains that may contribute to it, including meaning (**semantics**), how the structure is used in context (**function**), factors related to history (**language change**), the physical properties of sound (**phonetics**), and the structure of the human brain and how we learn and process knowledge (**neurology**, **cognition**).

On the other hand, ***no language is perfectly systematic***. Although there are sometimes patterns within patterns within patterns, there are often pieces that don't fit into any regular pattern, but which have idiosyncratic, or irregular, behavior. This is in large part due to language change. The irregularities are leftovers from older patterns that have been obliterated, as new structures emerge and spread through the language.

As an example, consider the English verb *shine*. This verb is a bit irregular as it has two forms of the past tense, *shined* and *shone*. The form *shined* is constructed by adding the regular past-tense suffix to the verb stem and following the rule of past-tense formation we just discovered (*shine* ends in a D-List consonant). The form *shone* is a reflection of an old pattern where past tense was indicated by changing the vowel in a verb's root. This pattern was inherited from an earlier stage in the language. It has largely died out, but traces of it remain in a handful of verbs, especially those that are used frequently and are therefore resistant to change (for example, *take/took, drive/drove*). In the development of English, the marking of past tense by -*ed* gradually spread through the vocabulary, supplanting the older forms. This process has not been completed with the verb *shine*, and both past-tense forms coexist in the modern tongue. Thus, this

irregularity of the language has a historical explanation. ***Irregularities in language usually result from language change***.

1.2 Languages

1.2.1 Languages of the world today

Languages are spoken across the globe. People are spread over the earth from the tip of Tierra del Fuego to the Arctic North, and wherever there are people, there are languages. Think for a minute about each of the continents and their communities. How many languages do you think are spoken in the world?

The question is more difficult than it first appears. The truthful answer is that ***we don't have an exact count, although we are able to make an educated guess***. There are two primary reasons why counting up languages is tricky. One is that ***linguists haven't identified all the languages of the world yet***. There are still speech (and sign-language) communities that follow their traditional ways of life and who have had little interaction with larger population groups or researchers. The languages of these groups are still undescribed. However, there is also a more funda-mental problem in counting up languages, which is that ***it is difficult to decide which speech varieties should be counted as languages and which should be counted as dialects of a single language***.

Let's consider possible criteria for distinguishing languages from dialects. One obvi-ous place to start is **mutual intelligibility**: can the speakers of the two language varieties understand each other? The criterion of mutual intelligibility, taken to its logical conclusion, suggests that if they can understand each other, the two varieties are to be considered dialects of a single language; if they cannot understand each other, the varieties are to be considered distinct languages. One problem with this criterion is that there are often multiple varieties of a language, and while speakers of adjacent varieties can understand each other, speakers of geographically separated varieties have a much harder time. This situation is schematized in Figure 1.1:

A   B   C   D   E

Figure 1.1 Schematization of language varieties

In Figure 1.1, each letter represents speakers of different varieties and the arrow represents geographic distance. While speakers of A might easily understand speakers of B and C, it might take effort to understand speakers of D, and it might be quite difficult to converse with speakers of E. Similarly, speakers of E might have no problem speaking with those of D and C, but might have more difficulty with speakers of A. So, are A and E different languages? If so, where does one draw the dividing line? This situation is known as a **dialect continuum**, and it represents a common situation throughout the world.

Of course, Figure 1.1 is highly idealized. Communities aren't usually ranged along a straight road with distinct boundaries, and there is often movement and intermarriage between the various groups. However, the problem remains of whether mutually unintelligible A and E should be counted as one or two languages. We can see that the question itself is overly simplistic and obscures the more complex reality of the dispersion of language varieties and their speakers.

Another problem with the criterion of mutual intelligibility is the word "mutual." This implies that speakers of both speech communities are equally at ease or equally perplexed when hearing the speech of the other. However, there are many cases of unidirectional intelligibility, that is, speakers of Group A can understand the speech of Group B, but not the other way around. This situation especially occurs when the Group A variety is spoken by a minority group and the Group B variety is a **standard language**, taught in schools and used in print and broadcast media. In this situation, the Group A speakers have repeated exposure to the B variety and so can understand it. The Group B speakers, on the other hand, may never have heard the speech of Group A, so find it surprising and difficult. It is not always clear whether these varieties are different dialects or different languages.

Another reason why it is difficult to count up languages is that ***there is a complex relationship between language and ethnic identity***. Consider the case of the Newars, an ethnic group which traditionally ruled the Kathmandu Valley in Nepal. While the largest concentration of Newars is in the Kathmandu Valley itself, there are other Newar communities scattered throughout the country. One variety of Newar is spoken in a village called Dolakha, quite a distance to the east. The Dolakha and Kathmandu speech varieties are truly mutually unintelligible. People from these two Newar communities cannot speak to each other in Newar, but must use the national language Nepali to converse. If the question of language versus dialect were to be based solely on mutual intelligibility, then these two varieties would count as separate languages. However, the Dolakha Newars are ethnically Newars in every sense of the word. They have the same customs, social structures, festivals, and traditions, and they intermarry with Newars from other parts of Nepal. Crucially, their language, even though mutually unintelligible with the other varieties, still serves to distinguish the group ethnically from non-Newars, so is a marker of Newar ethnic identity. The language is thus Newar in a very real and relevant sense to the speakers of the language itself. The function of the language as a marker of ethnic identity would suggest that the Dolakha variety is a Newar dialect, not an independent language. The criteria of mutual intelligibility and ethnic identity thus lead us to different conclusions on the question of language versus dialect.

The opposite situation can be found with Swedish and Norwegian, two of the Scandinavian "languages." These two speech varieties are easily mutually intelligible. However, a national boundary and ethnic identity divide the two groups, hence they are considered to speak distinct languages rather than dialects of a single language. Such circumstances motivated the famous quip by the Yiddish linguist

Figure 1.2 Newars at the temple complex in Patan, Nepal

Max Weinreich: "A language is a dialect with an army and a navy." ***Sociopolitical and ethnic considerations clearly have significant weight in the language/ dialect debate***.

While acknowledging that there are inherent difficulties in counting up the languages of the world, we still want to know roughly how many there are. The most current compilation of statistics on the world's languages is found in *Ethnologue: Languages of the World* (available online at www.ethnologue.com). My source for the statistics in the following discussion is the Internet version of the seventeenth edition (Lewis, Simons, and Fennig 2013), which puts ***the total number of known languages at 7,105***. How close was that to your own estimate?

The distribution of languages across continents is given in Table 1.1 (note that "the Americas" include North, South, and Central America, and "the Pacific" includes Australia, New Zealand, and the Pacific Islands). Table 1.1 shows the number and percentage of the world's languages spoken or signed on each continent.

Note that the languages of Europe account for less than 4 percent of the total number of languages of the world, while Asia and Africa have more than 30 percent each.

Table 1.2 presents statistics on the world's languages in relation to the size of the speech communities of native speakers.

TABLE 1.1 Distribution of languages across continents

Area	Number	Percentage
Africa	2146	30.2
The Americas	1060	14.9
Europe	284	4.0
The Pacific	1311	18.5
Asia	2304	32.4
Total	7105	100.0

TABLE 1.2 Number of languages by size of speech community

Number of speakers	Number of languages	Percentage
100 million to 1 billion or more	8	0.1
10 million to 100 million	77	1.1
1 million to 10 million	308	4.3
100,000 to 1 million	928	13.1
10,000 to 100,000	1798	25.3
1,000 to 10,000	1984	27.9
100 to 1,000	1054	14.8
10 to 99	340	4.8
1 to 9	134	1.9
0	188	2.6
Unknown	286	4.0

Table 1.2 shows that there are very few languages with very large numbers of speakers; only 5.5% of the world's languages have more than a million speakers. On the other hand, 56% of the world's languages have fewer than 10,000 speakers. When we combine these numbers with population statistics, the results are quite striking. ***Roughly 94% of the world's population speaks only 6% of its languages. The remaining 94% of the languages are spread over only 6% of the population***. Thus, we have a handful of languages with enormous speech communities and a very large number of languages with quite small speech communities.

1.2.2   Languages of the world tomorrow

While there are around 7,000 languages spoken or signed on the globe today, not all languages are equally robust. Over time patterns of language use in multilingual communities can shift so that a socially dominant language comes to be used more frequently and less-dominant languages are used in fewer social contexts and among fewer people. Such languages are described as **endangered**, at risk of ceasing to be spoken in the absence of conscious efforts to keep them vital. ***According to the Ethnologue, about 45 percent of the world's languages are endangered***.

The endpoint of the language endangerment process is **language death**, which occurs when a language ceases to have speakers and no longer serves as a symbolic marker of identity for the community. Normally the process of endangerment occurs gradually, over three or more generations. It involves a cessation in **language transmission**, the passing on of a language from one generation to the next. When children don't learn the language, the only remaining speakers are adults. That population naturally ages and declines until only a handful of speakers remains. In the absence of community efforts to reverse the trend, the language can cease to be spoken. If it ceases to be a cultural resource for the community, it is classified as dead (or extinct).

TEXTBOX 1.3

Does language death matter? Linguists and members of many speech communities answer with a resounding "yes." Each language is a testament to the ways in which a unique group of people has understood and interacted with their environment and has come to terms with the human condition. Each is a unique inheritance from countless generations of forebears, the encapsulation of their wisdom and knowledge. Each language reflects and instantiates the culture of the speakers. Each contains knowledge, traditions, and history. Each represents what a language can be and so enriches our understanding of this central aspect of our humanity.

"Surely, just as the extinction of any animal species diminishes our world, so does the extinction of any language. Surely we linguists know, and the general public can sense, that any language is a supreme achievement of a uniquely human collective genius, as divine and endless a mystery as a living organism. Should we mourn the loss of Eyak or Ubyky any less than the loss of the panda or California condor?"

– Professor Michael Krauss
Alaska Native Languages Center

There are a number of reasons why languages become extinct. Sometimes the process of language death has been brought about by explicit government policies designed to keep children from learning their native language. However, language extinction is not limited to communities targeted by such policies. ***Language endangerment and death appears to be primarily fueled by the broader process of globalization, including a shift from agrarian to urban lifestyles, and the increasing dominance of a small number of languages for the purposes of commerce, education, and the media***. These include both the truly widely spoken languages,

like Mandarin Chinese, English, Spanish, Hindi-Urdu, and Arabic, and smaller national languages, like Nepali, Greek, Georgian, and Thai. Often acquisition of such languages is necessary for anyone wanting to pursue an advanced education or a career in modern society. Thus, parents are under pressure to have their children educated in these languages and therefore choose to transmit these languages as opposed to those of the heritage communities.

Another element that can contribute to the loss of a language is the ***loss of the coherence and vitality of the speech community***. If the members of a small speech community become absorbed into a larger group through intermarriage, the community can become dispersed. Where there is no viable speech community, there is little reason to pass the language on to the children; neither will the children hear the language spoken with sufficient frequency to acquire it.

The recognition of the scope of the problem of language endangerment has led to significant work by members of endangered-language speech communities and linguists to record, preserve, and revitalize languages. **Language documentation**, the creation of an extensive record of a language and its community, is an important part of this process. **Language conservation** is also being undertaken in many communities, which are developing materials to be used in the education of children and to promote language use in the speech community. **Language revitalization** is

Figure 1.3 Members of the Gusii community in Kenya record traditional songs and dances as a component of their documentation of the Ekegusii language and Gusii culture (photo by Kennedy Bosire)

undertaken by speech communities whose language has been entirely lost or significantly reduced. Such projects can do much more than simply teach the language; they can play significant roles in strengthening communities and in promoting the preservation of traditional knowledge, practices, cultural values, and institutions.

1.3 Linguistics

1.3.1 The scientific study of language

Now that we have learned a bit about language and about the world's languages, we turn at last to the topic of linguistics. ***Linguistics is the scientific study of language***. By "scientific," we mean that the study is both **empirical** (based on observable data) and **objective**. Empirical data is critical for any scientific discipline, as it ensures that others can verify or replicate the findings. The term **linguist** refers to a person who examines the structures and principles underlying languages. Note that this is different from a **polyglot**, a person who speaks many languages. For more on this distinction, see Textbox 1.4.

TEXTBOX 1.4 **LINGUIST VERSUS POLYGLOT**

The longer you study linguistics, the more likely it is that someone will ask you the question: "How many languages to do you speak?" This question illustrates the commonly held misconception that linguists are polyglots. It is important to distinguish between the two. A **linguist** is a person who examines the structures of languages and the principles underlying those structures. A **polyglot** is a person who speaks many languages. Many linguists are, indeed, polyglots, but you don't have to be a polyglot to study linguistics. A nice analogy can be made to pilots and airplane mechanics. A pilot knows how to fly an airplane, based both on training and on an instinctive sense of flight and how a plane responds to a particular manipulation of the controls. An airplane mechanic looks inside a plane and knows how each part contributes to the workings of the whole. One doesn't need to be an airplane mechanic to be a pilot. Neither does one need to be able to fly a plane in order to be a mechanic. A linguist is like a mechanic, looking inside to see how the parts of the language fit together so that the language can function in human communication. The speaker is the pilot, able to use the language efficiently and effectively, but without necessarily knowing how it works.

   Probably the best airplane mechanics are also pilots, and in the same way, the most insightful analysis of the language will come from someone who speaks it, but a linguist can make a tremendous amount of headway on the analysis of a language without speaking it.

In linguistics, ***empirical data are recordings of spoken or written language, collected into a*** **corpus**. The nature of the recordings and how they are collected will depend on the goals of the study. For example, if one wishes to study the physical properties of sounds, the best recordings might be those produced in the isolation of a sound booth. If one wishes to study sentence structures and how they are used, the best recordings are likely to be natural conversations or narratives, supplemented by the comments of native speakers that reflect their intuitions about the structures and their

meanings in that particular context. If one is studying the language and society, one might choose to make video recordings of authentic interactions. In any case, recorded data, preferably of speech or writing produced in a natural setting, and not constructed by or for a linguist, are the most highly empirical and can be verified by subsequent researchers. This is not to say that this is the only type of useful data in linguistics. Speakers' intuitions about their language, particularly regarding subtle distinctions in meaning, add a depth to our understanding that we could not possibly obtain otherwise.

When we say that a science is objective, we mean that ***our analysis is not biased by any preconceived notions, or judgments of "good" and "bad."*** Human beings are prone to prejudice, and this can be directed at speakers of languages just as it is directed at ethnicities, religions, sexualities, styles of dress, or any other characteristic by which people are subgrouped. It is not uncommon to find languages described as "primitive," "corrupt," "illogical," "ugly," or just plain "bad." By contrast, other languages can be described as "perfect," "logical," or "beautiful." To take an example from the United States, some speakers of American English believe that the dialect of English spoken in certain African American communities (referred to as African American English, or AAE) is "corrupt" or "ungrammatical." People with this view cite AAE sentences like *She sick* and *She be sick*, and claim that they are "incorrect" since they differ from the Standard American English sentence *She is sick*. In actuality, AAE is making a grammatical distinction in these two sentences that is not marked in the grammar of Standard American English. The sentence *She sick* refers to a present situation; it simply states that the person is sick now. This sentence could be used, for example, to explain why someone is unexpectedly absent. The sentence *She be sick* means that she is often sick or has a long-term illness. The implication is that the illness is ongoing, and lasts for an extended period of time. This meaning distinction between a present state and an ongoing state is systematically made by the grammar of AAE (as well as by many other languages in the world). Of course, speakers of Standard American English can still signal this meaning if they want to, for example, by using an adverb such as *always*, but its use is not grammatically required. This doesn't mean that AAE is any "better" than Standard American English; the two dialects are just different. ***Every language or dialect is unique in the types of distinctions it makes. Every language is equally able to convey all of the complex meanings that humans communicate to each other in the course of a lifetime***. Languages differ in which distinctions they grammatically require their speakers to make, and in which meanings can be expressed by other, non-grammatical, means.

An important distinction can be made between **prescriptive** and **descriptive** approaches to language. ***A prescriptive approach to language is one that teaches people the "proper way" to speak or write***. Many children are exposed to prescriptive grammar in school, where they are taught, for example, not to split infinitives (e.g., *to boldly go*) or to end a noun phrase with a preposition (e.g., *the man I saw you with*). Prescriptive grammarians choose a set of forms that they enjoin others to adhere to.

These forms represent a (slightly) older stage of the language when the rules were regular, so the establishment of prescriptive rules reflects a resistance to the natural forces of change. In actuality the set of forms chosen for prescription are ultimately arbitrary; ***there is no logical reason why one should not split an infinitive or end a sentence with a preposition***. Prescriptive rules may still have social ramifications, however, and there are environments (such as academic writing) where ignoring these conventions can have negative social consequences (such as lower grades).

   A descriptive approach to language is one that describes how people actually use language. Descriptivists are not interested in telling people what is right or wrong, but in observing, describing, and explaining actual linguistic behavior. In line with the objective nature of linguistic science, linguistics is a descriptive enterprise.

1.3.2   Fields of linguistics

The field of linguistics is as broad and multifaceted as language itself. The following paragraphs provide a very brief orientation to the primary subfields of the discipline. As in the rest of this book, this presentation will take the traditional hierarchical approach to language, beginning with the smallest units and working up to larger and larger levels.

   We will begin with the study of speech sounds. The physical properties of sounds – how they are articulated and perceived, and the acoustic signatures of the sounds themselves – are the subject of study in the field of **phonetics**. We will then examine the systematic use of speech sounds in language, or **phonology**. (See Sidebar 1.2 for a brief explanation of the distinction between phonetics and phonology.)

   From the study of sounds we move to the study of words themselves. The ways in which words are structured and created are the purview of the field of **morphology**. Morphologists look at all the pieces of words (roots, prefixes, suffixes, etc.), their sounds and meanings, and the principles of their combination. Sidebar 1.3 presents one example of a topic in the area of morphology. The study of how words combine into phrases, clauses, and sentences is the study of **syntax**. Morphology and syntax are tightly integrated and are often referred to as **morphosyntax** or (in some uses) **grammar**.

A critical aspect of language that interacts with all of these levels is **semantics**, meaning in language. The study of semantics includes the study of words (**lexical semantics**) and the study of how meanings combine in clauses and sentences (**propositional semantics**).

   When we look at how speakers use linguistic structures in larger stretches of speech, we are studying **discourse**. This

field takes into account the interactional nature of language, for example, how speakers need to present their ideas in a way that allows hearers to understand them. With the help of computers, linguists can now look at statistically significant patterns over very large sets, or **corpora**, of discourse data; this methodology is referred to as **corpus linguistics**. The role of the broader context in interpreting linguistic form and meaning is examined in the field of **pragmatics** (see Sidebar 1.4). A large part of the context of speech comes from its embedding in the society and culture of its speakers. This field of study is **sociocultural linguistics**.

The field of **historical linguistics** examines how languages change over time. This historical perspective can be applied to all levels of language: sounds, words, structures, and meanings. Historical linguists are also interested in determining which languages are related and how they have descended from a mother language, which was spoken in the distant past (see Sidebar 1.5 for one such example). But languages don't evolve in isolation. Instead, they often influence each other as their speakers interact over time. The study of such **language contact** is a subfield of historical linguistics.

Our linguistic capabilities are critically embedded in our neurology and our ability to think. The field of **language and the brain** examines the physical and neurological basis of language, while **cognitive linguistics** looks at how language is instantiated by our broader cognitive processes. A related field is **language acquisition**, which studies how language is learned by children (**first language acquisition)** and by adults (**second language acquisition**).

Computational linguistics is a field at the intersection of linguistics and computer science that deals with the statistical or rule-based modeling of natural language. It is concerned with applying methods from artificial intelligence and machine learning to problems involving language. The recent acceleration of our technological abilities has led to a greater application of computational methods to a wide range of linguistic questions, such as how languages are learned.

When we look at sentence structures across languages, we notice that languages differ in the relative ordering of the subject (*Chris* in *Chris ate the apple*), the object (*the apple*), and the verb (*ate*). There are six logically possible orderings of these three categories:

Subject-Object-Verb   *Subject-Verb-Object*
Object-Subject-Verb   *Object-Verb-Subject*
Verb-Subject-Object   *Verb-Object-Subject*

However, all six orderings are not equally instantiated in the world's languages. A famous study of these orderings found that languages which put the subject first are very common, those that put the verb first are much less common, and those that put the object first are very few indeed. Why this should be, and the theoretical implications of this fact, is a question addressed by **linguistic typologists**.

We find languages throughout the world. The field of **typology and universals** looks at how the world's languages are similar and different. See Sidebar 1.6 for an example of this. Typologists are interested in developing a classification of languages based on how they are structured, and in looking for relationships between certain structural language types.

There are many applications of linguistics to situations in the world around us. The field of **applied linguistics** includes a number of subfields, including language teaching and **forensic linguistics**. Recently, there has been a strong move toward **language documentation**, the creation of a record of a language that can be used by speech communities and others in the face of possible endangerment or language death. Of course, linguistics is also a key part of the field of **speech pathology** and **speech and hearing sciences**.

This list of subfields of linguistics is fairly representative but is certainly not exhaustive. While we will not be able to touch on all of these fields in this book, we will cover most of them. The fields are diverse enough that there is usually something to interest everyone, and some readers will find that they are interested by everything. I find that the longer I study linguistics, the more interesting the field becomes. I hope you have the same experience.

CHAPTER SUMMARY

Human languages are complex, structured, and dynamic systems of human communication, which change over time under a variety of influences. While it is impossible to exactly count the number of languages of the world, our current estimate is in the range of 7,000. However, these are not evenly distributed, as most of the world's population speaks one or more of a small number of dominant languages, while a small percentage of the population speak one of many languages with comparatively few speakers, many of which are endangered.

Linguistics is the scientific study of language. It is empirical and objective. Linguists seek to state succinctly the structural properties of languages, and to understand their interactions, how they change, and how they serve the broader functions of language as a tool of communication that is embedded in human physiology, cognition, interaction, society, and culture. Explaining how individual languages work and how language works more broadly constitutes the aim of linguistic theory.