

Questions linguists ask

What is linguistics?

Linguistics is the scientific study of language. Linguists study language just like other scientists study other natural phenomena, such as photosynthesis, the solar system, DNA, or climate change. Though language has been an object of study since ancient times, from ancient Egypt, Greece, and Rome to India, China, and beyond, language scholars have had somewhat different aims, exploring different questions. Linguistics as we know it today is a relatively new field of study, based on the ideas of American linguist Noam Chomsky's 1955 *The Logical Structure of Linguistic Theory*, and subsequent works (see especially Chomsky, 1957 and 1965). The goal of what has come to be called **generative linguistics** is to understand the nature of human knowledge of language; that is, to understand that grammar is a system of rules that 'generates' all and only the grammatical sentences in a language. Linguists seek to discover what we, as speakers/signers, must know in order to produce and understand a language.

Note: When we talk about language here and throughout the book we mean both spoken and signed languages, and when we refer to a 'speaker' of a language generally, we mean any user of a spoken or signed language. We also discuss aspects of sign language specifically in several sections.

There are many ways to pursue the scientific study of language, and as in any scientific discipline, researchers have different goals,

different questions to pose, and different directions of research. Some linguists study the grammar, or rule system, of a language, and others are more interested in the social factors, such as gender, age, ethnicity, and other variables, that influence how we use language. Still others study how languages change over time, how children acquire language, or how our brains process and produce language. These are just a few of the diverse ways that we can study language scientifically, and we will explore them in more detail as we progress through this book.

“Studying linguistics has helped me to be more aware of all the amazing, intricate, and subconscious systems that allow communication to happen.”

We start with the basic but important question at the heart of linguistics: *What is language?*

What is language?

A typical answer to the question *What is language?* is that language is a mode of communication. This is true, of course, but species other than humans certainly communicate as well. So do bees have language? Do whales and dolphins, wolves and chimpanzees? By exploring how some animal communication systems are similar to and different from human language, we gain some insights into the nature of human language.

Human language versus other animal communication systems

When discussing different species' communication systems it is important to first clarify the question we are asking. Here are two questions that approach the issue of animal versus human communication from different perspectives.

How are animal communication systems similar to or different from the human communication system (language)?

Can animals understand/learn/produce human language?

It is one thing to examine ways in which communication systems are similar and different (the first question), and quite another to investigate whether one species can be taught to produce and understand the communication system of another species (the second question). There are many examples of researchers who have tried to teach human language to animals (in particular to primates such as Nim Chimpsky, Washoe, Kanzi, and Koko the gorilla, with different levels of success, as in Terrace, 1979, Fouts, Fouts, & Schoenfeld, 1984, Savage-Rumbaugh & Lewin, 1994, Patterson 1978, 1981). In no known case, however, has a primate fully mastered language in the way that a human child does by age 4–5. (See, for example, Brown, 1973; Crain & Lillo-Martin, 1999.) We will explore the first question, as it provides insights into the features of human language, and helps us understand the larger question, *what is language?*

The linguist Charles Hockett came up with a list of *design features* (or characteristics) of communication systems that we can use as a starting point (Hockett, 1960). He argued that while all communication systems have some of these features, only human language has all of them. (Hockett proposed 16 features in all; we provide only a representative sample here).

Hockett's design features of language

Semanticity: signals in the system have meanings

Words have meanings in human language, and so do some of the sounds that animals make. Vervet monkeys have different calls to alert each other to different predators (leopards, eagles, and snakes: see Seyfarth, Cheney, & Marler, 1980). Prairie dogs have different barks for different predators (coyotes, skunks, and badgers: see Slobodchikoff, 1998).

Arbitrariness: there is no logical connection between the form of the signal and what it means

There is an arbitrary connection between the word *dog* in English and the animal it refers to. We know this because the word for *dog* is different in other languages (*sobaka* in Russian; *aso* in Tagalog; *sq^wabay?* in Lushootseed; *mwba* in Swahili). Animal systems also have arbitrariness: there is no logical connection between the vervet monkey's call and a snake, nor between a prairie dog's bark and a badger.

Discreteness: messages in the system are made up of smaller parts that can be combined, rather than of single, indivisible units

Words are made up of smaller parts. For example, in human language we combine words and parts of words into larger words: *dog* + *s* = *dogs*; *dis* + *crete* + *ness* = *discreteness*. We combine single sounds into larger units: *s* + *t* + *o* + *l* combine to form the word *stole*, which we can transcribe (using the International Phonetic Alphabet) as /stol/. Sentences in turn are made up of combinations of words: *The dog stole the bone*. Vervet calls and prairie dog barks, on the other hand, can't be further broken down into meaningful parts; these sounds also cannot be combined with other sounds to express different meanings.

Productivity: users can produce and understand messages never heard before

You have likely neither heard nor produced the sentence *There is a beautiful unicorn on my yacht*, but you understand it. There do not seem to be any animal communication systems in which signals can be recombined in ways that produce novel messages.

Displacement: users can communicate about things that are not present

We can talk about things in the past or future, things that we hope will happen and things that happened long before we were born. Animal systems seem restricted to communicating about things in the present; Vervets don't communicate about snakes they may encounter tomorrow, and prairie dogs don't communicate about a badger they ran into last week.

It appears, then, that although animal communication systems are fascinating, and certainly worthy of study, these systems differ from human language in important ways.

Although we might be able to broadly characterize human language in terms of a series of design features, this is not the whole story, and not a complete description of our linguistic behavior. There are many other aspects of language to investigate and explore.

“Studying linguistics has helped me to understand humans better.”

What is a language?

We have briefly explored the very broad question *What is language?* We have seen that we can make some important distinctions between human language and other communication systems. But we also talk about *languages* as specific entities; we all know someone who speaks a language other than our own, and many of us have studied a language other than our own in school. The question that arises in this context is: *What is a language?*

Language variation: languages and dialects

We all know that although different speech communities may speak a single language, the actual varieties of that language spoken in each community—and even among individuals within that community—might be quite different. This makes it difficult to draw the line between what we call **a language** and **a dialect**, or

variety, of that language. For example, although we might say that people in France speak French, they of course do not all speak the same variety, or dialect, of French. French varies across France and even within communities in the same region. French is also spoken in many other countries in Europe and around the world (in Belgium, Switzerland, Canada, Mauritania, Morocco, Haiti, and others). What, then, do all of these language varieties have in common and how are they different from each other, and why do we (do we?) consider them all to be varieties of the same language, *French*?

Linguists generally consider different language varieties to be dialects of the same language if the speakers of those different varieties can understand each other, if there is—as linguists call it—**mutual intelligibility**. Languages, on the other hand, are varieties that are not mutually intelligible. But this way of distinguishing languages from dialects is not always useful. Speakers of what might be called two different languages might actually be able to understand each other, but the two are nevertheless called distinct languages for political, social, or geographic reasons. This is the case for speakers of Dutch and Flemish, Swedish and Norwegian, Macedonian and Bulgarian, Malaysian and Indonesian—each is considered a different language because it is spoken in a different country, and is an important part of a speech community's identity. Serbs and Croats can understand each other perfectly well but consider their languages distinct and even use different writing systems. (The Eastern Orthodox Serbs use the Cyrillic alphabet, while the Roman Catholic Croats use the Latin alphabet.) Other speech communities may be united by a shared writing system. In China, more than a billion people speak at least seven mutually unintelligible varieties of Chinese; thus, the varieties could be considered different languages. Speakers of Cantonese, for example, cannot understand speakers of Mandarin unless they learn it as a separate language, and vice versa. However, because the people are in a single nation and because they use a single writing system, most speakers consider the various varieties to be Chinese.

Language variation therefore raises a number of interesting questions, and is the focus of much study in linguistics. For example, linguists might study how and why a language may vary over space—from region to region. Linguists have determined, for instance, that some of the **regional** differences among dialects of American English can be attributed to the different colonial settlement patterns up and down the east coast of the United States. The slave trade and the contact between English and African languages also contributed to the rise of different **ethnic** varieties of English, including Gullah, spoken in the coastal areas of South Carolina, Georgia, and Florida, and African American English, spoken all over the United States. Many other variables can influence language variation, including **gender** and **sexuality**, **age**, **socioeconomic status**, and others. In fact, it is difficult if not impossible to study language variation in terms of a single variable or factor. We may make broad generalizations about pronunciations and vocabulary that distinguish different regional or ethnic varieties of American English, for example, but when we examine speech communities more locally, what are referred to as **communities of practice**, the linguistic landscape becomes quite a bit more complex. (See Eckert & McConnell-Ginet, 1992.)

For example, African American English (AAE), an ethnic dialect, is spoken all over the United States, and has a range of features and rules that distinguish it from other varieties of English. In AAE ‘habitual’ *be* indicates an ongoing state of affairs: *The coffee be cold* means *The coffee is always cold*. *The coffee cold*, where *be* is deleted, however, is equivalent to *The coffee is cold right now* in other varieties of American English. But AAE also varies from region to region; in Los Angeles, for example, AAE is influenced by Chicanx English (and vice versa), but the AAE spoken in Chicago is not (Wolfram & Schilling, 2015).

AAE also varies in terms of local communities of practice. Mallinson and Childs (2007) studied eight AAE-speaking women in a rural Appalachian community whose linguistic practices divided them into two groups, the ‘church ladies,’ who avoided certain AAE features, and the ‘porch sitters,’ who did not. These

linguistic differences aligned with each group's different ideologies about religion, education, and femininity. The church ladies gather at church to discuss family, the Bible, and church business. Their speech is more aligned with notions of what is considered proper English and proper decorum outside of the community. The porch sitters talk about men and their neighborhood, and identify more closely with race and the activities of the local community. The language of these eight women therefore cannot be described strictly in terms of norms of ethnicity or region, but rather in terms of the complex interplay of variables that characterize their different speech communities.

The study of language variation, or **sociolinguistics**, focuses on language variation in any kind of social context. Research in sociolinguistics includes a wide range of topics and variables, and we have only provided a few examples here. But this short discussion provides us with some insights into the question *What is a language*, by illustrating how the labels we use for different languages and language varieties can obscure the dynamic and complex nature of actual linguistic practice. We discuss aspects of sociolinguistic research in more depth in Chapter 8.

“Studying linguistics encourages welcoming and understanding of all other language varieties and their accompanying cultures without criticism of any kind, but rather analysis.”

Language change

It is also hard to pin down exactly what a language is because not only do languages vary from speech community to speech community, they also change over time. Taking English as an example, we might refer to the varieties of English spoken around the world today as varieties of Present Day English (PDE), even though they vary from each other in many ways. We also assume that the language

Table 2.1 Time periods of the English language

Old English	500–1050
Middle English	1050–1450
Early Modern English	1450–1700
Present Day English	1700–present

varieties from which PDE developed over time are all also English, even though they bear little resemblance to the English of today.

Old English, *Middle English*, *Early Modern English*, and *Present Day English* refer to English spoken in the British Isles during particular historical periods. These labels and dates are arbitrary; Old English didn't change into Middle English overnight in 1050, and Middle English didn't magically morph into Early Modern English in 1450. Rather, English changed gradually over time, for a variety of reasons (social, political, cultural, and linguistic). Nevertheless, we can still study these earlier varieties of English and determine that English—at least the varieties we have records of—during these historical periods shared certain grammatical characteristics, and that these characteristics gradually changed, often as a result of historical events.

The following selection from *Beowulf* (which appeared in written form around 1000 CE) illustrates how different Old English (also referred to as Anglo-Saxon) is from Present Day English.

hwæt	we	gardenas		in	geardagum
Listen!	we	of the Spear-Danes		in	days-of-yore

þeodcyninga	þrym	gefrunon
of those clan-kings	the glory	have heard

hu	ða	æþelingas	ellen	fremedon
how	those	nobles	courageous deeds	performed

'Listen! We have heard of the courageous deeds of the Danes (and) their kings in earlier times, how the noble ones accomplished courageous deeds.'

One is hard-pressed to find more than a few recognizable words in the *Beowulf* passage (*we, in*), and if you were to hear this passage read aloud (or sung, as it typically was at the time), you might notice that <hw> in *hwæt* is pronounced as <wh>, that <g> in *geardagum* is pronounced as <y>. You might also notice that *Beodcyninga* is a single word rather than, as in PDE, a phrase with separate words ‘of those clan-kings.’

Now jump ahead several hundred years to Chaucer’s time; here we can see how the language has changed, due in part to the effects of the Viking Invasions of the British Isles from the eighth to eleventh centuries, and of the influence of Norman French, spoken by the aristocracy after the Norman Invasion of England in 1066. The following passage is from the *Prologue* to *The Canterbury Tales*, written at the end of the fourteenth century, during the Middle English period.

A knyght ther was, and that a worthy man,
That fro the tyme that he first bigan
To riden out, he loved chivalrie,
Trouthe and honour, fredom and curteisie.
Ful worthy was he in his lordes were,
And therto hadde he riden, no man ferre,
As wel in cristendom as in hethenesse,
And evere honoured for his worthynesse.

Though the word order is still different in some ways from PDE, as is some of the vocabulary and spelling, the passage is far more understandable to contemporary English speakers than the excerpt of *Beowulf*. If you were to hear this passage read aloud in Middle English, you would notice that the vowel sounds are quite different from today’s, though again, you would still be able to understand quite a bit of what is being said. Some words are even identical to their PDE counterparts: *and, as, for, in, no, were, and worthy, tyme, loved, honour*, etc.

Consider too how word meanings have changed over time. In Middle English, a *knave* was simply a male child, not a rascal.

Nice meant ‘foolish’ and *silly* meant ‘weak.’ Around 1600 *decimate* meant ‘to kill one in every ten,’ and 600 years ago *myriad* meant a specific number (10,000 to be exact), rather than meaning ‘a lot,’ as we use it today. These words are all still in the English language, which suggests that at some point, some speech community adopted the new meaning. So we can study how a language has changed over time, what we call studying language **diachronically**, or we can study a language as it exists at a certain point in time, **synchronically**.

The study of language change includes the study of language attitudes about change. There have been attempts to create language academies to protect languages from changing; *l’Académie Française* was formed in 1635, and the Italian *Accademia della Crusca* in 1582, but these efforts rarely succeed. Proposals for an English Academy arose around this time as well, championed by authors including Daniel Defoe and Jonathan Swift, but the attempt ultimately failed. Though Samuel Johnson’s dictionary in 1755 was recognized as a kind of authoritative work on the English language, Johnson himself later recognized and embraced the inevitability of language change. Today, dictionaries, from the venerable *Oxford English Dictionary* to the *Dictionary of American Regional English*, to the online *Urban Dictionary*, reflect how word meanings shift and change over time. Modern lexicographers are language historians, cataloging what we say and how we use language.

There are many different topics that linguists study when they study language change, or **historical linguistics**. They may trace the origins of language families, to determine how languages are related to each other. They might study the origins of the words in a language, or etymology. Still other topics include how languages change in terms of syntax or word order, or how pronunciations change over time. What precipitates change? In addition to historical events such as wars, exploration, and colonization, disease, climate, population, famine, migration, and other factors shape language change.

A *language*, then, is not a static thing that we can pin down and define; rather, languages are dynamic systems that change over

time and vary over space. What we call a ‘language’ is actually a continuum of language varieties that have changed over time. In Chapter 7, we’ll explore more of the ways in which linguists study **historical linguistics** or **diachronic linguistics**. We leave the question of *What is a language?* and turn to a related question of great interest to linguists, namely, *What does it mean to know a language?*

What does it mean to know a language?

As we mentioned in the Introduction to this book, one of the central questions linguists ask is *What does it mean to know a language?* That is, what is it that we humans need to know in order to produce and understand whatever language we speak or sign?

Our linguistic intuitions

Any English speaker, no matter where they are from, knows that the following sentence is not a natural sentence of English. (We use the linguists’ convention of * to mark words or phrases that do not naturally occur in a particular language.)

*The dog at girl the barked.

Though a native English speaker can probably figure out what the sentence means based on the words in the sentence and on what they know about the world (about people and dogs), they also know that there is something odd about the order of words in this sentence, and that a more natural order would be:

The dog barked at the girl.

This suggests that native English speakers have some knowledge of **syntax**, or word order, knowledge that allows speakers to distinguish between a syntactically possible (natural) sentence of English and a syntactically impossible (unnatural) one.

There are many other ways to illustrate our intuitive knowledge of language. English speakers know, for example, that even though the sentences below look completely parallel in terms of word order, their interpretations are quite different.

The teacher is easy to please.

The teacher is eager to please.

In the *easy* sentence, the teacher is easy for someone else to please. In the *eager* sentence it is the other way around; the teacher is eager to please someone else. These differences in meaning are illustrated below.

The teacher is easy (for someone else) to please (the teacher).

The teacher is eager (for the teacher) to please (someone else).

As illustrated above, each of these sentences includes elements that we understand to be there, even though those elements are not pronounced. If there were no such understood elements, we would not be able to explain the very different interpretations of each sentence. This suggests that part of our unconscious knowledge of syntax includes knowledge of the possible word orders in our language, and of **understood elements**—structure that is there, but just not pronounced or signed—in a sentence.

There are many other such syntactic patterns and puzzles, in English and in all languages, signed or spoken, and exploring them leads us to a deeper understanding not only of our own knowledge of language, but of the structure of other languages, and how languages are similar and how they are different.

For example, consider this sentence:

The fox chased the rabbit.

Any English speaker knows that we can rearrange the order of words in this sentence in the following way:

The rabbit was chased by the fox.

We have, in this case, taken an *active* sentence and made it *passive*, something English speakers intuitively know how to do, even though it might be hard to explain exactly how it's done. English speakers also know how to form a question out of the first sentence:

Did the fox chase the rabbit?

And out of the second:

Was the rabbit chased by the fox?

Discovering the knowledge underlying word order and how pieces of sentences can be rearranged, what linguists refer to as **movement**, is a topic of much study in syntax.

Speakers also know how to interpret this sentence:

The rabbit was chased by the fox but the squirrel wasn't.

This sentence again involves *understood elements*, in this case, some syntactic material that we might think of as missing, a process that linguists refer to as **ellipsis** or **deletion**. We can represent deletion as follows:

The rabbit was chased by the fox but the squirrel wasn't ____.
The rabbit was chased by the fox but the squirrel wasn't
(*chased by the fox*).

English speakers have no trouble understanding the meaning of the missing material here.

As you can see from this short discussion, speakers' knowledge of syntax includes far more than knowledge of basic word order. We also know how to interpret material that may go unpronounced; we know how to rearrange words into different possible orders; and we understand sentence structure even when material is missing or deleted.

These and other aspects of our linguistic knowledge are not the kinds of things about language that we learn from our teachers or parents or friends; this is **unconscious** knowledge of language that

we are never taught. Nevertheless, language users must have this knowledge in order to produce and understand their language(s), and no native speaker knows a language better than another.

“Studying linguistics has helped me to listen really well and pay attention when people speak. It’s truly amazing how much we know without knowing we know it.”

Linguistic competence and linguistic performance

Let’s explore this notion of our linguistic knowledge in a little more depth. Suppose a native speaker of English makes a slip of the tongue, saying *You hissed my mystery lecture* rather than *You missed my history lecture*. Suppose someone says *hot* when they mean to say *cold*, or someone can’t seem to find the right word, even though they know what they want to say. Do these things mean the speaker doesn’t really *know* English, and that someone else, who doesn’t make such ‘errors,’ knows English better? From the point of view of the linguist, not really; speech ‘errors’ such as slips of the tongue, errors you make when you are tired or nervous, starts and stops, and so on, are part of our **linguistic performance**. While speakers and signers may differ in terms of linguistic performance, we do not differ in terms of **linguistic competence**, our intuitive knowledge of the rule system of our native language.

Although we tend to value certain kinds of linguistic performance (Dr. Martin Luther King Jr., for example, is well known in the United States as an accomplished orator), and we may label someone as more “well spoken” than another person, these are judgments based on social criteria (incidental criteria that vary across individuals and cultures), rather than on language users’ unconscious knowledge of language. The idea that there is a distinction between our unconscious knowledge of language and the language we actually produce goes back a long way (at least from the French philosopher René Descartes to Swiss linguist Ferdinand de Saussure and beyond), and this important distinction has radically changed the questions we ask about language and how we study it.

Let's take a closer look at our intuitive knowledge of language (our linguistic competence). So far, we have discussed some of the things we know about **syntax**, or word order. But there are other things we need to know about language in order to produce and understand it, including rules of pronunciation (**phonetics** and **phonology**), word structure (**morphology**), and meaning (**semantics**). These different components make up what linguists call the **grammar** of a language.

What is grammar?

Here, we briefly explore this notion of grammar, with examples from English (though we could use any language, signed or spoken, to illustrate). As you'll see, *grammar* as linguists study it is quite different from the way you may have studied grammar in school or heard about it in wider public discourse.

Our knowledge of grammar

Consider the following well-known sentence, attributed to Noam Chomsky:

Colorless green ideas sleep furiously.

Although this sentence is nonsensical, there is still something about it that seems natural; all the words are in the appropriate order for a typical English sentence. Rules of meaning, or **semantics**, are broken here, but not rules of syntax, or word order. More specifically, native English speakers know that *ideas* is an abstract, inanimate noun, and therefore can't sleep, nor can abstract nouns have colors, and be green. *Colorless* and *green* present a contradiction; something both with and without color at the same time. And *ideas*, as inanimate abstractions without volition, can't do anything *furiously*.

There are many other semantic puzzles that offer insights into this component of our linguistic knowledge. We know, for example, that the following newspaper headline is ambiguous; it has more than one meaning. What are these meanings?

Milk Drinkers Are Turning to Powder

This headline exemplifies **lexical ambiguity**: the ambiguity can be attributed to ambiguous words. The phrase *turning to* can be interpreted as ‘choosing’ (powdered milk over liquid) or as ‘becoming’ (the drinkers themselves are turning into powder). The word *powder* can also be ambiguous, meaning powdered milk or powder of some other kind.

And what about the sentence below? Assume that this one is not lexically ambiguous. What other reason could there be for its ambiguity?

They love English cotton clothes.

This sentence is **syntactically ambiguous**; its different meanings are based on syntax, not on the ambiguity of a particular word. We can explain the ambiguity by dividing the words up into groups, or phrases, in two different ways.

They love [English [cotton clothes]].

They love [[English cotton] clothes].

In the first sentence, *English* modifies *cotton clothes* (the cotton clothes are English), but in the second, *English cotton* modifies *clothes* (the clothes are made of English cotton). This sentence has two different meanings because it has two different structures!

Here’s a fun one: *I love ambiguity more than most people*. Can you explain why this is ambiguous?

The study of semantics sometimes overlaps with the study of logic; we know, for example, that if a tulip is a flower and that flowers are plants, then a tulip is a plant (a logical syllogism that involves the semantic relationship of **entailment**):

A tulip is a flower

Flowers are plants

A tulip is a plant

We also understand complex relationships among words that express quantity, or what linguists call *quantifiers*. Consider the following sentence for example:

Everybody loves somebody.

This sentence has two interpretations—that everybody loves the same person (*Evan, Lee, and Cary all love Isabelle*) or it can mean that each person loves a different person (*Evan loves Lee, and Cary loves Isabelle, and so on*). These interpretations have their source in something called **quantifier scope**, the interaction between the two quantifiers *every* and *some*.

Even though we may not be familiar with the terms *ambiguity*, *entailment*, and *quantifier scope*, we, as language users, understand these and many other complex semantic relationships.

We also know quite a bit about the structure of the words in our nonsense sentence. We know, for example, that *-less*, the suffix on *color*, turns the noun *color* into an adjective, *colorless*. We also know that *-less* can be attached to other nouns like *fear*, *penny*, and *defense* (*fearless*, *penniless*, *defenseless*), but not to verbs, like *meet*, *teach*, or *resemble* (**meetless*, **teachless*, **resembleless*). We know that *-s* on *ideas* indicates plurality, and that the *-ly* on *furi-ously* makes the adjective *furious* an adverb. We therefore have extensive knowledge of **morphemes**, or parts of words, and of **affixation**, the rules of suffixation and prefixation in a language like English. Even though you may not be able to describe these relationships in technical terms, part of your knowledge of English includes the knowledge of word structure, or **morphology**.

Some other examples of our morphological knowledge include knowledge of how to form new words using **word formation rules**, as illustrated below with examples of various types:

Acronyms: FOMO, SARS

Initialisms: NBA, PTA

Clipping: totes (from *totally*), pants (from *pantaloons*)

Blends: ginormous (*gigantic* + *enormous*), apathetic (*apathy* + *pathetic*)

Coining: bling, google

Compounding: bailout, upgrade

Eponyms: Achilles' heel, cardigan

Conversion: troll (as a noun and as a verb), text (from noun to verb)

Backformation: *liaise* (from *liaison*), *enthuse* (from *enthusiasm*)

A type of misinterpretation, which nevertheless illustrates how we analyze word structure, is called an **eggcorn** (a mispronunciation of *acorn*, depending on your dialect). Some eggcorns that you may be familiar with include *doggy-dog world*, for *dog-eat-dog world*, *death nail* for *death knell*, and *all intensive purposes*, for *all intents and purposes*. (Linguists have created an eggcorn database to keep track of these misinterpretations, many of which make more sense than the original pronunciations! Take a look! <https://eggcorns.lascribe.net/page/1/>)

In addition to rules of syntax, semantics, and morphology, we also know how to pronounce words in our language. Speakers of a language know the inventory of sounds in their language, and how those sounds can be combined. The study of the inventory of sounds is called **phonetics**, and the study of the sound system, how we combine sounds in a language, is called **phonology**.

Languages may differ in terms of their phonetic inventories, but each chooses from a set of *possible* linguistic sounds produced by the human articulatory system, which are distinct from other sounds we can make that are not linguistically meaningful (*vroom vroom*, *tsk*, *tsk*, a kissing sound, a raspberry, etc.). The International Phonetic Alphabet, or IPA, has been developed as a way of graphically representing the sounds of human language.

A phonetic transcription of an English pronunciation of the sentence we gave above, *Colorless green ideas sleep furiously*, using the symbols for English, follows:

[kʌlɪʃəs ɡɪn aɪdɪəz slɪp fɹɪˈɹɪəsli]

(The brackets indicate that this is written using phonetic transcription. In Chapter 3, we discuss the use of brackets [] compared to slashes //.)

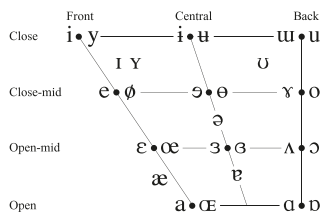
© 2015 IPA

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b		t d			ʈ ɖ	c ɟ	k ɡ	q ɢ		ʔ
Nasal	m	ɱ	n			ɳ	ɲ	ŋ	ɴ		
Trill	ʙ		r						ʀ		
Tap or Flap		ⱱ	ɾ			ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative			ɬ ɮ								
Approximant		ʋ	ɹ			ɻ	j	ɰ			
Lateral approximant			l			ɭ	ʎ	ʟ			

Symbols to the right in a cell are voiced, to the left are voiceless. Shaded areas denote articulations judged impossible.

VOWELS

Clicks	Voiced implosives	Ejectives
◌b Bilabial	◌ɓ Bilabial	◌' Examples:
◌d Dental	◌ɗ Dental/alveolar	p' Bilabial
◌ɟ (Post)alveolar	◌ɟ Palatal	t' Dental/alveolar
◌ɟ̥ Palatoalveolar	◌ɟ̥ Velar	k' Velar
◌ɬ Alveolar lateral	◌ɠ Uvular	s' Alveolar fricative



Where symbols appear in pairs, the one to the right represents a rounded vowel.

SUPRASEGMENTALS

- | | | | |
|---|-----------------------------------|---|-------------------------------------|
| Λ | Voiceless labial-velar fricative | Ç | Alveolo-palatal fricatives |
| W | Voiced labial-velar approximant | ɹ | Voiced alveolar lateral flap |
| ɥ | Voiced labial-palatal approximant | ɥ | Simultaneous ɥ and X |
| H | Voiceless epiglottal fricative | | Affricates and double articulations |
| ʕ | Voiceless epiglottal fricative | | can be represented by two symbols |
| ʕ | Epiglottal plosive | | joined by a tie bar if necessary. |

$$\underline{\text{ts}} \quad \widehat{\text{kp}}$$

- | | | |
|---|------------------------------|--------|
| | Primary stress | |
| | Secondary stress | |
| ː | Long | eː |
| ˑ | Half-long | eˑ |
| ̤ | Extra-short | ẽ |
| | Minor (foot) group | |
| | Major (intonation) group | |
| | Syllable break | ˌi.ækt |
| ˌ | Linking (absence of a break) | |

DIACRITICS Some diacritics may be placed above a symbol with a descender, e.g. $\overset{\circ}{\eta}$

◌ Voiceless	ᵀ ᵀ	◌ Breathy voiced	ᵇ ᵇ	◌ Dental	ᶑ ᶑ
◌ Voiced	ᶞ ᶞ	◌ Creaky voiced	ᵇ ᵇ	◌ Apical	ᶑ ᶑ
ᵇ Aspirated	ᵀ ᵀ	◌ Linguallabial	ᶑ ᶑ	◌ Laminal	ᶑ ᶑ
◌ More rounded	ᵇ	◌ Labialized	ᵇ	◌ Nasalized	ᶑ
◌ Less rounded	ᵇ	ᶑ Palatalized	ᶑ ᶑ	ᵇ Nasal release	ᵇ
◌ Advanced	ᵇ	ᶑ Velarized	ᶑ ᶑ	ᶑ Lateral release	ᶑ
◌ Retracted	ᵇ	ᶑ Pharyngealized	ᶑ ᶑ	ᶑ No audible release	ᶑ
◌ Centralized	ᵇ	◌ Velarized or pharyngealized	ᶑ		
◌ Mid-centralized	ᵇ	◌ Raised	ᵇ (ᶑ = voiced alveolar fricative)		
◌ Syllabic	ᵇ	◌ Lowered	ᵇ (ᶑ = voiced bilabial approximant)		
◌ Non-syllabic	ᵇ	◌ Advanced Tongue Root	ᵇ		
◌ Rhoticity	ᵇ	◌ Retracted Tongue Root	ᵇ		

TONES AND WORD ACCENTS

LEVEL		CONTOUR			
ẽ	or ˘	Extra high	ẽ	or ˆ	Rising
é	˘	High	ê	˘	Falling
ē	˘	Mid	ẽ	˘	High rising
è	˘	Low	ẽ	˘	Low rising
ẽ	˘	Extra low	ẽ	˘	Rising-falling
↓		Downstep	↗		Global rise
↑		Upstep	↘		Global fall

Figure 2.1 The International Phonetic Alphabet (revised to 2015)

Consider some of the ways in which linguistic inventories vary. The /r/ sound is pronounced differently in English, French, and Spanish. Arabic and Hebrew, but not English, have pharyngeal consonants. The Bantu languages Xhosa and Zulu have click consonants as part of their phonetic inventory. Samoan has neither /t/ nor /n/, sounds that occur in many other languages. Languages also have different vowel inventories; French, Portuguese, and Polish have nasal vowels, and in the Mon language of Myanmar and Thailand, vowels with high tones (tones are another phonetic distinction among languages) are pronounced with creaky or breathy voice, a pronunciation that does not exist in English in the same way.

Phonology, as we noted above, is the study of the way sounds from a language's phonetic inventory are combined. English speakers know, for example, that the plural *-s* at the end of *ideas* is actually pronounced more like the sound /z/ than the sound /s/, but that the *-s* at the end of *colorless* is pronounced /s/ rather than /z/. Speakers also know that *sleep* and *green* both begin with consonant clusters (/sl/ and /ɡr/) that we find in other English words (*slow*, *slovenly* and *grab*, *grovel*) but that we don't find words in English that start with /ls/ or /lg/. In fact, the longest possible sequence of initial consonants (or consonant cluster) in English is /stɹ/ as in *strike* and /spl/ as in *splat*, but in Georgian, a language spoken in the Caucasus mountains, words can begin with four, five, or even six consonants. Hawaiian and Hausa, on the other hand, have no consonant clusters at all!

Our phonological rule system accounts for other things about our pronunciations. For example, we might pronounce *I can bake* as *I cam bake* (because of a process called **assimilation**), and we might pronounce *cold cuts* as *col cuts* (due to a process of **deletion**). We also insert sounds into some words (through a process called **insertion** or **epenthesis**): *athlete* and *realtor* become *athalete* and *realator*, and we switch sounds around sometimes (**metathesis**) *cinnamon* → *ciminon*, *animal* → *aminal*.

Phonological rules also explain patterns of **stress**. Consider the following example, where stress shifts when we increase the number of syllables:

grámmar/grammátical/grammaticálicity

But that in other words, stress remains on the same syllable, no matter how many more we add:

forgíve/forgiveable/unforgíveable/unforgiveableness

Our phonetic inventory, and how sounds are combined in a particular language, account for our different **accents**, and also account for different pronunciations among speakers of the same language, who speak different varieties, or dialects, of that language (as we discussed briefly above). In most varieties of Canadian English *Mazda* is pronounced with the vowel sound in *cat*, but in most varieties of American English with the first vowel sound in *father*. In some parts of the United States, the words *cot* and *caught* are pronounced the same (as are *don* and *dawn* and *hock* and *hawk*), but in other parts of the country, these words are pronounced differently. Many dialects of English around the world drop /ɹ/ after a vowel, characterized by the sentence *Pahk the cah in Hahvahd yahd*. (Interestingly, this well-known sentence does not accurately characterize how ‘r-drop’ works; speakers would pronounce the /r/ in *car* here, because it is followed by a vowel in the next word, *in*—a little known component of this rule.)

Our knowledge of syntax, semantics, morphology, phonetics, and phonology makes up what linguists refer to as our knowledge of the **grammar** of a language. This definition of grammar is likely very different from other definitions you are familiar with, and it isn’t surprising, then, that the way linguists study grammar is quite different from how you may have studied grammar in school.

Though we have introduced the idea that grammar (in the sense we mean here) includes rules, these are not the kinds of rules that you may be familiar with. **Grammatical rules** in the linguist’s sense are more accurately defined as the natural laws and principles that govern our biological capacity to produce and understand language. We will continue to explore the notion of grammar, and of grammatical rules, in the following sections.

Prescriptive and descriptive grammar

The definition of grammar as a system of syntactic, semantic, phonological, and morphological rules is not the definition of grammar that most people learn about. Grammar in the linguist's sense is a *description* of a language user's unconscious knowledge of grammar (rules/principles of phonetics and phonology, morphology, syntax, and semantics).

Descriptive grammar: the set of unconscious rules that allow us to produce and understand a language.

Descriptive grammar differs quite dramatically from what we typically learn about grammar in U.S. schools or in popular discourse about language. You may have been told, for example, to use *whom* in certain positions instead of *who*; to write *To whom did you write the letter?* rather than *Who did you write the letter to?* You may also have been told to avoid 'double negatives' such as *I don't know nothing about that*, and to say instead *I don't know anything about that*. You may have been told to say *Owen and I went to the store*, even though it may seem more natural to you to say *Me and Owen* or *Owen and me went to the store*. We are typically taught—sometimes directly and sometimes indirectly—that certain ways of speaking and writing (what we might refer to as someone's grammar) are 'correct,' and that other ways are 'incorrect' or even 'lazy' or 'sloppy,' or flat-out wrong.

The idea that we should follow certain rules in order to speak and write English correctly has a long history, beginning in seventeenth-century England, when grammarians sought to 'fix' the language in some pure form (usually the form that corresponded with their own speech and writing, which is perhaps not surprising). These grammarians sought to *prescribe* rather than *describe* rules of language, and the sets of rules they came up with are rules of **prescriptive grammar**, rules that govern what we *should* say (according to some authority).

Prescriptive grammar: rules that prescribe what we should say, according to some authority.

Prescriptive rules are very different from descriptive rules; the latter seek to describe what we *actually* say (regardless of whether someone thinks what we say is correct or not). For many speakers, sentences like *Who did you talk to?* and *I don't know nothing about that*, or the use of *ain't* or of *like* are perfectly natural, suggesting that these sentences conform to those speakers' unconscious grammatical rules. Such utterances are, from a descriptive point of view, completely **grammatical**, even if from someone's prescriptive point of view these sentences might be considered incorrect or ungrammatical, and viewed as errors, because they don't match up with whatever is taken to be, by some language authority, to be acceptable, or correct. We have all been corrected for something we've said or written by teachers, grandparents, siblings, friends, or even strangers, and we have probably also corrected others for something they have said, signed, or written that we consider wrong. We all (even linguists) have our own language pet peeves! And we are also familiar with the notion of 'grammar' not only as a list of 'dos and don'ts' of spoken language, but also as writing conventions, spelling, and punctuation rules. This view of grammar is very different from descriptive grammar, which is (ideally) based on the unconscious grammatical system that underlies languages.

As we have said above, one of the central goals of linguistics is to study language scientifically, to determine the underlying system of rules that allows us to produce and understand language. To do this, linguists must gather language **data**, by observing what speakers actually say, rather than what speakers *should* say, according to some authority. It follows, then, that the scientific study of language, or more specifically grammar, is based on descriptive grammar rather than prescriptive grammar.

This is not to say that the study of prescriptive grammar isn't also interesting and valuable. Where did these prescriptions come from, and who decides what is 'correct' and 'incorrect'? What is the legacy of prescriptive grammar today? What is the role of prescriptive grammar in education? Who is a language authority and why? Has technology changed our perceptions of what is prescriptively grammatical? These are all interesting

questions, and exploring them tells us a great deal about **language attitudes** and how they shape how we perceive ourselves and others.

“Studying linguistics has allowed me to more clearly explain to others the importance of an egalitarian attitude towards language. It is astounding to realize how language snobbery contributes to discrimination, and the degree to which people make value judgments based on prescriptive grammar and dialect. I want to help people understand these perspectives by offering them persuasive evidence, which I now have the tools to do.”

Note, too, that there is not always a clear distinction between prescriptive and descriptive grammar. When you study another language in school, for example, you learn many of the descriptive grammatical rules of that language, along with some prescriptive ones. For some language users, prescriptive rules align with descriptive rules (which again tells us something about both). Suppose, for instance, that in your dialect, you would say *I didn't know anyone at the party*. A speaker of another dialect of English might say *I didn't know no one at the party*.

I didn't know anyone at the party.

I didn't know no one at the party.

Although each utterance is descriptively grammatical for the speaker of that language variety, the second version (using *no* words in negative sentences: *no one*, *nowhere*, *nothing*) can be socially stigmatized outside of the communities in which it is used, while the first version (using *any* words in negative sentences: *anyone*, *anywhere*, *anything*) may be less socially stigmatized.

I didn't go anywhere yesterday.

I didn't know anything about that.

I didn't go nowhere yesterday.

I didn't know nothing about that.

And if a speaker of a dialect that employs *I didn't go nowhere* uses the more 'school appropriate' *I didn't go anywhere* in their own speech community, this can also be stigmatized; here the speaker is adopting a structure that is not part of that speech community's dialect. Because of societal language attitudes, certain dialects are valued more highly than others, and it is often the case that the dialects more highly socially valued are those that we also assume to be aligned with prescribed rules about language.

Universal Grammar

The descriptive grammar of a language that we speak or sign can be characterized as our **mental grammar**, the set of rules that make up our unconscious knowledge of language. In fact, there is evidence that humans are 'hardwired' to acquire language, which suggests that there is some aspect of language that for humans is innate. This does not mean that we are born with the knowledge of Swahili or Ewe or Japanese imprinted in our brains, but rather that we all, under normal conditions, have the capacity to acquire the language, or perhaps more accurately, the *grammar* of the language, we are exposed to. This predisposition for language can be compared to birds' predisposition to fly; no one teaches baby birds to take wing and leave the nest, but at some point they do. Being predisposed to acquire language means that we have some innate ability to tease out the grammar of French or Bengali or Wampanoag if we are exposed to it.

Because we humans all have basically the same brain, which is hardwired in the same way biologically, linguists hypothesize that the languages that we acquire will also share the same basic blueprint, or set of basic grammatical principles. That is, all languages we acquire are *natural* languages, governed by the same set of natural laws and/or principles. This set of natural laws or principles that all languages share is sometimes referred to as **Universal Grammar (UG)**.

Assuming UG, the grammars of languages around the world will have the same basic architecture, even though on the surface they may look very different. Linguists therefore study the grammatical properties of different languages to discover the principles and properties those languages have in common, and to find explanations for the apparent differences among them.

In later sections we will provide examples of this kind of cross-linguistic research and what we learn from it. Before we do that, however, we'll briefly explore some evidence for UG and the human capacity for language.

First language acquisition

One strong piece of evidence that humans are hardwired to acquire language comes from first language acquisition. Children around the world, whether hearing, deaf, or Deaf (where capitalized Deaf refers to the Deaf culture and community), regardless of family structure and culture, acquire their first language in the same basic developmental stages as long as they are exposed to it.

Stages of first language acquisition

The babbling stage (4–8 months)

The one-word stage (9–18 months)

The two-word stage (18–24 months)

The early multiword stage (24–30 months)

The later multiword stage (30 months and older)

Children also, quite amazingly, acquire language without direct instruction; no one teaches us all the possible sentences in a language, yet we know how to produce and understand complex sentences we have never been exposed to by the time we are about 5 years old. Children's 'mistakes' also tell us quite a bit about our knowledge of language. Consider this example: suppose a child hears the sentences *We painted the green chair* and *We painted the chair green*. You might think that the child would infer that the adjective *green* and the noun *chair* can occur in two different orders (*green chair* and *chair green*). We might expect then, that

when the child encounters any sentence with the order adjective + noun, that this order may be switched. But this is not the case. Though a child might produce *We painted the chair green*, she will never produce **We saw the chair green*. Why don't children make this error? Because they do not learn by analogy; their acquisition is guided by linguistic rules (in this case, the varying syntactic properties of verbs; see Pinker, 1994). There are many such examples of how child language acquisition is guided not by instruction, correction, or reinforcement, but rather by innate knowledge of language. We discuss the stages of first language acquisition in more detail in Chapter 9.

Christopher the linguistic savant

The case of a man named Christopher, a linguistic savant, provides yet more fascinating evidence for the human capacity for language and UG. Neil Smith, of University College London, and Ianthi-Maria Tsimpli, of the University of Cambridge, have long studied Christopher, a linguistic genius who knows some 20 languages with varying levels of fluency (Smith & Tsimpli, 1995). Christopher has a 'perfect' ability to acquire language even as an adult, but he is in other ways severely cognitively impaired.

In one experiment, Smith and Tsimpli tried to teach Christopher an artificial language, Epun, with properties that do not exist in actual natural languages, but which is quite easy to learn. Christopher could not master this artificial language, because, Smith and Tsimpli argue, its rules did not conform to the blueprint of natural language, Universal Grammar. In another experiment they taught Christopher some Berber, a language in which speakers can drop the subject of the sentence in ways that are impossible in English. (Berber is a 'null subject' language, as are Spanish, Modern Greek, and many others. In Spanish, for example, one can say *tengo frio* 'have cold' rather than *yo tengo frio* 'I have cold/I'm cold.' The separate pronoun *yo* 'I' does not need to be pronounced.) However, the data Smith and Tsimpli provided to Christopher never included sentences with null subjects.

Nevertheless, Christopher began producing sentences with null subjects on his own. This suggests that Christopher was able to deduce the grammatical properties of Berber from the limited linguistic input provided to him, just as children deduce grammatical properties of the language without hearing every possible sentence in their language. Christopher's language ability provides compelling evidence for an innate human capacity for language and for UG.

Contact languages and language genesis

Pidgin and creole languages also offer insights into the human capacity for language and UG. A **pidgin** language develops when two or more different languages come into contact (pidgins and creoles are therefore referred to as **contact languages**), and speakers of those languages need a way to communicate with each other. Pidgins are not 'broken' languages, as they are sometimes portrayed in the media; they are simplified linguistic rule systems that share similar features regardless of the languages they emerge from.

Pidgins have sprung up all over the world, most prominently as a result of European colonialism and the slave trade of the eighteenth and nineteenth centuries. Pidgins have developed in West Africa and the Caribbean, often based on the colonial languages English, French, Portuguese, Spanish, and Dutch, in contact with African languages. Others have emerged in the Far East (Chinese Pidgin English, Yokohama Pidgin Japanese), and in the Pacific (the Micronesian pidgin, Tok Pisin, and Hawaiian Pidgin English). Many pidgins have developed into creoles, and are spoken as native languages. (Creoles will be discussed further below.) Others are no longer spoken; Chinook Jargon (or Chinuk Wawa) was once widely spoken in the U.S. Pacific Northwest, and took features and vocabulary from Chinook, Nootka, and other languages, as well as English and French later.

In the development of a pidgin, the language of the group with less power is the **substrate** language, and the language of the group

with more power is the **superstrate** language. Typically, the substrate language influences the structure and meaning of the words in the pidgin, but the words themselves are borrowed from the superstrate language.

Here are some examples of how speakers create meaning with limited vocabulary characteristic of pidgin languages. Examples are from Tok Pisin, a Melanesian creole (which began as a pidgin) spoken in Papua New Guinea. The superstrate language is English, and the substrate languages include a variety of the native Austronesian languages, especially Kuanua.

- Reduplication: *sip* ‘ship,’ *sipsip* ‘sheep’
- Extension: *stap* ‘stop’ means ‘be located, to remain, to continue,’ and also past tense
- Compounding: *maus gras* ‘mustache’ and *gras belong fes* ‘beard’

Some of the features (there are others) that all pidgin languages appear to share include:

- a smaller consonant and vowel inventory than most other languages
- absence of inflectional affixes (such as past tense and plural affixes)
- verb tenses are expressed by words rather than by verb conjugations
- subject-verb-object (SVO) word order
- limited vocabulary (that leads to reduplication, extension, compounding)
- simplified pronoun system
- simple sentence structure (few embedded clauses)

Here again we use examples from Tok Pisin (this example was recorded by Margaret Mead, as cited in Hall, 1966, p. 149). This is an early example of the language, in its pidgin form (written using IPA symbols, not the more standardized orthography).

naw mi stap rabawl. mi stap long biglajn, mi katam kopra. naw
Then I stay Rabaul. I was in workgroup, I cut copra. Then

wanfela mastər bılɔŋ kampani ɛm i-kɨɨm mi mi kɔk lɔŋ ɛm ɟem.
a white man from company he take me I cook for him again

mastər king. mi stap. naw ol mastər i-kɨk
Mister King. I stay. Then all white men were playing football

i- kɨkɨm ɛm. naw lɛɟ bılɔŋ ɛm i-swelup
They kick him. Then leg of him swell up.

You can see here that Tok Pisin has SVO word order (*mi katam kopra* / 'I cut copra') and that sentences are quite short and simple. Prepositions (*of, from*) are expressed by the single word *bılɔŋ*, and pronouns are simplified (*mi* = *I*, *ɛm* = *he* and *him*, *i-* = *they*). Past tense is expressed by the word *stap*, 'stay' rather than by a suffix on the verb such as *-ed*. Other verbs in the passage have no tense inflection: *katam*, 'cut,' *i-kɨɨm*, 'kick,' *kɔk*, 'cook,' and so on. You can also see that the words in Tok Pisin are similar to those in English, the superstrate language, but some have different meanings (*biglajn* 'big line' = 'workgroup,' *mastər* 'master' = 'white man,' etc.).

Over time, a pidgin can become more and more grammatically complex, and can eventually be acquired by children as a native language. At this point, the pidgin has developed into a **creole**, a language with a full-fledged grammatical system (Tok Pisin, Hawaiian Creole English, Jamaican Creole, Haitian Creole, Krio in Sierra Leone, among many others). Creole languages share some features with pidgins, and they also share features that are not found in non-creole languages. One of the most striking features of creole languages is the complex tense and aspect system. Below are some examples from three creole languages, Hawaiian Creole (English superstrate), Haitian Creole (French superstrate), and Sranan (English based, spoken in Suriname). (In French, *marcher* is 'to walk'.)

Table 2.2 Creole language comparisons

<i>Hawaiian Creole English</i>	<i>Haitian Creole</i>	<i>Sranan</i>
he walk (he walked)	il maché	a waka
he bin walk (he had walked)	il té maché	a ben waka
he stay walk (he is/was walking)	l'ap maché	a e waka

Source: (adapted from Bickerton, 1983)

There is much more to say about the features of both pidgin and creole languages, and about how these new linguistic systems emerge from language contact. What is important to note here is that pidgin and creole languages around the world share certain characteristics, regardless of the contact languages from which they emerge. One way to explain this striking similarity is to propose that in language contact situations, the human brain creates language based on certain universal principles of UG, and thus that these similarities are not accidental but rather are the result of our hard-wiring for language.

How do we study language scientifically?

As we have mentioned above, there are many ways to study language scientifically. We have briefly discussed some of the topics that sociolinguists and historical linguists study, and we have also seen that language acquisition and our biological capacity for language are important areas of research. The branch of linguistics that specifically focuses on studying the grammar of a language in order to discover the rules of syntax, semantics, morphology, phonetics, and phonology that make up speakers' unconscious knowledge of language is sometimes referred to as **theoretical** or **formal linguistics**. Linguists who focus on grammatical structure attempt to construct a model, or theory, of speakers' linguistic knowledge. Research in this branch of linguistic theory typically does not address social factors or context, but rather relies more on native speaker judgments about

natural language, though theoretical linguistics also uses corpus data and psycholinguistic experimentation.

“There are so many reasons I’m glad I’m studying linguistics! Linguistics has helped me to approach language scientifically, to appreciate the beauty of language, and to be excited by the possibilities of being surrounded by ambient language data that always leaves me with interesting questions to think about.”

In this section we provide only one example of how linguists study language scientifically. In later chapters, we discuss in more depth core areas of study in linguistics including phonetics and phonology, morphology and syntax, semantics and pragmatics, as well as how children acquire language, how language varies from region to region and community to community and even person to person, and how language changes over time.

Gathering and analyzing data: the scientific method

Thus far, we have discussed some of the questions linguists seek to answer, and we have asserted that linguists study language scientifically in order to find answers to these questions. We have yet to explain, however, how one actually does this. Researchers in other sciences, such as chemistry, physics, or biology, study phenomena that we can often observe and experiment with: melting polar ice caps, gravity, planetary motion, and so on. Scientists **observe** the world around them and ask **questions** about physical phenomena (Why is the sky blue? At what temperature does ice melt?) and **gather and study data** in order to formulate possible answers to these questions. Scientific inquiry therefore involves **forming** explanations, or **hypotheses**, about data, and then testing those hypotheses against additional data to see if they hold up. Scientists routinely **revise hypotheses** based on those data. Sets of hypotheses **make predictions**, and taken together,

one can then **propose a theory** of how to account for a particular phenomenon.

Below is a model of scientific inquiry, or the scientific method.

Scientific method

make observations → ask questions → gather data → form hypotheses → make predictions → gather more data → test and revise hypotheses → make predictions → propose theory → repeat

But how do we study language in this way? Language is, after all, not a physical object. Indeed, once we utter a word or form a sign, language seems to disappear into thin air. But in fact, there are limitless questions that we can ask about language, and there is a wealth of data to explore. Below, we provide an example of how we might study (a very specific) syntactic phenomenon in English using scientific inquiry.

“After studying linguistics, I finally understand what science really means—that it’s not biology or chemistry, but it’s a method of inquiry. And it’s been so exciting to discover that using language data!”

An example of scientific inquiry: english wh-question formation

One topic of scientific research in linguistics is syntactic theory, where linguists attempt to construct a model of sentence structure. The goal is to not only explain syntactic phenomena in a particular language, but to try to determine what syntactic properties all languages have in common. One phenomenon that occurs across languages is *question formation*; all languages seem to have a way of formulating questions, though they might differ syntactically and in other ways. As a starting point for researching question

formation, we will take a look at what we refer to as **wh-question formation** in English.

Here is an English sentence that any English speaker would consider grammatical:

Kim saw Cary with Lee.

And here is some more data—another natural English sentence:

Kim saw Cary and Lee.

These data seem pretty uninteresting. But we also might observe that we can form a question from the first sentence. In English, we do this via **movement**: we move an interrogative phrase, here, the word *who*, to sentence-initial position:

*Kim saw Cary with **Lee**.*

*Kim saw Cary with **who**.*

***Who** did Kim see Cary with___? (*who* = *Lee*)*

But notice that we can't form a question like this from the second sentence:

*Kim saw Cary and **Lee**.*

*Kim saw Cary and **who**.*

****Who** did Kim see Cary and___?*

Our two non-interrogative sentences look similar in terms of numbers of words and basic word order. The only difference is that in the first sentence the noun phrase *Lee* is preceded by *with*, and in the second, by *and*. So why can we form a question only from the first sentence but not the second?

We have determined that the only apparent difference between these two sentences is the words *with* and *and*. Linguists who study syntax know (based on much investigation of **syntactic categories**, or parts of speech) that these two words fall into different classes, or categories: *with* is a **preposition**, but *and* is a **conjunction**.

Prepositions: *to/at/by/to/with/for/about/after* etc.

Conjunctions: *for/and/nor/but/or/yes/so*

A linguist might **form the hypothesis**, then, that we can make questions with phrases following prepositions, but not with phrases following conjunctions, and that this has something to do with the syntactic differences between these two categories (preposition and conjunction). If this hypothesis is on the right track, it **predicts** that other prepositions will pattern like the *with* sentence and allow *wh*-question formation, but that sentences with conjunctions other than *and* will not.

Let's **gather more data** and **test our hypothesis**. We find that, as predicted, a sentence with a different preposition (*to*) allows *wh*-question formation, but one with a different conjunctions (*or*) does not.

Pete sells books to the students.

Pete sells books or magazines.

Who does Pete sell books to ____?

**What does Pete sell books or ____?*

By testing our hypothesis and the predictions it makes, even in this very limited way, we find further evidence that we can form *wh*-questions in sentences with phrases that follow prepositions, but not with those that follow conjunctions. These data support our hypothesis that prepositions and conjunctions, and the phrases they occur with, are syntactically different, even though in the sentences we are considering, this difference is not immediately apparent.

But now a new question emerges: what exactly is that difference? We want to *explain* why question formation works like it does, so we need to dig deeper.

As it turns out, there is a great deal of additional evidence that prepositions and conjunctions are syntactically distinct. Based on examining a wide range of additional data, linguists propose the schematic representation of this syntactic difference presented in Figure 2.2.

In the *with* sentence, the preposition forms a syntactic unit, or prepositional **phrase**, with its noun phrase (NP) object. This

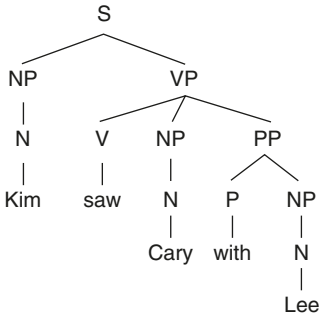


Figure 2.2

prepositional phrase is labeled PP, because the **head** of the phrase (the main word) is a preposition.

A *phrase*, then, is a syntactic unit that consists of at least a *head*, and possibly another phrase. In Figure 2.2 above, NP, VP, and PP are all phrases. (S, for ‘sentence,’ is also technically a phrase, but we’ll leave that aside for now.)

In Figure 2.3 below for the sentence with the conjunction *and*, the NP and VP are phrases (and there is no PP).

Figure 2.3 differs from Figure 2.2 with the PP *with Lee*. In the previous tree, the preposition *with* forms a phrase with the NP, *Lee*. In this tree, however, with the conjunction *and*, the NP *Lee* is conjoined to another phrase, NP *Cary*. The larger NP *Cary and Lee* forms a phrase with the V *saw*, resulting in the VP *saw Cary and Lee*.

We might hypothesize, then, that we form *wh*-questions only of NPs that form a phrase with another word (objects of prepositions, and objects of verbs, for example). We predict that the NP object of the preposition in Figure 2.2 can be questioned:

Who did Kim see Cary with____? (*who* = Lee)

And that an NP in a conjoined NP cannot form a question.

**Who* did Kim see Cary and____? (*who* = Lee)

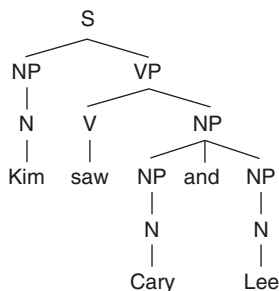


Figure 2.3

(We won't get into the details of the structure of questions here, such as what happens to the tense and why an auxiliary verb *did* is added when the statement becomes a question: Kim saw Cary with Lee → Who did Kim see Cary with?)

But note that we also predict that we should be able to form a question of the *entire* conjoined NP object of the verb *saw* in Figure 2.3—the (large) NP object [Cary and Lee]. This prediction is borne out:

Who did Kim see? (*who* = Cary and Lee)

Technical details aside, the important point here is that even though sentences may look superficially alike in terms of word order, on further investigation, we find that they can actually have very different internal structures, even though this difference is not immediately observable. And those syntactic differences provide us with a way to explain syntactic patterns, such as those we saw with *wh*-question formation in English. By studying linguistic data scientifically, we uncover the underlying structure of sentences and find explanations for phenomena that otherwise would remain mysterious.

This is just a micro-example of how a linguist might go about studying language (in this case, the syntax of *wh*-question formation) scientifically. Regardless of the technical details (the

process of any kind of scientific inquiry is a process, and therefore difficult to explain without some level of detail), this brief example illustrates how it is indeed possible to gather data, form hypotheses, make predictions, and test those hypotheses against additional data. We are therefore able to, ultimately, **construct a theory** of *wh*-question formation, a series of hypotheses that predict when question formation is possible, and when it is not.

“Studying linguistics has helped me to (1) analyze data, (2) become a better researcher, (3) appreciate diversity. I’m so happy I have come to learn more about this field.”

Exploring data from other languages

Linguists who study syntactic phenomena such as *wh*-question formation in a particular language are also interested in how that phenomenon works in other languages, and/or if other languages even have *wh*-question formation at all! Exploring data from a range of other languages provides us with a bigger picture of syntactic structure and UG, and of how questions operate *across* languages. Theories we might construct from such data tell us more about the common properties that languages share around the world, and help us explain why and how they are different.

For example, we’ve seen that in English, the *wh*-word or phrase appears in sentence-initial position in *wh*-questions. But now let’s look at how other languages form questions.

Consider French, a language that has SVO word order like English.

S	V	O	
Tu	manges	une	pomme.
You	eat	an	apple
‘You are eating an apple.’			

And like English, French forms *wh*-questions of the NP object of the verb with the *wh*-word or phrase in sentence-initial position:

O	S	V
Qu'est-ce que	tu	manges ____?
what is-it that	you	eat
'What are you eating?'		

We might hypothesize from these data that SVO languages form *wh*-questions as in English and French. We might also hypothesize that this is the only way languages form *wh*-questions—the question word or phrase occurs in sentence-initial position.

But when we look at more language data, we find that neither of these hypotheses hold up.

Aleut, or Unangam Tunuu, a language spoken in the Aleutian Islands, Pribilof Islands, and the Commander Islands, has subject-object-verb (SOV) word order, unlike English and French. (The object of the verb is in boldface here.) (Data from Michael Lestenkof, personal communication.)

S	O	V
Piitrax̂	Ivaanax̂	ilaġuliŋ.
Peter	John	help
'Peter is helping John.'		

Aleut does have *wh*-question formation, but not as in English. In Aleut, the *wh*-word or phrase occurs in the same position as it does in non-questions.

S	O	V
Piitrax̂	kiin	ilaġuliŋ?
Peter	who	help
'Who is Peter helping?'		

This suggests that our hypothesis that all languages form *wh*-questions with the question word or phrase in sentence-initial position fails to explain the data; in some languages (such as

Unangam Tunuu) the question word stays put. We might hypothesize, then, that this is because Aleut has subject-object-verb word order (SOV) rather than SVO. So we might hypothesize that only SVO languages form questions with the question word or phrase in sentence-initial position.

But consider Chinese, which, like English, has SVO word order.

S	V	O
ni	maile	pingguo
you	bought	apples
'You bought apples.'		

Even though Chinese has SVO word order like English, *wh*-questions are formed as they are in Unangam Tunuu, where the question word stays put, in the same position that the object would be in non-questions.

S	V	O
ni	maile	shenme?
you	bought	what
'What did you buy?'		

These data make us question our hypothesis yet again. Clearly, there is no one single way to form such information-seeking questions across languages, and it's also clear that word order alone does not predict how a language will form these questions. But what we do see is that there are at least two different ways to form questions: the question word or phrase occurs sentence-initially as in English and French, or it occurs 'in-situ,' as in Unangam Tunuu and Chinese.

As you can see from this short discussion, there is much left to discover about how question formation operates across languages, and the more language data we study, the more we gain insights into this phenomenon. As in other areas of scientific inquiry, there is always more data to study that leads us to revise our hypotheses. Studying language scientifically, just like studying other biological phenomena, is an ongoing process of discovery.

In this chapter we have explored some of the basic questions about language that linguists strive to answer, and we provided a brief overview of how linguists study language scientifically to construct theories of language that attempt to answer those questions.

In the course of exploring these questions, in order to provide you with a better idea of what linguistics is, we touched on some of the main areas of study in linguistics:

- the study of grammar (phonetics and phonology, morphology, syntax, semantics)
- the study of how languages change over time
- the study of how language varies by community, location, situation
- the study of how language is acquired, produced, and processed in the brain

In the next few chapters, we delve into each of these areas of study in more depth, to provide an overview of some of the diverse and overlapping kinds of research that linguists do. We have already introduced the notion of *grammar* as the system of rules of phonetics and phonology, morphology, syntax, and semantics, our unconscious knowledge of language. Suppose you wanted to study one or more of these areas in depth? What kinds of things would you study, and what kinds of questions would you, a language scientist, ask? We will return to each area of the study of grammar mentioned here to further explore how and why linguists study it.

References

- Bickerton, D. (1983). Creole languages. *Scientific American*, 249(1), 116–122.
- Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press.
- Chomsky, N. (1955). *Logical structure of linguistic theory*. Manuscript.
- Chomsky, N. (1957). *Syntactic structures*. The Hague, The Netherlands: Mouton.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

- Crain, S., & Lillo-Martin, D. (1999). *An introduction to linguistic theory and language acquisition*. Oxford, England: Blackwell.
- Eckert, P., & McConnell-Ginet, S. (1992). Think practically and look locally: Language and gender as community-based practice. *Annual Review of Anthropology*, 21, 461–490.
- Fouts, R., Fouts, D.H., & Schoenfeld, D. (1984). Sign language conversational interaction between chimpanzees. *Sign Language Studies*, 42(1), 1–12.
- Hall, R.A. (1966). *Pidgin and creole languages*. Ithaca, NY: Cornell University Press.
- Hockett, C. (1960). The origin of speech. *Scientific American*, 203, 88–96.
- Mallinson, C., & Childs, B. (2007). Communities of practice in sociolinguistic description: Analyzing language and identity practices among black women in Appalachia. *Gender and Language*, 1–2, 173–206.
- Patterson, F. (1978). Linguistic capabilities of a young lowland gorilla. In F.C. Peng (Ed.), *Sign language and language acquisition in man and ape: New dimensions in comparative pedolinguistics* (pp. 161–201). Boulder, CO: Westview Press.
- Patterson, F. (1981). Ape language. *Science*, 211(4477), 86–87. doi:10.1126/science.211.4477.86-a
- Pinker, S. (1994). *The language instinct*. New York, NY: Penguin.
- Savage-Rumbaugh, S., & Lewin, R. (1994). *Kanzi: The ape at the brink of the human mind*. New York, NY: John Wiley and Sons.
- Seyfarth, R.M., Cheney, D.L., & Marler, P. (1980). Monkey responses to three different alarm calls: Evidence of predator classification and semantic communication. *Science*, 210, 801–803.
- Slobodchikoff, C. (1998). The language of prairie dogs. In M. Tobias & K. Solisti-Mattelon (Eds.), *Kinship with the animals* (pp. 65–76). Hillsboro, OR: Beyond Words Publishing.
- Smith, N., & Tsimpli, M.I. (1995). *The mind of a savant: Language learning and modularity*. Oxford, England: Blackwell.
- Terrace, H. (1979). *Nim*. New York, NY: Knopf.
- Wolfram, W., & Schilling, N. (2015). *American English* (3rd ed.). Malden, MA: Wiley-Blackwell.