Chapter 2

Language

In this chapter, we outline how language works in very general terms. We begin with a tour of some of the more common and egregious misunderstandings about language and then progress to the key notions of any theory of language. Finally, we briefly review the principal areas of language study.

2.1 Common Misconceptions

Language is a wonderful domain to do science with. All humans have language in some sense, so the data one might need to test one’s hypotheses are ready to hand. Unfortunately, there is a great deal of misunderstanding about how language works. We need to break through these misunderstandings before we can appreciate the formal structures of language.

2.1.1 Learning is memorization

Let’s begin with a fairly common idea: children learn language by memorizing words and sentences. The basic idea here is that all there is to learning a language is memorizing the words and sentences one hears.

There are several problems with this view of language learning. First, it is not altogether clear what is intended by memorization, but the most obvious interpretation of this term would have to be incorrect. Memorization implies a conscious active effort to retain something in memory, but language learning seems to happen in a rather passive fashion. Children do not appear to spend any effort in acquiring the languages they are exposed to, yet they
CHAPTER 2. LANGUAGE

will learn any number of them simultaneously.

Memorization is certainly a tool by which things can be learned, but it is much more appropriate as a way of characterizing how we learn, say, the multiplication tables, rather than how we learn language.

Another problem with the memorization story is that children produce things that they haven’t been exposed to. For example, it is quite typical that at an early stage of acquisition children will produce forms like *foots* for the adult form *feet*. This has no explanation on the memorization view. The child has presumably only heard the correct adult form *feet*, yet produces a form that they’ve never been exposed to. Presumably, the child is attempting to produce the plural form by the general rule of adding an *–s*, despite having heard the irregular plural form *feet*.

Similar examples can be constructed at the sentence level. Imagine the child has been exposed to a novel situation, say, a new toy or a new person in their life. The child isn’t stumped by these sorts of situations, but constructs sentences appropriate to the new individuals. That is, the child has no trouble saying that *Ernie is in the kitchen*, even if the child has just met Ernie and it’s the first time he’s been in the kitchen.

Both sorts of example suggest that learning is at least partially a process of *generalizing* from what the child has been exposed to, not simple rote memorization of that exposure.

### 2.1.2 Correction is necessary to learn language

Another common misconception related to language learning is that children need to be corrected by adults to learn their language properly. We’ve already cited examples where children produce forms that are incorrect in the adult language, e.g. *foots*. It is reasonable—though incorrect—to suppose that they need explicit correction to produce the correct form.

There are two glaring problems with this view. First, not all parents correct their children, yet these sorts of errors disappear in the adult language, typically by age three or four.

Another perhaps rather surprising fact is that children typically ignore any attempt at correction. The literature is rife with anecdotes of children being painstakingly corrected for such forms and then blithely continuing to use them.

The point is that these sorts of “errors” are a natural part of language acquisition. They develop and fall away of their own accord, and parental
correction plays essentially no role in this.

2.1.3 Some people have bad grammar

Most of us believe that some people have “bad grammar”. This is a very tricky notion so it will be useful to go through some examples. Consider the following pair of sentences.

(2.1) You and me are studying logic.
     You and I are studying logic.

The two sentences mean the same thing. Both are used in English. Some English speakers prefer to use one over the other and some speakers will use both in different contexts. If a speaker does use both, then the difference is one of register or modality. The first is more typical of speech, rather than writing. The first is also more informal than the second.

Is either form better than the other? Not really. It is certainly the case that they are contextually restricted, but there is no objective scientific reason to maintain that either structure is “better” than the other.

One might note that the second sentence type is older while the first type is newer. This is indeed correct, but it does not make either structure better or worse than the other. In fact, there are cases that work the other way. For example, the word ain’t is generally thought to be “bad grammar”, but is well attested in the history of English.

The idea that some constructions are “bad” has led to some interesting developments in the history of English. In the case of the X and me pattern above, the idea that X and I is to be preferred to X and me has led to some strange reversals. For example, the first sentence type below is what is attested historically, yet the second sentence type now shows up quite often.

(2.2) Between you and me, logic is interesting.
     Between you and I, logic is interesting.

In fact, for some speakers, the avoidance of X and I goes further, including pairs like the following.

(2.3) Logic interests you and me.
     Logic interests you and I.
Here, as with the previous pair, the first of the two sentences is what is attested historically. What’s happened is that first \textit{X} and \textit{I} extends to new environments, as in the first pair. Presumed “language authorities” decry this development and speakers “overcorrect” to avoid the supposed bad structure. This is called \textit{hypercorrection}.

In all these cases, it is important to keep in mind that neither structure is intrinsically better or worse than the other. There are contextual and stylistic differences, especially in writing, but these are matters of style and custom, not a matter of goodness or badness.

\subsection*{2.1.4 Two negatives make a positive}

Consider a sentence like the following:

\begin{equation}
\text{(2.4) Ernie didn’t see nothing.}
\end{equation}

Sentences of this sort are often cited as examples of the illogic of bad grammar. A sentence like this—with the interpretation that Ernie didn’t see anything—is claimed to be an instance of bad grammar. It is bad grammar because it is said to be illogical. It is taken to be illogical because there is an extra negative, \textit{didn’t} and \textit{nothing}, and the extra negative should make the sentence positive.

It is indeed the case that two negatives \textit{sometimes} make a positive. For example. If I assert that \textit{Ernie saw no chickens when he was in the yard} and you know this to be false, you might then say:

\begin{equation}
\text{(2.5) Ernie did \textit{not} see no chickens when he was in the yard.}
\end{equation}

Here, the first negative \textit{not} denies the second negative \textit{no}. This is in contrast to the preceding example, where the first negative \textit{didn’t} emphasizes the second negative \textit{nothing}.

It is certainly the case that using a second negative for emphasis is stylistically marked in English. It is indicative of speech, rather than writing, and it is more casual. But is there something illogical about using a negative word for emphasis? No. Words and structures in a language can be used for many purposes and there is no particular reason why a word that has a negative meaning in one context cannot have another meaning in some other context.
In fact, this pattern is quite common in other languages. French, Spanish, and Russian all use two separate negative words to express *nothing*.

(2.6) French: Ernie n’a vu rien.
Spanish: Ernie no vio nada.
Russian: Ernie nichego ne videl.

Thus using two negative words to express a single negative idea is quite normal crosslinguistically.

It is true that the logical system we introduce in chapter 4 has the property that stacking up two negatives is equivalent to having no negatives, e.g. \( \neg \neg p \leftrightarrow p \). This simply means that logical negation is more restricted in how it is to be interpreted than negative words in human languages.

### 2.1.5 I don’t have an accent

Everyone speaks their own language differently from other speakers of the same language. These differences can be in what words a person uses, how they pronounce those words, how words are combined to form sentences, etc. Some of these differences are completely idiosyncratic and reflect individual variation. Some of these differences reflect a speaker’s geographic or social origins. A person’s completely unique language is termed an *idiolect*; a set of features that reflect a particular geographic area and/or social distinction is called a *dialect*.

Let’s first look at some pronunciation differences. Some speakers in the southern US pronounce words like *pen* and *pin* the same, while most northern speakers keep these distinct. Likewise, some northern speakers make a distinction between the initial sounds of words like *witch* and *which*, while other speakers do not.

There are lots of regional differences in terms of word choice. For example, dialects vary dramatically in how they refer to a “soft drink”: *soda*, *pop*, or even, used generically, *coke*. A very interesting example is *bubbler* which means a drinking fountain. This term is only used in western Wisconsin and New England. The term apparently was the brand name for a particular drinking fountain that was sold in just those areas of the country.

There are also differences in terms of word order or grammar. For example, there are dialects of English in which the following are acceptable and mean different things:
(2.7) Ernie eat it.
   Ernie be eating it.

The first simply means that Ernie is eating something. The second means
that Ernie habitually eats something. The structure is quite interesting be-
cause the distinction is not one that most dialects of English make and be-
cause the construction is stigmatized.

Another construction that shows up in some dialects is the double modal
construction, e.g. I might could do that. Finally, there are a number of dialects
that distinguish direct address to one person you vs. direct address to more
than one person: you all, y’all, you guys, yous.

The point here is that everyone’s language reflects their geographic and
social origins. Different dialects exhibit different degrees of social acceptabil-
ity, but there is no intrinsic qualitative difference between dialects, nor does
anyone not have a dialect.

2.1.6 Some languages are logical

It is common to ascribe different qualities to languages. For example: Ger-
man is ‘logical’, French is ‘beautiful’, or Russian is ‘harsh’. These terms have
nothing to do with the language in question and typically have much more
to do with our own prejudices about the people and culture.

2.1.7 Some languages are primitive

Various languages and dialects are often decried as ‘primitive’. Usually, this
is an instance of the type of misconception just above. For example, Native
American languages like Navajo or Tohono O’odham (Papago) are described
as primitive languages, not because there is anything especially primitive
about them, but because of unfortunate ambient attitudes that Native Amer-
ican cultures are primitive in some way.

Sometimes the argument is more sophisticated. For example, the dialect
of English cited above where a sentence like Ernie eat it is acceptable is cited
as primitive because it is “missing” the –s suffix that occurs in more accepted
dialects of English. Such arguments are specious and opportunistic. Thus,
in the case at hand, the dialect is cited as more primitive than other dialects
because it is missing the –s suffix, yet it could, by parity of argument, be
cited as less primitive than other dialects as it makes an aspectual distinction with be that other dialects do not make.

Sometimes the argument is even put in the opposite direction! Eskimo is cited as more primitive because it very quaintly has many words for snow, while English presumably has fewer. In fact, it’s been shown that Eskimo does not have any more words for snow than other languages. Moreover, people like skiers and snowplow drivers, whose jobs or recreation depend on snow, have many more words for snow. Is their English somehow more primitive?

Occasionally, the argument is made in terms of communicative limits. For example, one might argue that French is a highly evolved language since it has words for concepts like détente or ennui. This is specious too, however. French does indeed have the words cited, but the concepts are not difficult to express in words in English: “a release from tension” or “weariness and dissatisfaction”. Moreover, there are equivalent concepts that appear to be hard to express in French, but are easy in English: weekend, parking (lot), etc.

Every language will express some concepts as individual words and others as combinations of words. It’s not at all apparent that there is any rational basis to which is which in any particular language.

Finally, the character or existence of a writing system is sometimes cited as evidence of a language’s primitive nature. Thus, Navajo might be cited as “primitive” because it did not have a writing system until fairly recently, while English has had one for thousands of years.

Writing systems are certainly a valuable piece of cultural technology, but that is probably the best way to describe them: as technology. Thus a language with a writing system has a tool at its disposal that is quite useful. What’s important to keep in mind, however, is that the presence of a writing system does not appear to correlate with any aspect of language structure.

2.2 Key Notions

In the previous section, we spent a fair amount of time showing what language is not. In this section, we proceed to define language in a positive fashion.
2.2.1 What is Language?

What is language? Most people would define it as some sort of communication system. This is certainly true, but it is general enough to include other communication systems as well. For example, is Morse Code an instance of language? Are traffic lights—red, yellow, and green—language?

To distinguish what we think of as language from these other systems, we need to take account of the fact that language is a more complete system. In fact, language is arguably capable of expressing virtually any idea that a human being is capable of thinking of. We can term this power expressive completeness and the notion will rear its head again in chapters 4 and 5.

In fact, it has occasionally been argued that language determines what we can think about. That is, it is impossible to think of things that we cannot put into words. This latter position is somewhat different from expressive completeness and highly controversial. It is referred to as the Sapir–Whorf Hypothesis.

Notice that expressive completeness also rules out inadvertent communication systems. For example, we might conclude that Ernie is sick from him coughing or sneezing, but we would certainly not want to characterize those as instances of language.

Expressive completeness also rules out facial gestures as an instance of language, whether those gestures be unintentional, like a smile, or deliberate, like a wink.

Thus we characterize language as an expressively complete conventionalized communication system. The requirement of expressive completeness rules out miniature or “toy” systems. Conventionalization rules out unintentional communication systems. Language, on this definition, is a subcase of a more general notion of communication.

Notice that this definition does not entail that a language must include a written form. In fact, the definition is neutral with respect to modality, allowing for signed languages like American Sign Language (ASL). The language also allows for other modalities as well, e.g. an exclusively written language.\footnote{This departs from the usual linguistic notion of language.}

Given this definition of language, we can ask whether animals have language. The question is actually a rather odd one and rests on what we mean by expressive completeness. If we mean that a language is complete with respect to any message a speaker of it might want to purvey, then surely some
animals have language in this sense. If, on the other hand, we mean that a language must be complete with respect to any message that we might want to convey, then probably not, assuming that there are no animals communicating about the full range of topics that humans communicate about.

2.2.2 Creativity

In the previous section, we showed that language learning cannot be reduced to memorizing words and phrases. Rather, learning a language involves learning patterns and then exploiting those patterns in potentially novel ways. Thus, for example, our knowledge about what constitutes a well-formed sentence of English does not evaporate when we are confronted with a new name. A child confronted with a new individual with a name she hasn’t heard before is perfectly capable of uttering novel sentences describing the new individual, e.g. Ernie is in the kitchen.

Thus knowledge of a language is knowledge of patterns.

One might believe that these patterns are simply the sensible ones, but this would be incorrect. One piece of evidence against this idea is that the patterns that work in a language are different from the patterns that work in other languages. For example, we saw that some languages use double negatives, and others do not.

Another argument against this idea comes from the fact that we have contrasts like the following

\[
\text{(2.8) Colorless green ideas sleep furiously.} \\
\text{Furiously sleep ideas green colorless.}
\]

Neither of these sentences make sense in any obvious way, yet the first is vastly more acceptable than the second. Both arguments support the idea that the patterns that are acceptable in our language are not governed by what makes sense.

These patterns are quite complex. For example, most speakers of English will prefer one of the following sentences to the other.

\[
\text{(2.9) Who did Ernie believe Minnie claimed that Bob saw?} \\
\text{Who did Ernie believe Minnie’s claim that Bob saw?}
\]

Both sentences, though long and a little complicated, make perfect sense,
yet one is much better than the other.\textsuperscript{2} The point is that our knowledge of our language depends on knowing some rather subtle generalizations about what makes a sentence acceptable. These generalizations have to do with the structures in a sentence and not with whether the sentence makes sense in some intuitive way.

### 2.2.3 Prescriptive vs. Descriptive Grammar

Another key notion in understanding language is understanding the difference between what a speaker knows about their own language and what so-called experts tell us about language. For example, as a speaker of English, I know that I can say \textit{You and me are studying logic}. However, as an educated English speaker, I know that we are not supposed to write such things down and instead are supposed to write: \textit{You and I are studying logic}.

When we study what speakers actually do, we are interested in \textit{descriptive grammar}. If, on the other hand, we are interested in the rules that are imposed on speakers, then we are interested in \textit{prescriptive grammar}. Both are quite reasonable areas of study. The first is more about individual psychology, what it is a person actually knows about their language, albeit unconsciously. The second is more about social systems, what aspects of language are valued or not in the society at large.

### 2.2.4 Competence and Performance

An extremely important but controversial distinction is that between knowledge of a language, or \textit{competence}, and the use of that language, or \textit{performance}.

We can make sense of this distinction by imagining what it would be like to study language if we did \textit{not} make this difference. We would study language by observing what people said. There are two problems with this. First, people occasionally make mistakes when they speak, or occasionally change their minds in the middle of a sentence. For example, in conversation, one frequently hears sentences like these:

\textsuperscript{2}Most prefer the first to the second.
(2.10) I think..., uh, what did you say?  
Did you read the..., oh yeah, now I remember.  
Ernie likes..., oh, hey, Hortence!  
I bed that rook, I mean, I read that book.

Any speaker of English would recognize these as not acceptable instances of English word order, yet we utter these sorts of things all the time. If our theory of language was based purely on the sentences we observed in the real world, we would have to account for these.

Another problem is that there are sentences that we find acceptable, yet do not utter. We’ve already treated some of these. There are acceptable, yet nonsensical sentences like the following.

(2.11) Colorless green ideas sleep furiously.

There are also sentences that refer to possible situations that we simply haven’t been confronted with:

(2.12) Hortence loves Ernie.

In this latter case, we may simply not know individuals with those names, or have contemplated their emotional attachment.

In fact, one can argue that there are an infinite number of these sentences. Consider the following series of sentences:

(2.13) Sophie likes logic.  
Ernie knows Sophie likes logic.  
Sophie knows Ernie knows Sophie likes logic.  
Ernie knows Sophie knows Ernie knows Sophie likes logic.  
...  

Or this one:

(2.14) What is one and one?  
What is one and one and one?  
What is one and one and one and one?  
...  

In each case, the series begins with a completely acceptable sentence. We can add words in a simple way up to infinity. Of course, eventually, these
become too long to understand or too long to utter before falling asleep, but there is no principled upper bound on acceptability. If we were to force ourselves to restrict our language data to only the set of observed utterances, we would completely miss the fact that the set of possible utterances is infinite.

These problems are what leads to the distinction between language competence and language performance. A science of language can then be based on either sort of data. We might choose to investigate language competence by looking at the judgments of well-formedness that a speaker is capable of. We might instead investigate language performance by looking at what type of utterances speakers will actually produce and comprehend.

This distinction held sway in linguistics for almost fifty years and is still generally held to be useful. It is, however, becoming more and more controversial. The principle problem is that closer investigation of judgment data shows that those data exhibit variability as well.

2.3 General Areas

Let’s now consider the basic areas of language study. We can divide these into two broad categories. First, there are the areas that concern the actual structure of a language. Then there are the areas that concern how those structures play out in various ways.

2.3.1 Structures of Language

The areas of language study that focus on the structures of language include: phonetics, phonology, morphology, syntax, and semantics.

Phonetics

Phonetics is concerned with the sounds of language. The empirical goal is to discover what sounds are possible in language and to try to explain why only certain speech sounds occur and not others.

The method of explanation is the physical setting of speech. Patterns of speech sounds are explained in terms of how the articulatory system or the auditory system works.

Let’s consider a couple of examples of Phonetic facts and phonetic explanations.
One phonetic fact that we have already mentioned is that all languages appear to include the vowel sound [a], as in the first syllable of English father. The usual explanation for this is that this is an extremely simple sound to produce. Oversimplifying somewhat, it is produced by just opening the mouth fully and voicing. If we assume that languages make use of sounds that are easy to produce before they make use of sounds that are hard to produce, this follows.

Another kind of phonetic fact concerns [t]-like sounds. In English, the sound [t], as in toe, is produced by putting the tip of the tongue against the alveolar ridge, the bony ridge behind the upper teeth. In Spanish and Russian, on the other hand, [t] is produced by putting the tip of the tongue slightly forward, against the back of the upper front teeth. Other languages include other [t]-like sounds where the tip of the tongue is placed further back on the roof of the mouth against the hard palate. There are no languages, however, where [t]-like sounds are produced by placing the tip of the tongue against the soft palate.

This should not be too surprising and follows directly from the physiology of articulation. The connections between the tongue and the floor of the mouth prevent the tip from reaching that far back (unless someone is unusually gifted).
Let’s consider one more example. English includes the sound [v], for example, in the word *van*. This sound is produced by bringing the upper teeth in close proximity to the lower lip and voicing. Spanish, on the other hand includes the sound [β], as in the word *cabo* [kaβo] ‘end’. This sound is very similar to English [v] except that one brings the two lips in close proximity, rather than lip and teeth.

What’s striking is that there is only one language in the world that appears to have both sounds: Ewe, spoken in Ghana and Togo. Why is this combination so rare? The explanation is that the sounds are so very similar that it is too hard to distinguish them. The sounds of a language tend to be distributed so that they are maximally distinct acoustically.

### Phonology

Phonology is similar to phonetics, except the focus is on the distribution of sounds and sound patterns as instances of cognitive organization, rather than physiology. There is therefore a natural tension between phonetics and phonology in terms of explanation. Both disciplines deal with sounds and generalizations about sounds, but differ in their modes of explanation.

The two disciplines differ in their methodologies as well. Phonetics is an experimental field and makes liberal use of technology to understand the details of articulation, audition, and acoustics. Phonetics largely focuses on the performance of sound systems. Phonology, on the other hand, makes much less use of technology and largely focuses on the competence underlying sound systems.

We’ve already discussed a range of sound system facts that seem quite amenable to phonetic explanation. What kinds of facts require a phonological story?

One very good candidate for phonology is syllable structure. All languages parse words into syllables. In English, for example, *hat* has one syllable, *table* two syllables, *banana* three, etc. It’s not at all clear how to define a syllable phonetically—some phoneticians even deny that there are syllables—so the syllable seems a reasonable candidate for a cognitive unit. The idea is that our psychology requires that we break words up into these units.

There are some interesting typological generalizations about how syllables work. For example, while all languages have syllables that being with consonants, e.g. in both syllables of English *happy*, not all languages have syllables
that begin with vowels, e.g. in both syllables of *eon*. So there are two broad categories of language along this dimension. First, there are languages where all syllables must begin with a consonant, e.g. Hawaiian. There are also languages where syllables can begin with either a consonant or a vowel. There are, however, no languages where all syllables must begin with vowels. This generalization too is thought to be a fact about our cognitive organization.

Phonology is also concerned with relations between sounds in utterances. For example, there is a process in many dialects of English whereby a [t] or [d] sound is pronounced as a *flap* when it occurs between appropriate vowels. A flap is produced by passing the tip of the tongue quickly past the alveolar ridge and sounds much like the [r]-sound of Spanish or Russian. We will transcribe a flap like this: [ɾ]. This process causes items to be pronounced differently in different contexts. For example, *write* and *ride* in isolation are pronounced with [t] and [d] respectively, but when they occur before a vowel, the final sounds are pronounced as flaps.

\[
\begin{array}{ccc}
\text{in isolation} & \text{write} & \text{ride} \\
\text{before a vowel-initial word} & [\text{rayt}] & [\text{rayd}] \\
\text{before a vowel-initial suffix} & [\text{rayr} \text{er}] & [\text{rayr} \text{er}]
\end{array}
\]

Not all languages do this and so this cannot simply be a function of the physiology. Characterizing these sorts of generalizations and the search for explanation in the domain of cognition are part of phonology.

**Morphology**

Morphology is concerned with the combination of meaningful elements to make words. It can be opposed to phonology, which we can characterize as the combination of *meaningless* elements to make words. For example, a word like *hat* is composed of three sounds, three meaningless elements: [h], [æ], and [t]. There is only one meaningful element: the word itself. A word like *unhappy* has a complex phonology—it has six sounds: [ʌ], [n], [h], [æ], [p], [i]—and a complex morphology: it is composed of two meaningful elements: *un-* and *happy*. The element *un-* is a prefix, an element that cannot occur alone, but can be attached to the left of another element. It expresses negation.

There are also suffixes. For example, the word *books* is composed of a
stem *book* and a suffix *–s*, which expresses plurality. Prefixes, suffixes, stems, etc. are called *morphemes*.

The most important thing to keep in mind about morphology is that it can, in some cases, be boundless. Hence the set of possible words in a language with boundless morphology is infinite. Here is an example from English where a set of suffixes can be attached without bound.

\[(2.17)\]

\[
\text{nation} \\
\text{national} \\
\text{nationalize} \\
\text{nationalization} \\
\text{nationalizational} \\
\text{nationalizationalize} \\
\ldots
\]

Another extremely important point is that elements—*morphemes*—are not combined in a linear fashion, but are nested. For example, a word like *nationalize* has three morphemes that are grouped together as represented in the tree below.

\[(2.18)\]

\[
\text{nationalize}
\]

This structure can be important to the meaning of a word. Consider a compound word like *budget bottle brush*. This has two different meanings associated with two different structures.

\[(2.19)\]

\[
\text{budget} \quad \text{bottle} \quad \text{brush} \quad \text{budget} \quad \text{bottle} \quad \text{brush}
\]

The structure on the left is associated with the meaning where it is an inexpensive brush for bottles; the structure on the right is associated with the meaning where it is a brush for inexpensive bottles.

The two ideas come together in the following examples. A word like *unforgivable* has a single meaning: not able to be forgiven. A word like
unlockable actually has two meanings: not able to be locked and able to be unlocked. The ambiguity of the second word correlates with the two different possible structures for it.

(2.20)

\[
\text{unlockable} \quad \text{unlockable}
\]

The word unforgivable only has one meaning because only one structure is possible:

(2.21)

\[
\text{unforgivable}
\]

No other meaning is possible because the other structure below is not possible.

(2.22)

\[
\text{unforgivable}
\]

The reason this latter tree is not possible is that to produce it, we would have to first combine un– and forgive into unforgive, and that is not a word of English.

Syntax

Syntax is concerned with how words are combined to make sentences. We’ve already cited examples in sections 2.1 and 2.2 above that establish key properties we’ve ascribed to language. First, syntax showed us that language learning is more than memorization. Second, it established that knowledge of language is knowledge of the generalizations that underlie what is well-formed in the language. Third, it established that what is well-formed in a language is not determined by what “makes sense”.

A remaining essential point is that words are not combined in a linear fashion. Rather, as with morphology, words are combined in a nested fashion. Consider, for example, a sentence like the following:
(2.23) Ernie saw the man with the binoculars.

There are two possible meanings for this sentence. First, it could be that Ernie used binoculars to see the man. Alternatively, the man has binoculars and Ernie saw him. These two meanings are based on two different structures. The first has the man and with the binoculars as sisters within a larger grouping, labeled VP here for Verb Phrase.

(2.24)

Ernie  
  |  
  VP  
    |  
    saw  
      |  
      NP  
        |  
        the man with the binoculars

The other structure groups with the binoculars directly within the same phrase as the man, labeled NP here for Noun Phrase.

(2.25)

Ernie  
  |  
  VP  
    |  
    saw  
      |  
      NP  
        |  
        the man with the binoculars

In the first structure, with the binoculars modifies the verb; in the second structure, it modifies the noun.
CHAPTER 2. LANGUAGE

The point is that words are grouped together into structures and those structures contribute to the meaning of sentences. Syntactic competence includes knowledge of what structures are possible in the formation of sentences.

Semantics

Semantics is concerned with the meanings of words and sentences. One kind of semantics deals with how the meaning of a sentence is computed from its words and the way those words are grouped together. As we have already shown, the groupings can make dramatic contributions to the meaning of a sentence or of a word, as in the binocular example above. We also considered examples where there are multiple meaning differences that are quite subtle, but are not obviously associated with structural differences: the example treated in chapter 1 of Two women saw two men.

We will see that logic, treated in chapters 4 and 5, is quite useful in semantics.

2.3.2 Other Areas

There are a number of other really interesting areas of language study and we list just a few of them here (in no particular order).

Psycholinguistics is the study of performance, how language is actually used. The methodology is typically experimental. Psycholinguists study language production, language comprehension, speech perception, lexical access, etc.

Neurolinguistics is concerned with language in the brain. At the theoretical level, it focuses on how language is processed in actual brain structures. At a more applied level, it deals with various sorts of cognitive disorders involving language.³

Sociolinguistics is concerned with the relationship between language and society. It is concerned with how social forces are reflected in language and with how language affects social variables.

³There is something called Neurolinguistic Programming, but this is a misnomer. This is not a field of language study. Occasionally, one sees the abbreviation NLP for this field. NLP is also used for Natural Language Processing, which is a field of language study.
Language Acquisition One can study how language is acquired. This is done in several ways: following the development of individual children or running experiments on sets of children.

Writing Systems can be studied for their own intrinsic interest, but also as a window into the structure of a language or as evidence for how language changes over time.

Literature also provides an interesting vantage point on language. Again, it can be studied for its own intrinsic interest, but also for the information it provides about the structure of language.

Applied Linguistics refers to several different fields, all of which use the study of language in some concrete application. These include, for example, language teaching and forensic linguistics.

Computational Linguistics includes two broad domains. One is the use of language in computational tasks, e.g. machine translation, text mining, speech understanding, speech synthesis, etc. The other area is the mathematical modeling of language. Many of the foundational areas we treat in this book are part of this latter area.

Discourse Analysis is the investigation of how units larger than a sentence are constructed. This can include conversation or texts. The latter can bring this area close to the study of literature.

Historical Linguistics is the investigation of language change. This can be done by comparing modern languages, looking at historical records, or looking at the language-internal residue of historical changes.

2.4 Summary

This chapter has introduced the study of language generally. We began with a refutation of some of the more egregious misconceptions about language.

We then established a number of key properties of language. These include a definition of language as an expressively complete conventionalized system of communication. We also argued at several points that the set of words and sentences is infinite and that knowledge of language is more than just knowing what those words and sentences are.
Finally, we reviewed the main areas of language structure and many of the areas of language study.

2.5 Exercises

1. We’ve seen that languages are sometimes characterized in almost anthropomorphic terms, e.g. “harsh”, “beautiful”, “logical”, etc. Explain why.

2. We’ve seen that sometimes a concept that is encoded with a word in one language is expressed with a phrase in another language. Why do you think this happens?

3. Does the difference between competence and performance apply to written language? Explain why or why not.

4. We’ve cited several structures in English that show that the set of words and the set of sentences are infinite. Find another that’s different from the ones cited in the text.

5. Find a description of how the “nothing” construction works in some language we have not discussed. Describe the pattern and provide some examples. Does the language have double negation?