4 Overview

Language as a conversion between thought and sound

Let's return to the basic communicative act sketched in Chapter 2—one person saying something to another. To begin with, I'll simplify it even more.

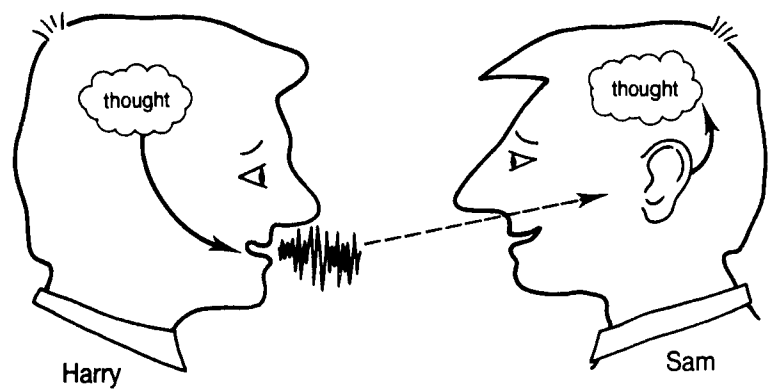


Figure 4.1 *A speech event*

Something in Harry's brain that we might as well call a "thought" results in movements of his vocal tract (lungs, vocal cords, tongue, jaw, and lips), which in turn create a sound wave that is transmitted through the air. This sound wave, striking Sam's ear, results in Sam's having the same "thought" (or a similar one) in his brain.

Of course the brain doesn't move the vocal tract by magic. It employs its usual means of bringing about motor activity: neural activation of the muscles. In other words, Harry's brain creates

patterns of neural firings that drive the muscles of his vocal tract. His vocal tract, by being so activated, creates and broadcasts the acoustic patterns that we hear as speech.

Noises going into Sam's ears don't create thoughts in his brain by magic either. Rather, the inner ear turns acoustic waveforms into patterns of neural firings that are transmitted to the brain.

Consider now the patterns of neural firings that activate the vocal tract, and those that reach the brain from the ears. Neither of these patterns can in itself be the thought transmitted by the speech event. In case this isn't ridiculously obvious, here are three reasons why.

1. The form of the thought must be neutral between spoken and heard language—we must be able both to speak the thought and to hear it—so it had better not be a pattern of neural firings that pertains specifically to one or the other of them.
2. One can "have a thought" without choosing to speak it, so the thought had better not directly drive the vocal tract. Similarly, one can "have a thought" without having heard anyone else speak it, so thought can't be driven directly by auditory neurons.
3. Thought must be (at least mostly) independent of what language it is spoken in. The patterns of motor neuron firings that produce the noises *Le chien est mort*, *Der Hund ist tot*, and *The dog is dead* are not at all alike; and since these sounds are quite different, so are the patterns of auditory neuron firings they evoke. Yet they all convey (pretty close to) the same thought to speakers of French, German, and English respectively. In order for us to be able to translate among languages, there must be a constancy in the thought being expressed that isn't present in the motor or auditory patterns. In short, we have to identify the thought with a pattern of neural firings that is distinct from both the motor and auditory patterns.

(By the way, what can it mean to say a thought is a pattern of neural firings? If you find this idea troublesome, please bear with me until Chapter 14, when we confront this issue a little more closely.)

This conclusion requires us to add a further piece to the picture: the brain has to have a way to convert the patterns of neural firings

that constitute the thought into the patterns that drive the vocal tract, plus a way to convert the patterns of firings produced by the inner ear into those that constitute thought. Figure 4.2 sums up our analysis so far. (The heads in Figure 4.1 are cute but a nuisance to draw, so I'll schematize a little: the solid line indicates the boundary between the person's body and the outside world; the dotted line indicates the boundary between the brain and the rest of the body.)

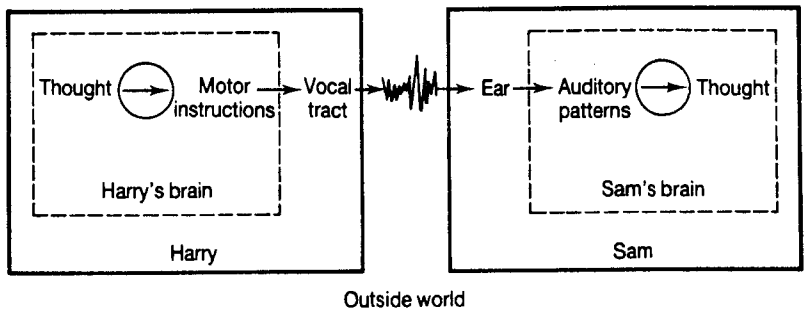


Figure 4.2 *Conversion of thought into sound waves and back again*

We now begin to home in on where language fits in: it is the brain's means of translating in a principled way between thoughts on one hand and auditory and motor patterns on the other. In other words, we can locate it within the circled arrows in Figure 4.2. (Of course, Figure 4.2 is incomplete, since each person has the capability of performing both translations. We'll fix that in a moment.) The process of going from thought to motor instructions is *speech production*; that from auditory patterns to thought is *speech perception*. We can think of different languages as different ways of converting thought into motor patterns and auditory patterns back into thought.

One of the basic insights of linguistic theory is that these translations or conversions are not carried out by the brain in one fell swoop. Rather, two major intermediate steps are involved in the conversion: *phonological structure* (or sound structure) and *syntactic structure* (or phrase structure). Zooming in more closely on Harry's and Sam's brains in Figure 4.2, the general configuration looks like Figure 4.3: the circled arrows of Figure 4.2 have been elaborated to include these intermediate steps of conversion.

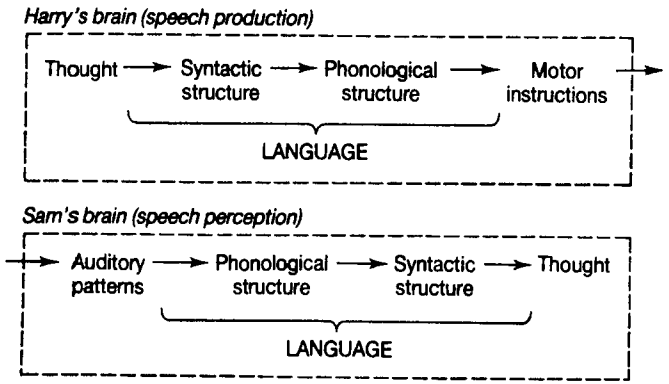


Figure 4.3 *The place of language in the conversion of thought into sound waves and back again*

Notice that most parts of these diagrams are shared. We can therefore combine them, and portray the brain of a single person who is able both to produce and to perceive language, simply by allowing some of the conversions to go in either direction, as in Figure 4.4.

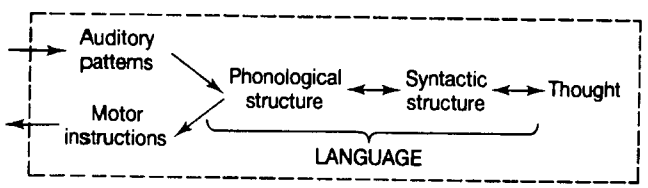


Figure 4.4 *Information flow in the brain of a combined speaker-hearer*

How are we to understand Figure 4.4? To help get a sense of it, think of what goes on when you play a videotape on your VCR. The videotape contains information stored as patterns of magnetization on the oxide of the tape; its organization is essentially linear and spatial. The VCR playing the tape translates this organization into a temporal pattern of electrical impulses that it feeds into the television set. In turn, the television set translates this pattern of impulses into a pattern of dots on the screen, arrayed in two dimensions in space plus one dimension in time.

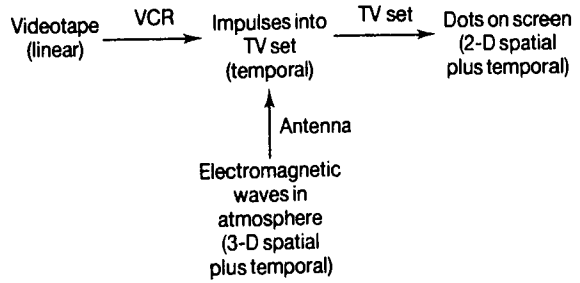


Figure 4.5 *Information flow in a VCR–antenna–TV configuration*

Each of the information formats—videotape, electrical impulses, and dots on the screen—has its own intrinsic organization; a specialized device is necessary to turn each of these types of information into one of the others.

Figure 4.5 shows another format of information relevant to this system. TV broadcasts involve a signal being transmitted in the form of an electromagnetic wave, which is three-dimensional and temporal in its organization. The antenna is a device that converts this signal into one of the formats already mentioned: the electrical impulses entering the TV set. Videodiscs are yet another format for information, requiring yet another specialized device to convert their patterns into inputs for the TV.

Finally, we should remember that there exist several different formats for videotape—VHS and Betamax, for example. Each of these needs a slightly different device to transform it into properly coded impulses for the television set. (One could think of these as different "languages" or "dialects" of videotape.) Similarly, the patterns of dots on the screen can be organized differently, depending on whether one is dealing with black-and-white or color television, or with ordinary versus high-resolution video.

Going back to Figure 4.4, a similar story emerges, except that each of the formats of information consists of patterns of neural firings. Again, each format has its own characteristics, and specialized devices—different parts of the brain this time—are required to convert the information from one format to the next. Again, each pairing of formats requires a different conversion device. For example, converting auditory patterns into phonological structure is quite a different process from converting phonological structure into motor patterns, even though the two processes share phonological structure as a common format of information. And different

conversion procedures are needed if the phonological and syntactic structures are those of Vietnamese or Hopi instead of English.

We could add yet another component to Figure 4.4: the relationship to written language. Alphabetic written language is basically an encoding of phonological structure—one learns to "sound out words"—though of course the encoding is not perfect, and English spelling is especially notorious for its idiosyncrasy. In order for such an encoding to be useful to us, we must have another conversion process that gets information into the format of phonological structure, through the visual system rather than the auditory system—namely, reading. That is, reading can be seen as an additional branching in Figure 4.4, analogous to the antenna in Figure 4.5.*

It is worth pointing out an important psychological distinction between the reading/writing system and the speaking/hearing system. Acquiring the ability to read and write usually requires extensive instruction and practice, and not everyone achieves this ability; by contrast, every normal person learns to speak fluently. This helps to underline the important point in Chapter 3 that spoken language is not explicitly taught: compare the relatively laborious task of teaching reading with the effortless task of letting children learn to talk. The contrast highlights how special the learning of spoken language is.

The most prominent issues of modern linguistic theory concern the organization of phonological and syntactic structure, the mental codes that serve as way-stations between thought and the auditory and motor codes. Not surprisingly, these structures are far more complex than the ones employed in the VCR–TV system of Figure 4.5. The next two chapters will present a very sketchy idea of what these structures are like.

Functionalism

But before discussing phonological and syntactic structures, I had better say a word about how we manage to study them, even though we don't know very much about how they are physically encoded in the brain as patterns of neural firings. This general approach to studying mental capacities goes under the name of *functionalism*, and

* Nonalphabetic writing systems such as Chinese characters also require a conversion process whose input is the visual system, but the output of the process is possibly syntactic instead of phonological structure.

it is a leading strategy in much of cognitive psychology and artificial intelligence as well as linguistics.

To help us see what functionalism is about, let's think about the videotape again. In order to store TV pictures, a videotape must carry a code that expresses certain distinctions, and this code must be stored in terms of basically one-dimensional patterns of magnetization on a tape. So we can ask how the code could be organized so that the videotape can do its job. As a pattern, it doesn't matter too much whether we put the code on a magnetic tape or on something else comparably one-dimensional, say a punched paper tape or a barcode that can be read by an optical scanner: it's the pattern that counts. Similarly, we can study the patterning of speech sounds—their order, the differences and similarities among them, and their contributions to understanding—to a certain degree independently of the neural medium in which they are physically encoded.

In a functionalist theory, what does it mean to say we have a certain principle in our mental grammars, as part of the equipment we bring to understanding and creating novel sentences of English? Let's take a very simple principle of English—for instance, that the subject of a sentence (normally) precedes the verb. If this principle is somehow in our heads, then the terms "sentence," "subject," and "verb" must be too. What does it mean to say we unconsciously know and use these terms?

Well, for one thing, it doesn't mean that we know them *in English*! It's in the nature of the communicative situation that we as theorists have to state the principles in English, or in Japanese, or, if we want to get fancy, in terms of some mathematical formalism. That's the only way we have of stating theories about anything. But if we're developing a theory of physics and we happen to use the symbol *e* for an electron, it's not as though we expect actually to find electrons with little *e*'s on them. Similarly, we don't expect to look in the brain and find the sentence "Amy loves Uncle Sparky" stored with a little sign on "loves" that says "verb." In fact, we don't even expect to look in the brain and recognize the word "loves."

Suppose we think of our stored knowledge as the contents of some curious sort of filing cabinet in the brain. The information in the filing cabinet isn't stored in a form readable by us outside observers. Why should it be? It's not there for the benefit of outside observers, it's there for the use of the rest of the brain.

So when we state a rule of mental grammar, we're doing something like this: We use a term like "verb" to distinguish some class of words from everything else. The claim is that, whatever way

the brain uses to store words, it has a way of distinguishing these particular words from everything else. We use a term like "sentence" to distinguish a particular class of word *sequences* from everything else; the claim is that the brain makes the same distinction. We use terms like "subject" and "object" to pick out word sequences that are particular *parts* of sentences; the claim is that the brain can pick out the same parts. Finally, the whole condition, "the subject of a sentence precedes the verb," states a relation among various parts of the sentence; the claim is that the brain—however it identifies and stores these parts—imposes the same relation.

We can't find out if our claim is right by going in and looking at the brain, because we don't know how. However, if the claim is right, it also has certain consequences for how the brain is going to regard certain sequences of words—and we can test these consequences by doing experiments.

What is a linguistic experiment? As in other sciences, the strategy is to study unobservable phenomena by relating them to things that *are* observable. If we want to measure the mass of an electron or the sun, we can't just weigh them on a scale. We have to use some sort of indirect means to get at what we want to measure—we have to think of something else we *can* measure that is connected to what we really want to know, in what we think is a reliable way. The same is true with the mental principles behind language. The only difference is that linguistic experiments have to do with the inside of our heads instead of external objects.

It turns out that among the kinds of experiments that can be done on language, one kind is very simple, reliable, and cheap: simply present native speakers of a language with a sentence or phrase, and ask them to judge whether or not it is grammatical in their language, or whether it can have some particular meaning. In fact, we have already done a number of these experiments in the course of the previous chapters. I presented various strings of words such as "Harry thinks Beth is a genius" and "Amy nine ate peanuts," and I judged whether they were or were not possible sentences of English. If all went well, you had no trouble agreeing with my judgments. That's all there is to it. The idea is that although we can't observe the mental grammar of English itself, we *can* observe the judgments of grammaticality and meaning that are produced by using it.

This experiment is so simple that you may have hardly even noticed it as such. But it isn't so different from other experiments that study what's going on in the head—for instance, such well known visual phenomena as the Müller-Lyer illusion, the duck–rabbit, the

Necker cube, and various "impossible figures." In the Müller-Lyer illusion, shown in Figure 4.6, the horizontal line on the left looks shorter than the one on the right. But if we measure them with a ruler, they turn out to be the same length.

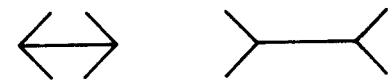


Figure 4.6 *The Müller-Lyer illusion: The two horizontal lines don't look the same length*

Figure 4.7 is the duck–rabbit, which can look alternately like a duck or a rabbit, and as you look at it, it can periodically switch back and forth.

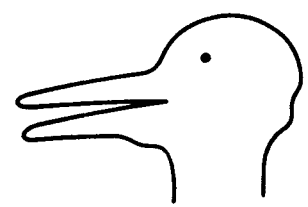


Figure 4.7 *The duck–rabbit ambiguity*

Figure 4.8 is the Necker cube, which looks alternately like a wire cube seen from the right and above or one seen from the left and below.

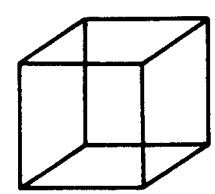


Figure 4.8 *A Necker cube*

Figure 4.9 presents three "impossible figures"—line drawings that can't be interpreted as drawings of actual objects.

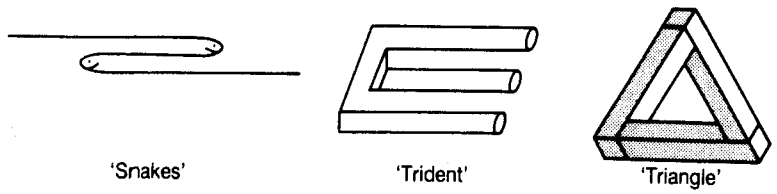


Figure 4.9 *Three impossible figures*

These experiments are so simple and reliable that all we have to do is present them to observers and ask them what they see. Moreover, it is clear that our judgments of these figures have nothing to do with what we were "taught about seeing"; that these judgments require no conscious thought; and that at the same time it's very hard to be explicit about why the figures look the way they do. That is, these visual judgments have all the same symptoms as judgments about sentences. I'm suggesting, then, that the two kinds of judgments have similar status as experimental evidence.

A linguistic example very much like Figures 4.7 and 4.8 is the ambiguous sentence (1a). On one interpretation, your relatives are visiting you, and the sentence means about the same as (1b). On the other, you are visiting your relatives, and the sentence means about the same as (1c).

(1) *a* Visiting relatives can be boring.
b Relatives who are visiting you can be boring.
c Going to visit relatives can be boring.

Parallel to the impossible visual examples in Figure 4.9 are all the ungrammatical sentences of the last chapter, such as "An oboe not is octopus an." We know immediately, intuitively, that there is something the matter with them. ("Intuitive" judgments are just judgments that follow from unconscious principles. We make the judgment but can't say exactly why.)

Ideally, we might want to check these experiments out by asking large numbers of people under controlled circumstances, and so forth. But in fact the method is so reliable that, for a very good first approximation, linguists tend to trust their own judgments and those of their colleagues. So in a couple of hours it is possible to dream up and perform dozens of these low-tech experiments.

On the other hand, there are cases where these experiments aren't so simple. The most obvious is when you are working on some language other than your own. You then have to find an informant, a native speaker of Turkish or Kwakiutl or whatever, and ask him or her for judgments. In this situation it's much harder to make up sentences to check, and harder to guess what to try next; the process is trickier, but it can be done. (It works both ways. I once spent an entire Kyoto-to-Tokyo train trip being asked by a Japanese linguist for judgments about hundreds of English sentences.)

Another problematic situation arises when the judgments of crucial sentences aren't so crystal-clear and reliable. For instance, a certain line of recent research on so-called "long-distance dependencies" (see Chapter 6) depends on judgments of sentences like those in (2).

(2) *a* What did he wonder whether to fix?
b These are the only vegetables which I don't know where to find out how to plant.

Here I'm not so sure what your judgments are going to be like. To me, the sentences are pretty awkward. On the other hand, they sound better than the slightly different sentences in (3), which I find horrible.

(3) *a* What did he meet a man who can fix?
b These are the only vegetables which I don't know anyone who planted last year.

In such cases the researcher tries to proceed with sensitivity, perhaps trying more possibilities and consulting more people before making a judgment on what the mental grammar is telling us about the sentences.

Of course, other kinds of experiments can be used to explore properties of the mental grammar, including computer simulations, various reaction-time procedures borrowed from experimental psychology, and even measurement of brain waves during language processing. (People who do these kinds of experiments are called *psycholinguists*.) Often these other procedures provide crucial evidence. Their disadvantage is their relative inefficiency: it takes a great deal of time and energy to set up the experiment. By contrast, when the experiment consists of making judgments of grammaticality, there is nothing simpler than devising and judging some more sentences. (Unless, of course, you're working on Inuit and have to go back up to Hudson Bay to see your informant.) So grammaticality judgments

remain the most widely used experimental technique in contemporary linguistics.

You have to bear in mind, of course, that it is possible to make up and judge sentences from now until doomsday without getting any closer to understanding mental grammar. But the same is true in any experimental science: experiments are worthwhile only if they help us to figure out what's going on.

Drawing this all together: the functionalist approach to mental grammar is to make experimentally testable hypotheses about the organization of information and knowledge in the brain, without too much concern for the moment about how the brain physically encodes this information.

Some researchers have gone even further and suggested that functionalist research should divorce itself from *any* consideration of how the brain actually functions: we should be interested in intelligence in the abstract, in organizations of information that could be equally well embodied in computers or in Martians. From this point of view, it is essentially beside the point how the human brain in particular encodes language.

I would rather not buy into such an extreme version of functionalism. I am interested in how *we* work. If someday someone shows that the neural circuitry of the brain is capable of encoding certain kinds of patterns and not others, I will be damned sure that this constrains my hypotheses about the principles of language. The problem is hard; I need all the help I can get. Conversely, I would hope that an understanding of the functional organization of language would inform research into how the brain encodes information: however the brain works, it must be able to encode information with *these* sorts of patterns.

The Modularity Hypothesis

The organization of language sketched in Figure 4.4 already illustrates an important hypothesis of modern cognitive science and neuroscience: the differentiation of the brain. Think again of the VCR–TV system. Each translation from one format into another requires a specialized device. We need a VCR to translate from videotape into electrical impulses; we can't use the circuitry of a TV set or a computer printer. And it can't be just any old VCR, it has to be one that can read the kind of videotape we happen to have. Likewise, if the production and perception of language require several

complex specialized codes, the brain must include complex specialized devices to deal with each of them, and with the translations among them.

This hypothesis has been called the *Modularity Hypothesis*: the idea that the brain is divided into many separate units or *modules*, each with the capacity to deal with a specialized kind of information. It is not just the way the modules are connected up with each other that makes the brain function as it does. The nature of the circuitry inside each module is absolutely crucial, for it is the particular circuitry that makes a module act as a phonology processor or a visual shape identifier or a finger wiggler or whatever.

Notice how the Modularity Hypothesis connects up with the Genetic Hypothesis of Chapter 3—the idea that a great deal of the structure of language is transmitted genetically, through the inheritance of brain structure. The more the brain consists of specialized parts, the less likely it is that these parts are acquired through learning, especially learning of a simple stimulus–response nature. The brain thus comes to look a lot like the rest of the body, with exquisitely specialized complex physical structure.

To repeat Chomsky's point from Chapter 3: we aren't tempted to think that, by trying hard enough to fly, we can learn to have wings rather than arms. We can strengthen our arms and learn to use them in complex tasks such as doing gymnastics or playing the cello, but we're not going to be able to alter their basic organization. Likewise, under the Modularity Hypothesis, we shouldn't believe that our brains can develop entirely new specialized parts in response to new and unusual tasks. We can strengthen and refine the use of the parts that are there—say, by learning mathematics or chess or real estate law—but the basic functional organization can't change. (It *can* change, of course, through evolution. Just as birds evolved a structural specialization of the forelimbs for flying, humans evolved a structural specialization of the brain for learning language.)

This way of looking at learning is even more radical than the view developed in Chapter 3. Language perception and production require specialized devices for processing information in different formats and for translating information from one format to another. We're now saying that in learning language, children don't build these specialized devices from scratch. Rather, they just "tune up" or strengthen or adjust devices that are already present by virtue of biological structure. (And a speaker of two languages may be thought of as having the ability to "change channels" within the same specialized device.) Under this view, our solution to the Paradox of

Language Acquisition is perfectly natural. Compare the problem of adjusting your TV to the problem of inventing it (or figuring out how it works). The former is more like the child's task; the latter more like the linguist's. No wonder there's a disparity.

In Part IV we will come back to the Modularity Hypothesis in the context of abilities other than language. Right now, our task is to get a sense of (1) the complexity and specialization of the codes employed in speech production and perception and (2) the relative contribution made to these codes by Universal Grammar (or prebuilt structure) and by learning (or tuning). This will give us an overall idea of the degree to which the brain is specialized for language, and it will set a standard by which we can judge the plausibility of the Modularity Hypothesis for other domains of human nature.

5 Phonological structure

We now delve into some of the actual content of the organization of language. This chapter deals with phonological structure, one of the two intermediate steps of conversion between thought and sound. The next chapter deals with the other, syntactic structure. The structure of thought (or meaning) is more controversial, and I have put it off until Chapter 14.

Phonological structure is neither an auditory nor a motor pattern

One of the primary intuitions we have about language is that it comes divided into words, and that the words can be neatly divided into syllables and individual speech sounds. The phonological structure of language is an encoding of this sequence of sounds. It turns out that this sequence is a considerable abstraction of what physically takes place in speech. The acoustic stream we hear as speech shows no such neat divisions. To understand why, it's useful to see how speech is produced.

For a convenient analogy, think about how a trumpet works. The player presses his or her lips together and forces a stream of air through them, producing a sort of buzzing sound. When the trumpet is placed against the vibrating lips, the air column in the trumpet is forced into vibration as well. The way the air column vibrates is a function of the vibration of the lips interacting with the resonant frequencies of the tube; the tone quality we hear has to do with which natural harmonics of the tube are produced in what proportions. If we change the tightness of the lips, the pitch of the whole system changes; if we place a mute in the trumpet's bell, we change the mix of resonant frequencies of the tube, and so the tone quality changes.

Now let's imagine a trumpet whose tube is made out of rubber instead of brass, so it can be stretched and pinched in various ways.

Language Acquisition is perfectly natural. Compare the problem of adjusting your TV to the problem of inventing it (or figuring out how it works). The former is more like the child's task; the latter more like the linguist's. No wonder there's a disparity.

In Part IV we will come back to the Modularity Hypothesis in the context of abilities other than language. Right now, our task is to get a sense of (1) the complexity and specialization of the codes employed in speech production and perception and (2) the relative contribution made to these codes by Universal Grammar (or prebuilt structure) and by learning (or tuning). This will give us an overall idea of the degree to which the brain is specialized for language, and it will set a standard by which we can judge the plausibility of the Modularity Hypothesis for other domains of human nature.

5 Phonological structure

We now delve into some of the actual content of the organization of language. This chapter deals with phonological structure, one of the two intermediate steps of conversion between thought and sound. The next chapter deals with the other, syntactic structure. The structure of thought (or meaning) is more controversial, and I have put it off until Chapter 14.

Phonological structure is neither an auditory nor a motor pattern

One of the primary intuitions we have about language is that it comes divided into words, and that the words can be neatly divided into syllables and individual speech sounds. The phonological structure of language is an encoding of this sequence of sounds. It turns out that this sequence is a considerable abstraction of what physically takes place in speech. The acoustic stream we hear as speech shows no such neat divisions. To understand why, it's useful to see how speech is produced.

For a convenient analogy, think about how a trumpet works. The player presses his or her lips together and forces a stream of air through them, producing a sort of buzzing sound. When the trumpet is placed against the vibrating lips, the air column in the trumpet is forced into vibration as well. The way the air column vibrates is a function of the vibration of the lips interacting with the resonant frequencies of the tube; the tone quality we hear has to do with which natural harmonics of the tube are produced in what proportions. If we change the tightness of the lips, the pitch of the whole system changes; if we place a mute in the trumpet's bell, we change the mix of resonant frequencies of the tube, and so the tone quality changes.

Now let's imagine a trumpet whose tube is made out of rubber instead of brass, so it can be stretched and pinched in various ways.

Let's also imagine that a second tube branches off somewhere in the middle, and that air can be directed out of either tube or both. Such a horn probably won't have the clear resonances of a regular trumpet, but it will be able to produce a much greater variety of tone colors, because its resonant frequencies can be altered so much.

Next let's shove this trumpet down the player's throat. Instead of the lips, the vocal cords down in the larynx set the tube into vibration. The tube consists of the throat and its branches into the oral and nasal cavities. The nasal cavity can be disconnected by raising the velum (or soft palate), which functions as a valve. The oral cavity can be constricted or closed off by closing the lips or by raising different parts of the tongue. The sound produced is a function of the vibration of the vocal cords coupled with the very complicated resonances of this tube. As the muscles of the vocal tract change the shape of the tube, the resonances change correspondingly, and these differences are perceived as different speech sounds.

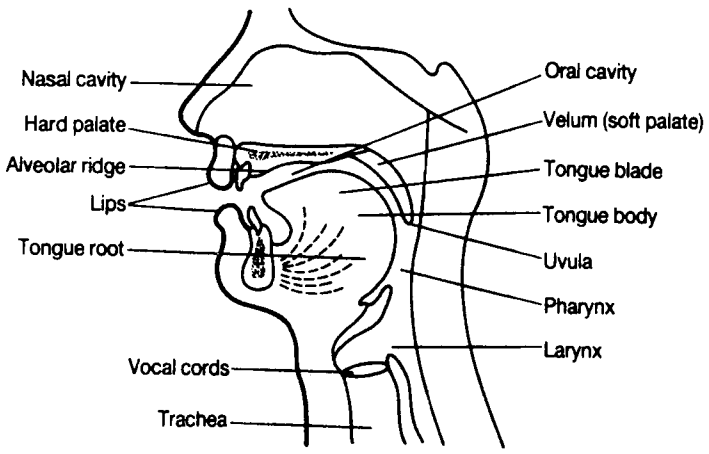


Figure 5.1 *The vocal tract*

During speech, the movements of the vocal tract are smooth and continuous. For instance, try saying the word *wow*, and pay attention to what your mouth is doing. You don't hold a *w* (or *oo*) sound, with pursed lips, then suddenly switch to an *ah* sound, with mouth and lips open, then instantaneously switch back to the *w*. Trying say it that way: *oo-ah-oo*, with sharp transitions between the sounds. You can hear how unnatural it is. In the normal pronunciation of the word, your lips

and mouth open smoothly and then close smoothly, with virtually no holding of either the closed *w* or the open *ah* sound.

As a result of the smooth transition between positions of the vocal tract, the acoustic signal produced by the vocal tract also shows a smooth transition from one sound to the next, without any abrupt boundaries. Consequently, the signal that the hearer *perceives* as neatly divided speech sounds is actually far from it. The waveform undergoes continuous change as a result of continuous change in the shape of the vocal tract.

But then what *do* the separate speech sounds correspond to? As it turns out, they correspond rather closely to beginning and end configurations of vocal tract movements—for instance, in the word *wow*, the sequence lip-pursing followed by open mouth followed by lip-pursing. This sequence is obviously different from the auditory information. But it is also different from the motor instructions driving the vocal tract: ultimately, the vocal tract must be told not what sequence of positions to attain, but what muscles to tense and relax in order to obtain that sequence.

Nevertheless, in order to structure speech, it is necessary for the brain to code the sequence of speech sounds and their combinations into words and sentences. This code of speech sounds is referred to as *phonological structure*. In this section we have seen that phonological structure is a distinct kind of mental organization from either vocal tract instructions or auditory patterns (see Figure 5.2).

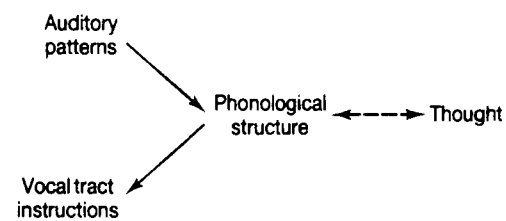


Figure 5.2 *The place of phonological structure in the information flow of language*

When speaking, then, one has a phonological structure (as well as a thought) in mind. The phonological structure specifies a sequence of vocal tract configurations, and the brain must convert this sequence into instructions that tell the muscles of the vocal tract how to move. When *hearing* speech, the brain must convert the continuous, smeary information coming from the auditory nerve into

such a sequence of vocal tract positions—in effect reconstructing configurations of the speaker's vocal tract, that is, perceiving the phonological structure that the speaker has in mind.

Psychological reality vs. physical reality of phonological structure

Alert readers may be getting slightly uneasy at this point. At the outset of this chapter I observed that we intuitively sense utterances as divided into words, and words as divided into syllables and individual speech sounds. Yet we've spent the last few pages showing that the acoustic signal out in the world shows no such neat divisions—that the actual sounds of speech undergo continuous change, so that each one smears into the next.

To make this more vivid, think about the divisions between words. In writing, of course, we leave spaces between words, and this mirrors our perception: we (almost) never have trouble hearing where one word ends and the next begins. But listen to yourself say the following pairs of sentences:

(1) *a* I don't really think it's a parent.
b I don't really think it's apparent.
(2) *a* Have you looked at this guy yet?
b Have you looked at the sky yet?
(3) *a* We needed a cantor.
b We need a decanter.*

At a normal conversational rate, these sentences can be spoken in such a way that they are acoustically indistinguishable. Yet they require the breaks between words to be placed at different points in the sequence of sounds. More generally, you—don't—leave—an—audible—space—after—each—word—when—speaking. Youjustjamthewordstogetherwithoutpauses. In other words, the boundaries between words, though undeniably part of phonological structure, are not present in either the motor instructions or the auditory information.

In case these examples from English don't convince you, think about the last time you heard people speaking a language you didn't know. Could you tell where one word ended and the next began?

* This is a real-life example. During my friend Michael Bennett's wedding ceremony, the rabbi was having trouble pouring the wine, and Michael quipped, "We have a rabbi, but we . . ."

There were pauses here and there, but for the most part it probably sounded like just a great rush. This shows that you have to know the language in order to be able to perceive the word boundaries.

A profound point is lurking here. We have discovered that the words we consciously hear and pronounce are not in any physical sense "out in the world." We can't find them through physical measurement of acoustic waveforms. Rather, the way we experience the stream of language seems to have more to do with the patterns of phonological structure in our heads, in which the speech sounds and words are clearly demarcated.

I won't stress this observation too much for now. But we will come back to it in Chapter 12, where we will ask: Are there more general implications about the relation between the physical world and what we perceive? This question will serve as the starting point for the third Fundamental Argument, the Argument for the Construction of Experience.

The internal structure of speech sounds

We now go back to speech sounds themselves. Recall that phonological structure encodes speech sounds as a sequence of vocal tract configurations—successive positions of the larynx, jaw, lips, tongue, and velum. How does the brain specify these configurations, and how are they stored in memory?

Yet again we have to break with intuition. Intuitively, speech sounds are unitary: the sound *z* is just that, a single sound. However, one of the major discoveries of phonological theory, originally developed by Nicholas Trubetzkoy and Roman Jakobson in the 1920s and 1930s, is that speech sounds are encoded in the brain in terms of more primitive specifications called the *distinctive features* of speech sounds. Some of these features are listed in (4). If you pay very careful attention to what is going on in your mouth and throat as you make different sounds, you can verify the features for yourself.

(4) *a* Significant constriction of the vocal tract (consonants)
vs.
Vocal tract unconstricted (vowels)
b Vocal cords tensed and therefore in vibration (sounds such as *b*, *g*, *d*, *z*, *n*, *th* as in "there," all vowels)
vs.
Vocal cords relaxed and therefore not vibrating (*p*, *t*, *k*, *ch*, *f*, *sh*, *th* as in "thin")

c Velum lowered so air passes through nose (*m*, *n*, *ng* as in "sing," vowels in French "on" and "fin")
vs.
Velum raised so air passes only through mouth (all other sounds)

d Air flow through mouth completely blocked (*p*, *b*, *m*, *t*, *d*, *n*, *k*, *g*, *ng*, *ch*, *j*)
vs.
Air flow through mouth not completely blocked (all other sounds)

e Most constricted part of mouth at lips (*p*, *b*, *m*, *f*, *v*, *w*, *u*)
vs. at tip of tongue (*t*, *d*, *n*, *s*, *z*, *l*, *th*)
vs. at body of tongue (*ch*, *sh*, *j*)
vs. at back of tongue (*k*, *g*, *ng*, German *ch* as in "Bach")

Each speech sound can be described in terms of a combination of the distinctive features. The sound *d*, for instance, codes a configuration in which the vocal tract is significantly constricted, the vocal cords are tensed, the velum is raised, air flow through the mouth is blocked, and the main constriction is at the tip of the tongue. The sound *t* differs minimally from *d* in that the vocal cords are relaxed; *n* differs minimally from *d* in that the velum is lowered so air passes through the nose. The sound *z* differs minimally from *d* in that the mouth is not completely blocked, so air continues to pass through; *b* differs minimally from *d* in that the constriction is at the lips. Thus the similarities and differences among speech sounds can be specified in terms of distinctive-feature analysis.

The relations among sounds provided by distinctive features enable us to explain many curious aspects of pronunciation. For a simple example, the plural suffix for English nouns is pronounced three different ways: as a *z* sound in words like "dogs," as an *s* sound in words like "cats," and as an *uhz* sound in words like "horses." (It is always *spelled* "s" or "es"; but I want to pay attention to how it is actually *pronounced*. For simplicity, I am disregarding irregular plurals like "oxen," "mice," and "sheep.") What decides among these three choices in any given word?

It turns out to depend on the final sound of the noun to which the plural is attached.

(5) *a* If the noun ends with one of the sounds *s*, *z*, *sh*, *ch*, *j*, the plural is pronounced *uhz*.
b If the noun ends with one of the sounds *p*, *t*, *k*, *f*, or *th* as in "death," the plural is pronounced *s*.
c If the noun ends with anything else, the plural is pronounced *z*.

These classes of sounds may look arbitrary, but in fact distinctive-feature analysis shows us the method behind the madness. The class in (5b) consists of sounds that are "unvoiced"—that is, in which the vocal cords are not vibrating. The class in (5c) consists entirely of sounds that are "voiced"—in which the vocal cords vibrate. Two of the pronunciations of the plural, *s* and *z*, are articulated identically except for this feature of voicing; we can see the pronunciation of the plural as "agreeing" in voicing with the end of the word it is attached to.

What about the class in (5a), which includes both voiced sounds (*z* and *j*) and unvoiced ones (*s*, *sh*, and *ch*)? This class includes sounds whose articulation is very close to that of the plural ending: *s* and *z* themselves plus the sounds articulated with the body of the tongue. Here a vowel sound is inserted before the plural ending to prevent interference between the two consonants—and the plural itself is pronounced *z* to agree with the voicing of this vowel.

There is a small class of words ending in *f*, such as "wolf" and "half," in whose plurals the final consonant changes to *v*: "wolves" and "halves." What has happened here? Notice that the sound *v* is articulated exactly like *f*, except that it has vocal cord vibration. So in these words, exceptionally, the last consonant of the word changes its sound slightly in the plural, becoming voiced. The plural ending is then pronounced *z* to agree with it.

Not only can the choice among the three pronunciations of the plural be predicted from the features of the final sound of the word, but the choice is *productive*—that is, we can produce plurals for new nouns on the spot. We don't just memorize the plural ending for every noun we know (though we probably *do* memorize the irregular cases). To see this, suppose I introduce new words to you and ask you what their plurals are. If you had to learn the plural form by memorization, you couldn't carry out this task. Here are some words (they happen to be borrowed from Yiddish).

(6) *a* kvetch
b dybbuk
c shmeggeggie

Whatever "kvetch" might mean, you know its plural is "kvetches," with the plural pronounced *uhz*. This follows from the fact that the last sound of the word falls into group (5a). Similarly, the plural of "dybbuk" is "dybbuks," with the plural pronounced *s*, because the word ends in an unvoiced sound; that of "shmeggeggie" is "shmeggeggies," with the plural pronounced *z*, because the word ends in a voiced sound.

Even more striking is what happens when a new word ends with a sound not present in English. "What a big family of composers! There was Johann Sebastian Bach, and Carl Philip Emmanuel Bach, and Wilhelm Friedemann Bach, and even P.D.Q. Bach. There sure were a lot of __!" Fill in the blank. The word is "Bachs," with the plural pronounced *s*. Why? Because even though the *ch* sound is not a sound of English, English speakers can tell (unconsciously or intuitively) that it is unvoiced, and that it is not made with the body of the tongue. It therefore falls in with the sounds in group (5b), whose plural is pronounced *s*. In other words, we don't even memorize the plural rule in terms of the list of sounds—we use the distinctive-feature analysis, which can extend to sounds not even present in our own language.

These observations deal with only a tiny aspect of the sound pattern of English. But they are a microcosm of the larger picture: using distinctive-feature analysis, we have been able to describe the pronunciation of the plural fairly simply, rather than as a complex list of special cases. This kind of result is replicated in case after case in many different languages, suggesting that this is indeed the way the brain encodes speech sounds.

You probably had no conscious awareness of distinctive features before I pointed them out a moment ago. The conscious awareness of the sound pattern of language extends as far down as individual speech sounds (and perhaps not that far in nonliterate societies), but it does not include the further analysis of the speech sounds themselves. Nor do we have any conscious awareness of principles such as the one governing the pronunciation of the plural; yet everyone adheres to them. Thus the features and the way they are put to use in language had to be discovered by experiment. In fact, in our discussion above, we have replicated some of the simpler experiments in this domain. More generally, I would consider the discovery of distinctive features, and the continual refinement of their formulation over some decades, to be a scientific achievement on the order of the discovery and verification of the periodic table in chemistry.

There is a great deal more to phonological structure than just distinctive features. In addition to a sequence of speech sounds, words and phrases carry stress, or relative emphasis among their syllables. In many languages (such as Chinese, Vietnamese, and many West African languages), words carry with them an inherent melody (or sequence of tones) that is as much a part of the word as its speech sounds. Words and phrases also have an inherent rhythm that determines their temporal flow. For each of these aspects of the sound structure of language there is a growing literature of analysis and theory, with much lively dispute. I have tried to illustrate here only the basic spirit of the analysis.

The Paradox again

If the system of distinctive features is such a great scientific discovery, what does this say about the child's learning of language? How did we as children learn the principles of the English plural, including its application to "shmeggeggie" and "Bach"? We had to figure out at least three things.

1. We had to notice that describing a multiplicity of objects is correlated with changes in the way the names of the objects are pronounced. What in the world do three cats, five apples, and fourteen clouds have in common? Not much other than multiplicity (or plurality). Yet this really very abstract notion is correlated with the use of some special noise tacked onto the names for the animal, the fruit, and the meteorological phenomenon respectively. This has to do with meaning, so it's not, strictly speaking, within the ambit of this chapter. But it *is* part of the learning process.
2. We had to guess that there is some relation among the noises tacked onto the different names for objects—that is, that there is a standard way to form plurals and not just a million different ways. (Why couldn't the plural of "cat" be "cattis," the plural of "apple" be "akkle," and the plural of "cloud" be "cleez"? That *could* be memorized, though it would be a pain.) Guessing that there is a regularity is probably not specific to language, though—children try to find regularities in everything.
3. We had to figure out that the regularity in use of the sounds *z*, *s*, and *uhz* to denote multiplicity has to do with

the last sound of the words they are attached to, and in particular to the distinctive features of those sounds. And here it is necessary to invoke Universal Grammar. For in order to figure out that the pronunciation is dependent on distinctive features, the child must either figure out the distinctive features or else know them in advance. Given that linguists just figured them out sixty years ago, but children have been learning the English plural rule for hundreds of years, we have to suppose that children have access to distinctive-feature analysis of sounds in order to be able to learn to speak.

Naturally, since the child doesn't know in advance which of the languages of the world he or she is going to be learning, the system of distinctive-feature analysis had better be able to accommodate all the speech sounds of all languages, including the German *ch* sound, the French nasal vowels in "fin" and "on," and even the click sounds of the South African languages Xhosa and Zulu. Any single language chooses only a subset of these possibilities. That is, we can think of distinctive features as a part of Universal Grammar that provides a "menu" of speech sounds, plus the relationships among these sounds that come from shared features.

In learning a language, then, the child selects certain speech sounds from this menu to match those in the environment. Having selected these sounds, the child already "instinctively" or "intuitively" (i.e. unconsciously) knows how to sort them out—for example, which ones are voiced and which unvoiced, which are made with the tip of the tongue and which with the back, and so forth. It is this classification that makes it possible to figure out the principles that govern regularities of pronunciation like the English plural.

I've dwelt on the plural not because it's such a big deal in and of itself. It is one of the simpler cases of hundreds (maybe thousands) of phonological regularities that have been studied in the languages of the world, each of which presents similar problems to the child. And, as it happens, the English plural rule has played an interesting role in studies of language acquisition, which we will come back to in Part III. So it's worth concentrating on it a little bit here.

What else besides speech is in the auditory signal?

I've tried to show so far that speech is encoded in the brain as a sequence of distinctive-feature configurations. Recalling our function-

alist orientation from Chapter 4, we won't worry too much how the brain neurologically accomplishes this encoding. But it is worth remembering that contemporary neuroscience is silent on this problem. We know that speech is encoded in some particular *part* of the brain, but we have no idea of how the neurons in this part of the brain encode distinctive features and all the principles that depend on them. We'll return to this issue in Chapter 11.

Meanwhile, on to further complications. Let's return to speech perception—how the brain gets from the activation of auditory neurons to a perception of discrete speech sounds. We've already noted how smeared the auditory input is, and how this presents a severe problem for perception, which has to locate the boundaries between speech sounds and even between words. But that isn't the only problem that speech perception faces.

For one thing, everyone's voice is different. Because of differences in the size of the vocal cords and the anatomy of the vocal tract, there is considerable variation from person to person in the resonant frequencies that are crucial for defining vowel sounds. This means that speech perception can't just define a vowel in terms of some fixed set of frequencies. Rather, the frequencies have to be quickly adjusted to the overall range of frequencies we hear coming from the speaker. We obviously do this without thought.

In addition, the acoustic signal presented by someone speaking in a slow, sweet voice is quite different from that of the same person saying the same thing in a rapid angry voice or a measured ironic voice or a casual bantering voice. The possible variations are immense, and somehow the process of speech perception can ignore all that and pick out the speech sounds. Again, much more easily done than explained.

On the other hand, the brain doesn't simply regard these variations as noise that gets in the way of perceiving what the person is saying. Rather, the process of auditory perception analyzes the acoustic signal into three separate but simultaneous factors: *who* is speaking (voice recognition), *what* the speaker is saying (language perception), and *how* it is being said (the speaker's tone of voice or emotional affect). Each of these factors appears to be picked out by a separate module of the brain. For instance, we can recognize voices and tones of voice in a foreign film whose language we don't understand. We even "instinctively" (i.e. without conscious effort or intent) attribute emotional tone to acoustic events like a rumble of thunder or bird song, which in and of themselves are not meant to convey emotional tone. Moreover, each of these factors can be

disrupted by brain damage, leaving the other two intact: there are stroke victims who cannot understand language anymore but do respond to tone of voice, while others cannot recognize voices anymore but do understand what people are saying.

These extra factors in auditory perception can be added to our earlier diagram of the flow of information in language, as shown in Figure 5.3.

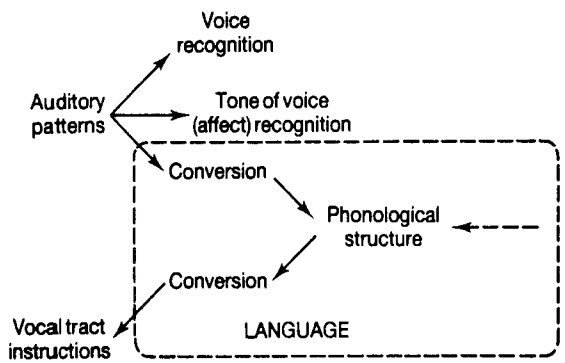


Figure 5.3 *The auditory signal feeds language and two other specialized processors*

The idea is that there are three different specialized processors awaiting the auditory signal. Each one is eager to find what it is prepared to find: the linguistic processor finds speech segments, voice recognition finds the overall mix of frequencies that identifies a speaker's voice, and affect recognition finds the variation in frequencies that characterizes tone of voice.

To help us see what this process is like, let's return to our VCR–TV assembly (Figure 4.5, p. 43), in which I ignored an important complication. The VCR feeds a unified train of electrical impulses into the TV set. But the TV set, in fact, has to sort this train into two independent factors, the picture and the sound, each of which requires its own specialized processor to convert the appropriate part of the electrical impulses into the appropriate kind of output. So it is with the auditory system, but instead of pictures and sound we get three different aspects of the same sound.

It is often said that the brain filters incoming information, attending to only a small part of what impinges on the sense organs.

This is taken to be an impressive feat. What we have found here is somewhat different: out of a unified, smeary acoustic signal, the brain derives three distinct kinds of information, at least one of which, speech, is discrete and highly complex.* Nothing goes to waste! Personally, I find this feat a great deal more impressive.

* The other two types of information may turn out to be equally complex, but so far hardly anything is known about them.

6 Syntactic structure

Syntactic structure is distinct from phonological structure

Phonological structure allows us to build up speech sounds into words and string them together. But it doesn't help us to describe the kinds of patterns we discussed in Chapter 2, shown again in (1).

(1) *a* An X is not a Y.
b Since an X is not a Y, a Z is not a W.
c X Verbs that S.

What fits into the slots marked "X," "Y," "Verb," and "S" in these patterns? We can't describe these just in terms of their sound. Rather, we need the notion, familiar from traditional grammar, of "parts of speech" such as noun, verb, adjective, and preposition—plus ways of combining them. X, Y, Z, and W in patterns (1a) and (1b) have to be filled by nouns such as "nunnery," "banana," "oboe," and so forth; *Verb* in pattern (1c) has to be filled by a verb such as "thinks," "believes," "expects," and so forth.

There is a further complication in pattern (1c). As we saw in Chapter 2, *S* has to be filled with another pattern, a sentence that can stand on its own. In the examples in Chapter 2, X was filled by a name (or proper noun) such as "Larry," "Moe," or "Curly." But it can also be filled by a larger pattern of words consisting of a common noun and a collection of modifiers—a so-called *noun phrase*. In the sentences in (2), I've underlined the noun phrase that takes the place of X in pattern (1c), and I've marked in bold the *head noun*, the noun that everything else modifies.

(2) *a* <u>The big black bear</u> thinks that you won't shoot him.
b <u>A woman in the lobby with a book under her arm</u> believes that an oboe is not an octopus.
c <u>The tall boy who Bill met yesterday</u> expects that the world economy will disintegrate within a year.

So, in order to describe X in (1c), we need a pattern made of further patterns, just as we do for *S*.

It should be fairly obvious that these syntactic categories and patterns—noun, verb, noun phrase, and so forth—can't be characterized in terms of phonological structure. For one thing, the very same sequence of sounds can serve as different parts of speech, as seen in the sentences in (3); you can doubtless multiply examples *ad libitum*.

(3) *a* We're going to **rock** around the clock. (rock = verb)
b We put some **rock** around the clock. (rock = noun)
c Beth **threw** the ball. (threw = verb)
d It went **through** the window. (through = preposition)

Consequently, if we (and our brains) are to be able to characterize the patterns of phrases and sentences, we need an additional layer of structure beyond that provided by phonological structure, a layer in which the basic units of analysis are parts of speech, and in which they are combined into phrases and sentences. Just as the auditory signal is factored into voice recognition, affect recognition, and speech perception, the speech analysis is factored into phonological structure and this further analysis, which we'll call *syntactic structure*.

On the other hand, syntactic structure can't be related as directly to the auditory signal as phonological structure is. In order to determine the part of speech of a word, first the word has to be identified. What word is being spoken obviously doesn't depend on who is saying it or their tone of voice—that all has to be filtered out already. But this is exactly what phonological structure does. That is, identifying the word depends not on its auditory characteristics but, rather, on its *phonological* characteristics. So our functional diagram of information flow in language is elaborated to Figure 6.1.

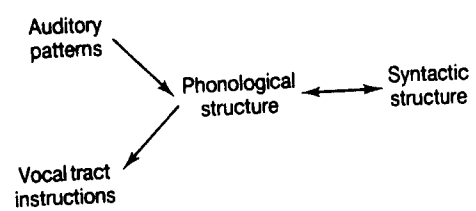


Figure 6.1 *The place of syntactic structure in the information flow of language*

Syntactic structure is distinct from meaning

School grammar defines a noun as "the name of a person, place, or thing" and a verb as "an action or state of being." According to these definitions, the units of syntactic structure are actually elements of meaning (or thought). And many people (some eminent psychologists and computer scientists included) think that syntactic analysis is nothing but a stripped-down description of meaning: if we characterize the meaning properly, there is nothing left to be said about syntactic structure.

It would indeed be nice if we didn't have to posit a level of syntactic analysis in between phonological structure and meaning—if the brain's analysis were maximally simple. But the facts of language don't let us off the hook so easily. Syntactic structure is closer to meaning than sound is—it's the last way-station enroute from sound to meaning—so it strongly reflects certain aspects of meaning. But, as I want to show, there are other properties of syntactic structure that don't have much to do with meaning. Rather, they have to do with organizing the elements of meaning into linear order so that they can be pronounced, and at the same time marking the relations among these elements so that they can be re-identified by the hearer.

To start with, let's ask whether each part of speech really denotes a consistent kind of meaning. Some of the most common matchings of entities of meaning with parts of speech are shown in (4).

(4) Object = Noun (*dog*, *skyscraper*, *ocean*, *molecule*)
Action = Verb (*breathe*, *enter*, *provide*, *interpret*)
Property = Adjective (*hot*, *jealous*, *quiet*, *insubstantial*)
Location = Preposition (or prepositional phrase) (*in the house*, *on the ceiling*, *between NY and LA*)

Now it is true that any word that names an object will be a noun. But on the other hand, not every noun names an object. "Earthquake" names, if anything, an action, as does "concert"; "redness" and "size" name properties; "place" and "location" pretty obviously name locations. In fact, for just about any kind of entity we can think of, there exist nouns that name that kind of entity. So the grammatical notion of noun can't be given a definition in terms of what kind of entity it names.

Similarly, prepositions can be used to name not only locations but also times ("after lunch," "through the night") and properties

("out of luck," "in a good mood"). So prepositions don't correspond to any fixed sort of entity either.

These examples also show that a particular kind of entity need not correspond to a single part of speech either. Actions can be named by either verbs or nouns; properties can be named by adjectives, nouns, or prepositions. In fact, the very same property can be expressed by an adjective or an adverb, depending whether it modifies a noun or a verb:

(5) *a* a **violent** earthquake, a **beautiful** concert
b The earth shook **violently**; The orchestra played **beautifully**

We conclude that parts of speech, the basic units of syntactic structure, are not definable in terms of meaning.

Here's another reason why syntactic structure isn't predictable from meaning. In Chapter 4 we noticed that meaning or thought is independent of the language that is being spoken. Otherwise it makes no sense to speak of translating from one language to another, conveying the same meaning.* It follows, then, that any difference between the original and the translation isn't part of the meaning they share.

Of course, languages don't share the *phonology* that goes with the same thought. That's why we have to study vocabulary like crazy when we're learning foreign languages: what noise means the same thing in Portuguese that "umbrella" means in English? But in addition to learning phonology, we have to learn what order to put the words in, and that's a *syntactic* property of the languages in question.

For example, English adjectives normally precede the nouns they modify, but French adjectives (with some exceptions) normally follow the nouns they modify.

(6) le chat noir = the cat black ("the black cat")

English verbs normally follow the subject and precede the object, but Japanese verbs always follow both subject and object.

(7) Bill-ga hon-o utta = Bill book sold ("Bill sold the book")

In English, one can form a question by placing an "auxiliary verb"

* *Almost* the same meaning, at any rate. In Chapter 14 I'll mention some circumstances where completely accurate translation is not possible.

("do," "will," "be," etc.) in front of the subject, but in German questions, the main verb can be placed before the subject.

(8) Liebt Wozzeck Marie? = loves Wozzeck Marie?
("Does Wozzeck love Marie?")

These patterns of word order depend on knowing the parts of speech of the words, so they have to do with syntactic structure. On the other hand, since they differ from language to language, they can't depend on the meaning. So again we see that syntactic structure has properties that are independent of meaning.

Some syntactic patterns

If parts of speech don't have to do with meaning, what *do* they have to do with? It should be evident by now that the classification of words into parts of speech determines their roles in patterns.

Let's briefly explore some syntactic patterns in English. As we go through them, it is important to bear in mind that these patterns are part of mental grammar—that we somehow have these patterns stored in our brains, and that we had to learn them.

As we saw in Chapter 5, a noun can appear with a plural ending: "dogs," "bananas," "earthquakes." A verb, on the other hand, can appear with a past tense ending: "helped," "believed," "procrastinated." Notice that our ability to use these endings is *syntactic* knowledge, and doesn't follow from the meanings of the words. In terms of meaning, it would make sense if nouns that name actions could appear with a past tense. But there are no words "earthquaked" or "concerted" which name an earthquake or a concert that occurred in the past. Likewise, in terms of meaning, it would make sense to be able to put a plural ending on a verb to mean that the action was performed more than once. But we can't say "Bill will dances" to mean he will dance several times. (The "-s" ending in "Bill dances", of course, indicates not plural, but that there is a third-person-singular subject.) In other words, the availability of past tense and plural endings correlates with the syntactic distinction between verbs and nouns, not with the distinction in meaning between objects and actions.

Another case: in English, a verb is preceded by its subject and followed by its object, if there is one. Both the subject and object are noun phrases. As we saw a minute ago, the verb can also be modified by an adverb. So we find sentences like (9).

(9) The enemy rapidly destroyed the city.
subject noun phrase — adverb — verb — object noun phrase

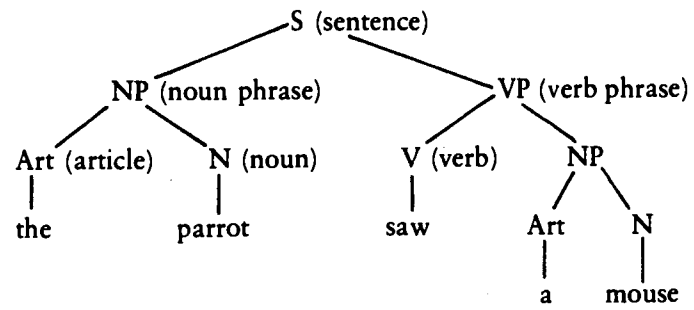
Now it so happens that there is a noun, "destruction," which describes the same action as the verb "destroy." But if we want to express something similar to (9) using this noun, the phrase comes out somewhat differently: not "the enemy rapidly destruction the city," which is gibberish, but (10).

(10) the enemy 's rapid destruction of the city
"subject" noun phrase — possessive — adjective — noun — preposition — "object" noun phrase

These differences can't have anything to do with meaning, since the meanings of (9) and (10) are parallel. Rather, they have to do with the syntactic patterns that go with verbs and nouns.

Let's look next at how subpatterns are put together into larger patterns. It is customary to notate the way a sentence or phrase is composed of patterns and subpatterns by drawing a "tree," like this:

(11)



This notation is similar to old-fashioned sentence diagrams in the way it breaks the sentence into parts. It is different in that it labels each part as belonging to a particular syntactic type such as N, VP, or Art. (11) says that the sentence "the parrot saw a mouse" is composed of two main parts, a noun phrase (the subject) and a verb phrase (the predicate). The noun phrase consists of two parts, the article "the" and the noun "parrot." The verb phrase also consists of two parts, the verb "saw" and the noun phrase "a mouse," which itself breaks into two parts, the article "a" and the noun "mouse."

Many variations in word order among the languages of the world become transparent when they are viewed in terms of tree structures. For example, the relative order of French adjectives and

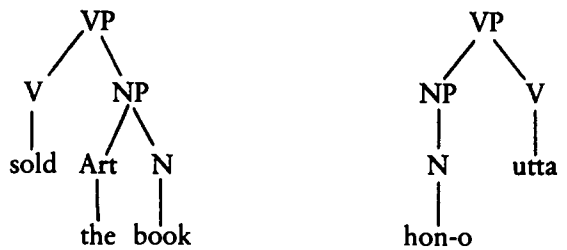
nouns differs from English in that the A under NP is swung around to the other side of the N.

(12) English French

(the cat black)

The relative order of the Japanese verb and direct object results from reversing the V and NP that hang from VP.

(13) English Japanese



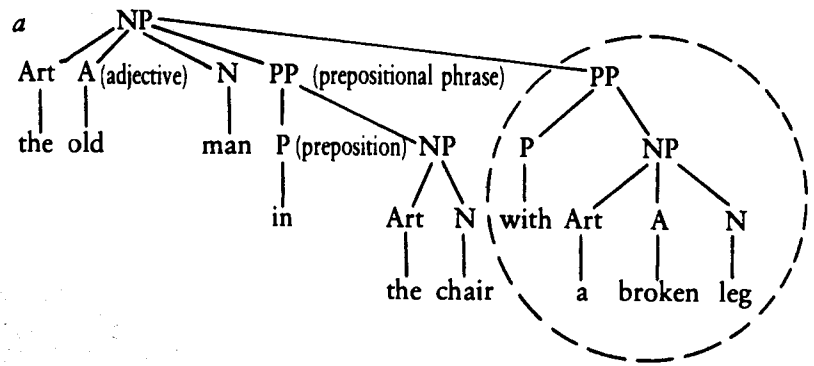
(book sold)

(On the other hand, the difference between English and German questions is more complicated, and I won't try to explain it here.)

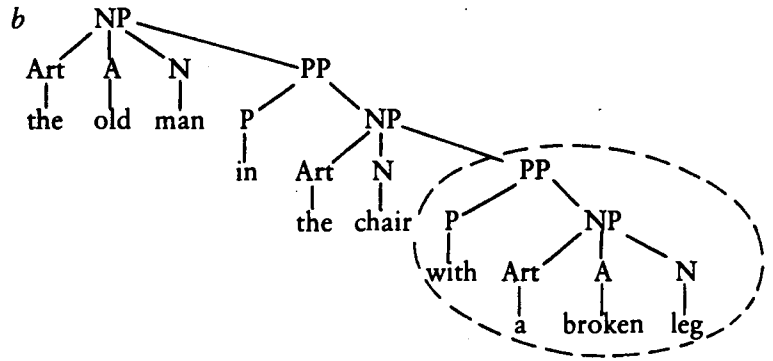
The tree notation also gives us a nice way to describe many cases of ambiguity. A typical case is shown in (14).

(14) the old man in the chair with a broken leg

a



b



The phrase in (14) is ambiguous: Does the old man have a broken leg, or does the chair? These two possibilities show up as the two trees in (14). The crucial difference is where the circled prepositional phrase "with a broken leg" is attached to the phrase. If it is attached to the upper noun phrase, it modifies "man"; if it is attached to the lower, it modifies "chair." The phrase is ambiguous because both structures result in the same string of words. In other words, the difference in meaning shows in the syntactic structure, but not in the phonological structure.

It is a continuing issue in linguistic theory to determine exactly which ambiguities in language are to be explained this way. For instance, the sentence in (15a) can be interpreted as (15b) or (15c).

(15) *a* Many people here have read two books.
b There are two particular books (say, *Gone with the Wind* and *Fear of Flying*) that many people here have read.
c Many people here have read two books, but not necessarily the same two.

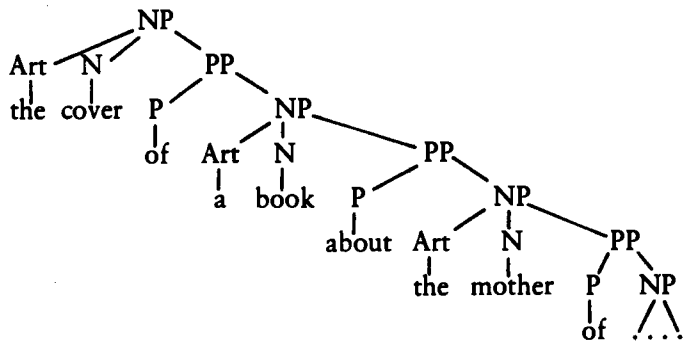
The question is whether sentence (15a), like the phrase in (14), has different syntactic structures for these different interpretations, or whether it has just a single syntactic structure that can be interpreted either way. I can't begin to go into the complexities of this case here; suffice it to say that the jury is out after nearly thirty years of investigation. This research has revealed a great deal about what the correct solution has to be able to account for, and about what possible solutions *won't* work, but there is still substantial disagreement among theorists on what the proper solution *is*. (A notable case of the Paradox of Language Acquisition: children trying to learn language probably don't have to make up their minds about this, even unconsciously!)

Recursion and syntactic movement

Let me continue a little more with tree structures.

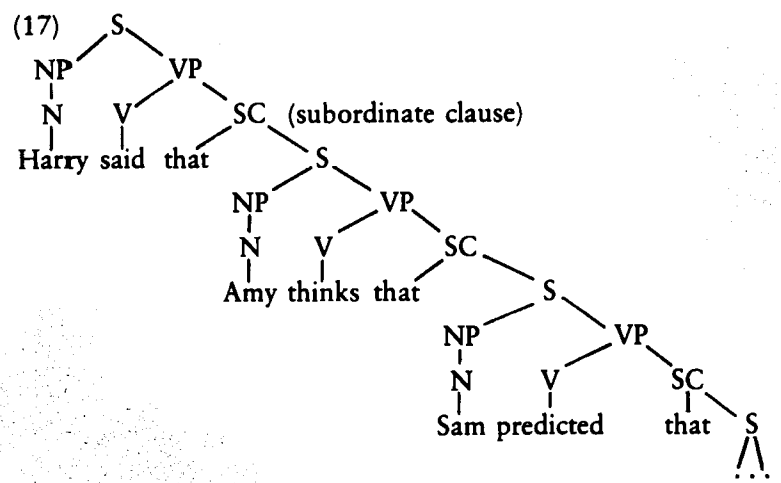
When we first discussed syntactic patterns in Chapter 2, we talked about the property of *recursion*, the ability of sentences to occur inside of other sentences *ad infinitum*, as in "Harry said that Amy thinks that Sam predicted that Mildred would believe that Beth is a genius." Tree structures give us a nice way to account for this pattern. Notice, for instance, that the trees in (14) have an NP inside another NP, and in fact (14b) continues the pattern to a second step. This pattern, too, can go on indefinitely, forming unwieldy but comprehensible phrases like "the cover of a book about the mother of the man in the chair with a broken leg." Here's a tree for part of this phrase.

(16)



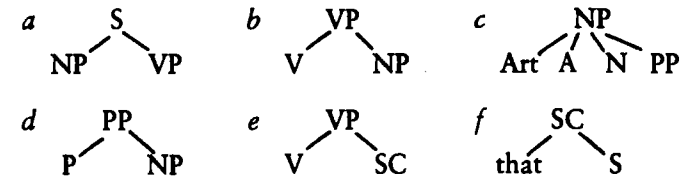
In similar fashion, our pattern "X Verbs that S" enters into tree structures like this:

(17)



We see, then, more explicitly than in Chapter 2, that long and complex sentences can be built up from a collection of relatively simple patterns. In fact, to a great extent, the basic syntactic patterns of the language can be thought of as subtrees only one layer deep, like these:

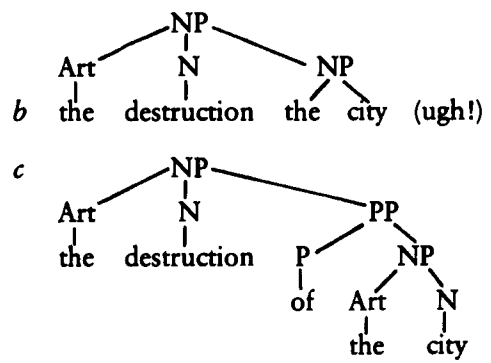
(18)



By putting these together like Lego pieces, trees of great complexity can be constructed. In addition, we can characterize the patterns of different languages by supplying them with a different stock of basic subtrees—as shown, for example, in (12) and (13).

This approach also enables us to show why not every combination is possible in every language. For instance, in English there is no subtree like (19a). As a result, a noun cannot occur with a "direct object" as in the ungrammatical (19b). Rather, the closest parallel to the verb phrase "destroy the city" is to use a prepositional phrase, as in (19c).

(19) *a* Not in English:



As usual, this organization of mental grammar is verified by extensive experimentation with speakers' intuitions about sentences, of the sort discussed in Chapter 4. In addition, this way of describing the patterns is claimed to be parallel to the way the brain organizes them, even if we are ignorant of how the neurons accomplish such organization.

Experiment also reveals that there are cases where subtrees like (18) are not enough to characterize our syntactic knowledge. The

examples in (20) are so-called *wh-questions*, which begin with one of the *wh-words* "who," "what," "where," "when," and "which."

(20) *a* What did Beth eat for breakfast?
b Whom did Harry think that Nancy met at the store?
c Which book does your mother say that the teachers think that the children will read over the summer?

In (20a), "what" is understood as the direct object of the verb "eat," even though it is at the beginning of the sentence instead of after the verb. We can see this because an appropriate answer is something like "Beth ate octopus for breakfast," in which the noun phrase that answers "what?" is in the usual position after the verb.

In order to account for the "understood" relation of "what" to the verb, linguists have proposed that in mental grammar there is a stage of analysis in which "what" really *is* the direct object of the verb—"Beth did eat what for breakfast." Then "what" in some sense "moves" to the front of the sentence. To help make sense of this idea, notice that there are two kinds of situations in which the wh-word actually is used as direct object instead of at the front. One is in a situation of incredulity (so-called "echo-questions"):

(21) Holy cow! Beth ate WHAT for breakfast?!
OCTOPUS??!!

The other is in a quiz-show sort of situation:

(22) Mr. Van Doren, for $64,000: on the morning of July 4, 1776, General Washington ate *what* for breakfast?

In addition, ordinary wh-questions in certain other languages do not place the wh-word at the beginning, but instead leave it in place, just like the wh-words in (21) and (22). (23) is an example from Korean, where, as in Japanese, the verb follows the direct object; notice that the phrase "which college" falls right in the middle of the sentence.

(23) Ne-nun Chelsoo-ka enu tayhak-ey kat-ta ko sayngkakha-ni?
You Chelsoo which college went that think*
'Which college do you think that Chelsoo went to?'

* I am grateful to Soowon Kim for this example. For clarity, I have omitted the suffixes from the translation. For completeness: "-nun" is a topic marker; "-ka" marks nominative case; "-ey" marks dative case; "-ta" marks indicative; "-ni" marks the sentence as a question.

The idea, then, is that echo-questions, quiz-show questions, and Korean questions show us the basic or "underlying" position of wh-words, from which they move to the front in ordinary English questions. This is one of the foundational ideas behind Chomsky's theory of *transformational grammar*: a sentence in the mind has an "underlying structure" or "deep structure" that is different from its surface form, and various principles of mental grammar can transform the sentence by moving certain parts such as wh-words around. This innovation, more than any other, is what permitted Chomsky's approach to syntactic structure to describe a wide range of linguistic phenomena that had remained relatively opaque to previous theories.*

According to current theory, when wh-words move, they leave behind a "trace," a sort of unpronounced pronoun, in the place where they came from. The idea of an "unpronounced pronoun" may seem rather mystical. But in fact, various techniques involving measuring reaction times and detecting brain waves have shown that people listening to sentences like (20) actually detect these traces of moved wh-words (unconsciously, of course). The trace is detected just at the point during the sentence where linguistic theory postulates it, and it has an interpretation appropriate to the wh-word that has been moved.

Constraints on long-distance dependencies

It turns out that wh-words can "move" to the beginning of the sentence from many different places. In "Whom did Harry think Nancy met at the store," "whom" is understood as the direct object of the subordinate clause "Nancy met ∧ at the store" (where the trace left by the moved wh-word is indicated by ∧). In the even longer sentence (20c), "which book" moves out of the doubly subordinate clause "the children will read ∧ over the summer." In fact, we can construct sentences in which a wh-word is understood as belonging

* Deep structure (now often termed "D-structure") has always been understood simply as an aspect of syntactic structure that expresses certain structural regularities. These regularities, since they include word order, have to be syntactic (i.e. distinct from meaning) and have to express variation among languages (i.e. they are not universal). However, because of the way deep structure was characterized at a certain stage of development of the theory during the mid and late 1960s, many commentators erroneously identified it with either meaning or Universal Grammar or both. However, this was never the intent of the term, except in certain circles for a certain brief time.

to a clause as deeply embedded as we want. (24) is very awkward, but nevertheless understandable.

(24) Whom did Sam say Harold thought the teacher had told us that Fred would get Susie to kiss ∧ last Tuesday?

Because it can involve many embeddings, the relation between a wh-word and its "original" or "understood" position cannot be expressed in terms of one-layer subtrees like those in (18). For this reason, the relation is called a *long-distance dependency*.

Oddly enough, though, a wh-word can't be moved from just anywhere in a sentence to form a question. Mental grammar is very particular on this point. To see what I mean, notice that the quiz-show questions in (25) are perfectly understandable.

(25) *a* For $64,000, Mr. Shmoo: General Washington ate kippers and *what* for breakfast?
b For a chance to go on to the bonus round, Ms. Glurk: The Empire State Building was completed in the year that *what actress* married a future Prime Minister?

But if I try to make the very same inquiries using garden-variety wh-questions, I am completely incomprehensible.

(26) *a* What did General Washington eat kippers and ∧ for breakfast? (ugh!)
b What actress was the Empire State Building completed in the year that ∧ married a future Prime Minister? (triple ugh!)

It's not that these are too long or complex to understand—they're shorter than the reasonably comprehensible sentence in (24). Moreover, their difficulty appears to lie in their syntactic structure, not their meaning. We know what the questions in (26) *ought* to mean—the same as those in (25)—it's just that we can't ask them this way. That is, these examples seem to reveal "imperfections" in mental grammar, situations where it doesn't allow us to express our thoughts as freely as a completely general system might.

The problem is not limited to wh-questions. (27) illustrates two other kinds of long-distance dependency in English.

(27) *a* Topicalization:
That kind of movie, I would never be caught dead sending my kids to ∧!

b Adjective + "though" construction:
Sophisticated though Susan thinks Bill is ∧, she'll still marry Clyde.

In example (27a), the topic of the sentence, "that kind of movie," appears at the beginning instead of in its normal position as the object of "to." (This construction may not be "proper" English, but we all say things like this in casual conversation.) In the more learned construction shown in (27b), "sophisticated" appears before "though" instead of in its normal position after "is." Like wh-questions, these constructions are thought of as involving a "movement" from the position indicated by ∧ to the front of the sentence.

However, the configurations from which movement can take place (indicated by ∧) are limited in the same way as in wh-questions. (28) provides examples analogous to to the unacceptable wh-questions in (26).

(28) *a* Topicalization:
That kind of cracker, I would never make my kids eat cheese and ∧ for breakfast! (ugh!)
b Adjective + "though" construction:
Sophisticated though Susan is aware of the fact that Bill is ∧, she'll still marry Clyde. (gack!)

Notice that we can see what these are supposed to mean: "I would never make my kids eat cheese and THAT KIND OF CRACKER for breakfast"; "Though Susan is aware of the fact that Bill is (very) sophisticated, . . ." It's just that we can't express these meanings using the long-distance-dependency constructions.

These sorts of situations were first pointed out (by Chomsky and, in much greater elaboration, by John Robert Ross) in the mid 1960s. It has turned out that, with minor variations, similar restrictions appear on *all* long-distance dependencies in *all* languages examined so far. Why should this be?

My imaginary skeptic breaks in:

Well, it's just that as children we never heard sentences like (26) and (28). If we had, we would have learned to make sense of them.

But look: we probably never heard sentences like (24) either, yet we can interpret (24) without too much trouble. The problem posed by sentences like (26) and (28) is: Exactly what patterns are in our mental grammars that make all these examples comprehensible *except* for those in (26) and (28)? Pushing it one step further, how did

we acquire patterns for such a great variety of cases without happening to include patterns for (26) and (28) among them? Furthermore, why do the restrictions on these patterns appear universally, as far as we can tell?

These questions should be reminiscent of our discussion in Chapters 2 and 3: we are back in the world of the Fundamental Arguments. Unfortunately, even to begin to answer them would take us far beyond the scope of this book. Suffice it to say that they have been a central preoccupation of syntactic research for three decades, so they constitute a prime instance of the Paradox of Language Acquisition.

Universal Grammar in syntactic structure

I've presented only the barest minimum of syntactic phenomena of English and the thousands of other languages of the world. But already the description has become fairly complex; perhaps it's even a little tough to follow. If so, that too is evidence for the Paradox of Language Acquisition—remember that you have unconscious command of all these patterns that I have had such trouble describing to you.

Again, if it is so complex, we are driven to invoke the Argument for Innate Knowledge and ask: With what aspects of these patterns is the child endowed in advance? That is, what is the Universal Grammar for syntactic structure?

Throughout this chapter, I have been at pains to show that syntactic structure is to a degree independent of both sound and meaning. The syntactic structures of sentences that we speak and hear are in no sense present in the physical reality of speech sounds or in the world of objects and events of which we speak. They are just part of the mind's internal accounting system, a step in the calculations that relate sound to meaning.

Let's think about how hard this situation is for the language learner. Prelinguistic children can observe objects and events in the world, they can hear speech sounds in the world, and presumably they can even form associations between speech sounds and objects ("CAR, Beth! See the CAR?"). But they can't observe *nouns* and *verbs* in the world—these are only internal classifications that correlate with syntactic patterns, which themselves depend on other internal classifications. So how is the child to figure out the system, if it can't be observed?

Current theorizing suggests, at the very least, that children don't have to figure out that there are such things as nouns and verbs that fit into hierarchical tree patterns. Universal Grammar provides the child with a substantial skeleton of syntactic structure that gets the acquisition process off the ground. Like the Universal Grammar for phonological structure, it is useful to think of this as a collection of universal stipulations of the possible units and relations available to all human languages, plus a kind of menu (like a computer software menu) that helps to guide the learner through the possibilities.

Here, stated very informally, are some aspects of UG that pertain to phenomena we have looked at in this chapter.

1. UG lets the child know that the expressive variety of language is made possible by combining local subtrees into larger assemblies. The child does not have to figure out that words are not just strung together one after another.
2. UG stipulates that a language contains a class of *nouns*, that the names of physical objects (among other things) are found in this class, and that a noun plus its modifiers constitute a syntactic unit *noun phrase*. It leaves open, though, where the modifiers are placed: is a modifying adjective before the noun, as in English, or after, as in French? The child has to figure this part out, but it's a lot less than inventing the whole idea of a noun phrase from scratch.
3. UG stipulates that there is a class of *verbs*, and that a verb can combine with a noun phrase to form a *verb phrase* (or predicate). It leaves open whether the verb precedes the noun phrase (English) or follows it (Japanese).
4. UG stipulates that there is a set of wh-words used to ask questions. It leaves open whether they can be left in place (Korean and special constructions in English), moved to the front (ordinary English questions), or moved directly before the verb (Hungarian).
5. UG stipulates that the patterns of a language may contain long-distance dependencies of various sorts, subject to a number of innate restrictions.

More generally, the linguist's strategy in forming hypotheses about Universal Grammar is to try to reduce the things the child must figure out to those that can be observed in the physical and linguistic

environment. Given a knowledge of words, a child can determine word order. But tree structures, and the categories present in them (NP, VP, etc.), can't be observed. They have to come from inside the mind, as it were "instinctively" or "intuitively." Universal Grammar is, if you like, the organization of that instinct.

7 American Sign Language

Basic facts about American Sign Language

I am concluding this part of the book with a brief discussion of American Sign Language (ASL), the language of the Deaf community in the United States and most of Canada. I have chosen to do so in part because of the intrinsic interest of signed languages, but more specifically because of the light they shed on language in general. In addition, research on ASL plays a prominent role in our discussion of the biological basis of language in Part III, so it makes sense for me to acquaint you first with some facts about the language.

The most important thing I want to stress is that *ASL is a language*. Of course, it seems completely different from familiar languages such as English, Russian, and Japanese. Its means of transmission is not through the speaker's vocal tract creating acoustic signals that are detected by the addressee's ears. Rather, the speaker's gestures create signals detected by the addressee's visual system.

Some people have found such communication altogether alien and magical. I'll try to show, though, that the differences are rather superficial. It is sort of like switching a VCR–TV system from videotape to videodisc—the peripheral system is different, but the inner workings are exactly the same.

Not exactly the same as *English*, though. ASL is emphatically not just a coding of English into manual signs. Such codings do exist, but they can be clearly distinguished from ASL. One is *fingerspelling*, in which each letter of the alphabet has a hand sign, so that English words can be spelled out letter by letter on the hand. This is used to sign names and technical terms of English for which there is no conventional ASL translation—much as words like "glasnost" and "perestroika" were borrowed whole into English some years back.

The other coding of English into signs is called *Manual English* (or, in one variant, Signing Exact English). It renders literally into