E 9.4

Speech Perception

9.4.1 Receiving Messages

As described in both File 1.2 and File 9.2, the language communication chain involves both sending and receiving messages. This file and the following two files are concerned with how we receive messages, that is, how we perceive and interpret spoken and written language.[1] The process of receiving and interpreting messages is called **speech perception.** Speech perception can be seen as the reverse of speech production: in speech production, we have an idea that we turn into an utterance, whereas in speech perception, we hear or see an utterance and decode the idea it carries.

Our ability to understand linguistic messages is quite remarkable. In a matter of milliseconds, we identify sounds, match them with words in our mental lexicon (see File 9.5), and apply syntactic rules to understand the meaning of the message (see File 9.6). We can do this even in a crowded and noisy bar. We can pick out relevant acoustic information (what someone is telling us) in the presence of other noises such as the person at the next table telling a joke, the waiter dropping a glass of beer, and the music playing in the background.

This file deals with the process of identifying the sounds of speech. This is a difficult task because no sound is ever produced exactly the same way twice. For example, if a person utters the word *bee* ten times, neither the [b] nor the [i] in each production will be physically identical. So how do we match, for example, a [b] with the category /b/ in our head if no [b] is physically the same as another? This is called the **lack-of-invariance** problem. This file introduces a number of speech perception phenomena that help explain how we deal with the lack-of-invariance problem and manage to match highly variable phones to phonological categories in our heads.

9.4.2 Categorical Perception

One phenomenon that helps explain how we deal with lack of invariance is **categorical perception**, which occurs when equal-sized physical differences are not equal-sized psychologically, that is, when we do not perceive a continuum as a continuum. Rather, differences **within** categories are compressed, and differences **across** categories are expanded. People come to perceive entities differently after they learn to categorize them. In particular, members of the same category are perceived to be more alike than members of different categories, even if they are equally different. For example, two different kinds of yellow are perceived to be more similar than a yellow and a red, even if the differences in wavelength between the colors in each pair are identical. This is a case of categorical perception.

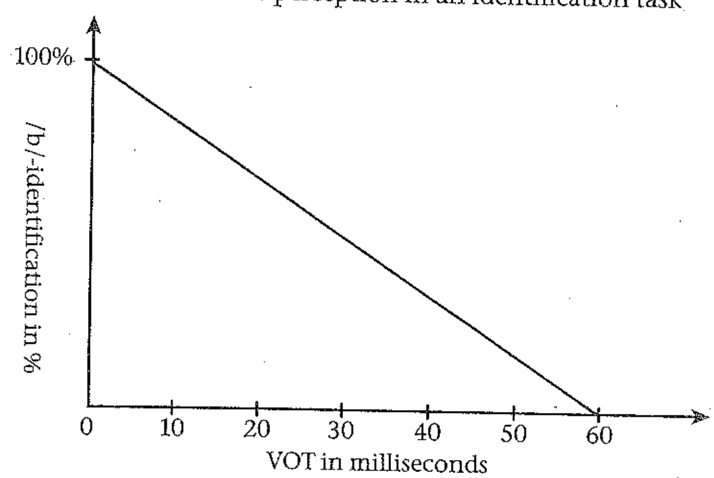
Categorical perception also occurs in language, particularly in consonant perception. Let's look at an example. The sounds [b] and [p] differ only in voicing: [b] is voiced, but [p] is voiceless (see File 2.2). Consider the syllables [bɑ] and [pɑ]. Physically, these sounds dif-

[1]The perception of signed language will not be considered here.

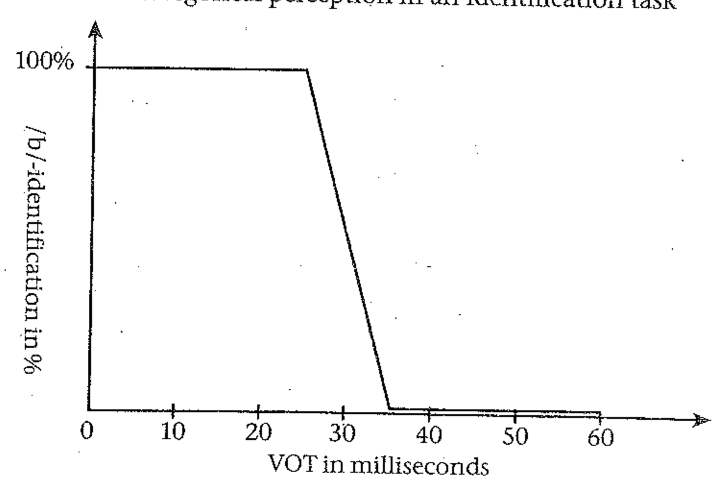
fer in their **voice onset time** (VOT), the time between the opening of the lips at the end of the stop and the beginning of vocal-fold vibration or voicing in the following vowel. In English, a bilabial stop with a 0-millisecond VOT would always be perceived as /b/. In this case, the voicing starts as soon as the stop is released (hence a value of 0 ms for the VOT). However, a bilabial stop with a 60-millisecond VOT would always be perceived in English as a /p/. Here the voicing for the /ɑ/ starts 60 milliseconds after the stop is released.

But what about a bilabial stop with a 10-, 20-, 30-, 40-, or 50-millisecond VOT? Would it be perceived as a /b/, as a /p/, or as something in between? The answer to this question can be determined by synthesizing bilabial stops with varying VOTs followed by the same vowel, for example, [ɑ], and asking people whether they heard /bɑ/ or /pɑ/. That is, we have a series of syllables of the form [bilabial stop]+[ɑ]; each bilabial stop has a different VOT value, ranging from 0 to 60 ms. If people listening to this continuum of VOT values were to perceive it as a continuum between /bɑ/ and /pɑ/, we would expect the results of such a task to look like the graph in (1). The larger the VOT, the more /p/-like the sounds would be perceived; we would see a gradual decline in /b/-identification and a gradual increase in /p/-identification as the VOT increases. But if /b/ and /p/ are perceived categorically, we would expect a graph like the one in (2). Sounds within one category (either the /b/ category or the /p/ category) would be perceived as similar or the same, but sounds across category boundaries would be heard as different. In this case we would expect to see a sharp drop in /b/-identifications at the category boundary.

(1) Schema of continuous perception in an identification task



(2) Schema of categorical perception in an identification task



As the McGurk effect illustrates, despite considerable variability in the acoustic signal, we are able to draw on both acoustic properties and visual information to identify speech sounds.

9.4.5 Other Factors Involved in Speech Perception

The previous sections showed that we are able to identify phonological categories despite high variability in the speech signal because our perceptual systems can accommodate many of the contributing factors. There are a number of additional factors that help us categorize sounds, such as our knowledge of phonotactic constraints, the words in our mental lexicon, and the context of an utterance.

As discussed in File 3.3, listeners have unconscious knowledge of the phonotactic constraints of their language. One source of evidence for this knowledge comes from perception errors, also called slips of the ear. Specifically, errors of this type always result in possible (though not always actual) words. For example, if we hear the beginning of a word and are not sure whether we heard the sequence [ʃk] or [sk], we can conclude that it was [sk] since English does not allow the consonant cluster [ʃk]. Listeners know what to expect in the way of sequences of sounds; if they did not have this knowledge, we would expect listeners to mistakenly hear words made up of sounds sequences that are impossible in their language.

The words in our mental lexicon can also help us identify individual sounds. For example, if we are not sure whether a sound we heard was an /m/ or an /n/, we can determine that it probably was an /m/ if it was preceded by the phones [kɹi] since *cream* is a word of English, but *crean* is not. On the other hand, if the sound was preceded by [kli], we can determine that it was probably an [n] since *clean* is a word of English, but *cleam* is not. Finally, the linguistic context of an utterance can help us identify sounds. For example, the word *heel* is more likely to appear in the context of shoes, whereas the word *peel* is more likely to occur in the context of oranges.

An effect called **phonemic restoration** illustrates how strongly both of these factors influence speech perception. In an experiment by Warren and Warren (1970), participants were played one of the sentences in (3). The * indicates that a sound was replaced with a cough. Interestingly, participants heard **eel* as *wheel, heel, peel,* or *meal,* depending on the context that followed. For example, for (3a), they heard *wheel,* while for (3d), they heard *meal.* This means that participants "heard" a sound that was actually not present in the acoustic signal because it fit into the context of the utterance. Furthermore, when they were told that a sound was missing from the utterance and were asked to guess which one it was, listeners were unable to identify the missing sound.

(3) a. It was found that the *eel was on the axle.
 b. It was found that the *eel was on the shoe.
 c. It was found that the *eel was on the orange.
 d. It was found that the *eel was on the table.

FILE 9.5

Lexical Processing

9.5.1 What Is Lexical Processing?

Lexical processing refers to the way that words are recognized. People can recognize words extremely quickly, even though the average college-educated adult knows between 40,000 and 60,000 words to select from (Miller 1996). In fact, it takes only about 250 milliseconds to "find" a word in the **mental lexicon**. This is remarkably fast since the task of finding a word is harder than it seems: many of the 40,000 to 60,000 words are similar given that they are constructed from a relatively small number of phonemes. For example, *state, snake, stack, stay, ache,* and *take* all sound similar to *steak.*

This file discusses how words are stored in the mental lexicon, the process of word recognition, and what word recognition reveals about the organization of the mental lexicon.

9.5.2 The Mental Lexicon

Think of the mental lexicon as a dictionary located in the brain, containing all the words an individual knows, including what each word means, how it is used in conjunction with other words, and how it is pronounced. Keep in mind, however, that this dictionary is not a tangible object but rather some abstract network of information in the brain. We cannot point to it, but we have strong reasons to believe that it exists.

We have just said that the mental lexicon contains all the words an individual knows. But given the productive nature of language, some linguists have questioned whether this is, in fact, the case. For example, they wonder whether complex words like *internationalization, generalization, visualization, globalization,* and *liberalization* are stored as whole words or are broken up into morphemes.

It seems reasonable to assume that the fastest way to access such complex words would be if they were all stored as separate, individual entries. This is called the **full listing hypothesis** (Butterworth 1983). In this case, *internationalization, generalization, visualization, globalization,* and *liberalization* would each be stored as a separate entry, as would the words *nation, national, international, nationalize, internationalize, nationalization, general, generalize, visual, visualize, globe, global, globalize, liberal,* and *liberalize.* However, storing every word individually, while allowing for fast access, takes up a lot of storage space; it also seems to miss the generalization that these words are in some way related to each other.

Alternatively, complex words could be broken down into morphemes, with each morpheme stored individually in the lexicon (see File 4.1). This is called the **affix-stripping hypothesis** (Taft and Forsters 1975). According to the affix-stripping hypothesis, complex words are built from the morphemes listed in the lexicon. For example, the words *internationalization, generalization, visualization, globalization,* and *liberalization* would be built from the entries *nation, general, visual, global, liberal, inter-, -al, -ize,* and *-ation.* Storing only morphemes would greatly reduce the number of entries in the lexicon. For example, if each of our 5 complex words and the 15 related words from the previous paragraph were listed separately, we would need 20 different entries. In contrast, we would need only 9 entries if only the individual morphemes were listed. This would not only reduce storage space, but it would also highlight the nature of the relationship between these words. Yet it would slow down lexical access since words would be constantly decomposed into and rebuilt from individual morphemes during comprehension and production.

Most recent models propose that the truth lies somewhere in between. That is, some complex words may be stored as whole units, while others may be stored partially decomposed or even fully decomposed. In most models, the degree of decomposition depends on the kinds of affixes a complex word contains. Some affixes are very transparent: attaching them to a root does not change the way the root is pronounced, and the meaning of the complex word is completely compositional. An example of such an affix is *-less* [ləs] 'without,' which can be attached to, for example, *hope* [hoʊp] and *clue* [klu] to create the words *hopeless* [hoʊp + ləs] 'without hope' and *clueless* [klu + ləs] 'without a clue.' Other affixes are not as easily decomposed since attaching them to a root frequently changes the way the root is pronounced. For example, when we attach *-ity* [ɪɾi] to *severe* [səviɹ], the pronunciation of the second root vowel changes from [i] to [ɛ], giving *severity* [səvɛɪ + ɪɾi]. Decomposing words with such affixes is not straightforward since the meaning is not always obvious from the parts. It is thus possible that a word such as *severity* is stored as in its entirety, whereas *clueless* is decomposed and stored as two elements: *clue* and *-less*. While there is experimental evidence supporting this intermediary hypothesis (see, for example, Vannest and Boland 1999), more research is needed to understand which affixes are listed separately in the mental lexicon and which ones are not.

Whether complex words are stored as a whole or are decomposed into individual morphemes may also depend on language and, in particular, on how many morphologically complex words a language has and how complex they are. Many analytic languages, such as Mandarin, have relatively few morphologically complex words. Thus, all the words of Mandarin could probably be stored in the lexicon as a whole without taking up too much storage space. However, agglutinating languages such as Turkish have many morphologically complex words. An example of how complex words in Turkish are formed is given in (1) (Hankamer 1989).

(1)	göz	=	"eye"
	gözlük	=	"glasses"
	gözlükçü	=	"seller of glasses (oculist)"
	gözlükçülük	=	"the occupation of oculists"
	gözlükçülükçü	=	"a lobbyist for the oculist profession"
	gözlükçülükçülük	=	"the occupation of being a lobbyist for the oculist profession"

Thus, for languages like Turkish, storing morphologically complex words as a whole could require more storage space than reasonably could be allotted to the mental lexicon. Hankamer (1989) even suggests that storing all the words of Turkish in their full form would require more storage space than the human brain has.

There are three major areas of study in the domain of psycholinguistics; so far we have considered how language is represented in the brain and how language is acquired. But another major area is how people use their knowledge of language. How do they understand what they hear? How do they produce messages that others can understand in turn?

We have already begun to answer these questions with regard to speech production (File 9.3) and speech perception (File 9.4). Much of the research on adult normal processes of the last twenty or thirty years, however, has focused on three areas: word recognition, syntactic analysis, and interpretation. This file will introduce the first of these, word recognition, and File 9.6 will deal with syntactic analysis and interpretation.

9.5.3 Word Recognition

An important step in understanding any linguistic message is being able to recognize the meaning of words. As seen in File 5.2, the meaning of a sentence is determined in part from the meanings of the words it contains. Identifying a word also tells us what syntactic category it belongs to, thus determining the range of phrase structures it can occur in (see Files 5.3 and 5.4). A word also provides information about the syntactic structure of the rest of the sentence. When we recognize a verb like *put,* for example, we expect certain types of information to follow. In particular, we expect to hear the object of the 'putting' (e.g., *the plate*) followed by where the object is being put (e.g., *on the table*). When we hear the verb *arrive,* we know that it does not take an object (we do not 'arrive someone' or 'arrive something'), but that it may take a prepositional phrase, as in *arrive at the hotel.* This kind of information is made available to us when we recognize a word and access the information pertaining to it in our memory.

Identifying words is such an effortless task under most conditions that we typically don't realize how difficult it really is; designing a computer system to accomplish this task, for instance, has proven very difficult. One reason for the difficulty is the sheer number of words that any given person knows: the average six-year-old already knows about 14,000 words, while a college-educated adult, as mentioned above, knows between 40,000 and 60,000 words. A second reason is the large amount of other linguistic information that is accessed during word recognition and stored along with each word, as illustrated just above.

9.5.4 The Cohort Model

One of the essential parts of understanding words is recognizing them as they come at us in a stream of speech. How do we go about recognizing words? One common-sense view that receives a lot of support from experimental evidence is that as soon as people hear speech, they start narrowing down the possible words that they may be hearing. If the first sound that they hear is /s/, all words beginning with any other sound are eliminated, but words like *summer, spring, stone, sister,* and *spine* remain on the "list of possible words." If the next sound is /p/, many other possible words are ruled out, including *summer, stone,* and *sister.* A word is identified as soon as there is only one possibility left. This is called the **uniqueness point** of a word, and we refer to this model of word recognition as the **cohort model** of lexical access. The word **cohort** refers to all the words that remain on the "list of possible words" as the auditory input progresses. The words that are activated as the first sound of a word is perceived are called the **initial cohort.** The cohort model hypothesizes that auditory word recognition begins with the formation of a group of words at the perception of the initial sound and proceeds sound by sound with the cohort of possible words decreasing as more sounds are perceived.

Several experiments have supported this view of word recognition. For example, one obvious prediction of this model is that if the beginning sound of a word is missing, recognition will be much more difficult, perhaps even impossible. As early as 1900, experiments showed that word recognition is impaired much more when the initial sound of a word is mispronounced than when the final sound is mispronounced. This supports the cohort theory: if the end of the word is missing, it can be predicted based on the initial portion, while it is much more difficult to use the end to predict the early part of the word.

Although this model makes a lot of intuitive sense and has some experimental support, it leaves several questions unanswered. One problem is that in listening to running speech, people can't always identify where a word starts. In written English, boundaries are fairly clearly marked, but this is not the case for spoken language: there is usually no pause between one word and the next. So, for example, when someone hears the sequence [pɑpəpoʊzd], it is not at all obvious whether the utterance should be understood as *Papa posed* or as *Pop opposed.* If people need to know a word's first sound in order to identify it, *opposed* might not even be part of the "list of possible words." This is because [ə] could be interpreted as the last sound of [pɑpə], rather than the first sound of [əpoʊzd]. More examples of this type of ambiguity relating to word boundaries are shown in (2). If you say these phrases aloud at a normal conversational speaking rate and without context, you may have difficulty identifying where the word boundaries are.

(2)	the sky	this guy
	a name	an aim
	an ice man	a nice man
	I scream	ice cream
	see Mabel	seem able

Some other problems for the cohort model are described in the following section; each suggests additional aspects of word recognition that need to be addressed for a theory of word recognition to be correct.

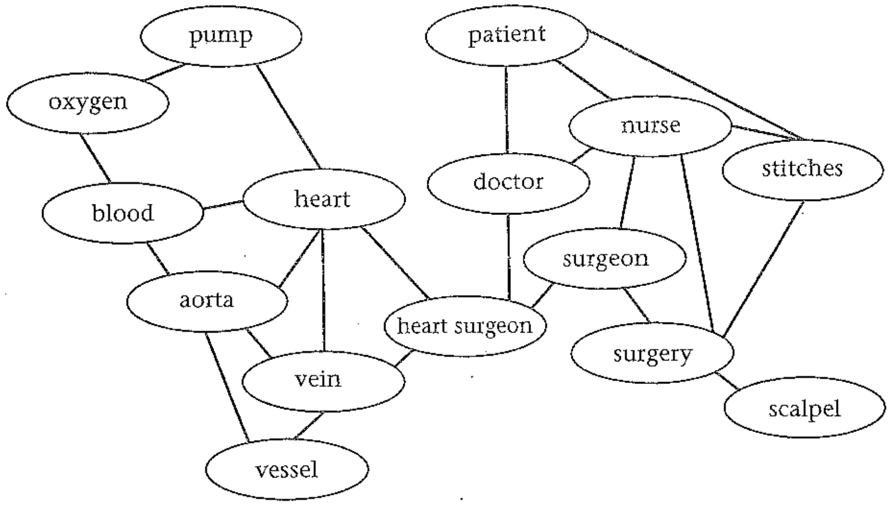
9.5.5 Frequency, Recency, and Context Effects

One of the most important factors that affect word recognition is how frequently a word is encountered in a language. This **frequency effect** describes the additional ease with which a word is recognized because of its more frequent usage in the language. For example, some words (such as *better* or *TV*) occur more often than others (such as *debtor* or *mortgage*), and words that occur more frequently are easier to access. This effect is not easy to explain assuming the beginning-to-end word recognition approach sketched out above. One possible explanation of the frequency effect is that the lexicon is partially organized by frequency and that in addition to being accessed from beginning to end, more-frequent words are "at the top of the list" of cohorts.

People also recognize a word faster when they have just heard it or read it than when they have not recently encountered it; this phenomenon is known as the **recency effect.** Recency effects describe the additional ease with which a word is accessed because of its recent occurrence in the discourse or context. Note that frequent words are likely to have been encountered more recently than infrequent words, so it may be possible to explain the frequency effect as a recency phenomenon, thus reducing the number of separate effects that have to be explained.

Another factor that is involved in word recognition is **context.** People recognize a word more readily when the preceding words provide an appropriate context for it. For example, in the sentence *This is the aorta,* people are not given any context that helps identify the word *aorta.* But in the sentence *The heart surgeon carefully cut into the wall of the right aorta,* many people would find that the cue of the *heart surgeon* helps them identify the word more quickly. One mechanism that has been proposed to account for this kind of **context effect** is a **semantic association network** (see (3)). This network represents the relationships between various semantically related words. Word recognition is thought to be faster when other members of the association network are provided in the discourse. It is obvious that the meaning of a word is tied to our understanding and general knowledge of the concept to which it refers. Thus, it is not unreasonable to suppose that hearing the words *heart surgeon* not only activates the direct meaning of the words *heart surgeon* but also makes a number of associated concepts more available to the hearer, such as those involved in the physiology of the heart, modern surgical procedures, and so on. These concepts are in turn linked to the words that are used to refer to them, making these words also more available.

(3) Example of part of a semantic association network



9.5.6 Lexical Ambiguity

Much research has centered on how ambiguous words are understood. How, for example, does a listener know whether *bug* refers to 'an insect' or 'a listening device'? This kind of ambiguity is called **lexical ambiguity,** meaning that a single lexical item—that is, a word—can have more than one meaning.

There are two main theories of how lexical ambiguity is resolved by listeners. The first claims that all the meanings associated with the word are accessed, with the correct meaning eventually being winnowed out (similar to the idea of the cohort model for word recognition). The second claims that only one meaning is accessed initially and that the listener holds on to this meaning until finding evidence that contradicts this first assumption. Support for the first position comes from experiments such as the following. When people are asked to finish a sentence, they take longer when the fragment to be finished contains an ambiguous word than when the ambiguous word is replaced by an unambiguous term, as in the sentences in (4).

(4) After taking the right turn at the intersection . . .	(*right* is ambiguous: 'correct' vs. 'rightward')
After taking the left turn at the intersection . . .	(*left* is unambiguous)

This delay suggests that all meanings of ambiguous words are accessed and that time has to be taken to decide among them; if only one meaning were accessed, we would not expect a difference in timing between ambiguous and unambiguous words.

However, other experiments suggest that under some circumstances, only one meaning is initially accessed. The frequency effect mentioned above has been shown to be important here. If a word has more than one meaning and one of the meanings is more frequent than the other, people tend to assign the more frequent meaning to the word. *Chair,* for example, has at least two meanings: 'an object to sit on' and 'the head of a department or committee.' The former meaning occurs much more often in speech than the latter, and, as a result, when people hear the word *chair,* they tend to associate it with the more frequent meaning and in fact may be slower at processing the sentence if the less frequent meaning was intended. This seems to suggest that only one meaning is initially considered, at least for words whose meanings differ markedly in frequency of occurrence. Similarly, if the conversation has centered on one particular meaning of an ambiguous word (e.g., if the conversation is about how the committee chairman lost his temper at a meeting), then a listener may access only the less frequent meaning of *chair* in the sentence: . . . *and then the chair was reprimanded by the supervisor.* This is an example of the recency effect influencing word ambiguity resolution.

In addition to frequency, semantic context also plays a significant role in determining which meaning is relevant. For instance, when a word like *bug* is seen in the context of *spy,* it is reliably identified as meaning 'a listening device,' but in the context of *flies and roaches,* it is identified as meaning 'an insect.'

In summary, lexical processing is concerned with how words are stored in the mental lexicon and how those words are accessed. There is a lot of evidence indicating that the process of lexical access starts as soon as the first sound of a word is heard; that is, we don't wait until we've heard the entire word to retrieve it from our lexicon. Rather, much like the cohort model suggests—but taking into account frequency, recency, and context effects—the list of possible words is narrowed down until the uniqueness point of the word is reached, and then the word is recognized. In addition, lexical access is incredibly fast: we can recognize a word in about 250 milliseconds, a remarkable feat considering the number of words in our mental lexicon.

FILE 9.6

Sentence Processing

9.6.1 How Do We Put Words Together?

The previous file described how we access words in our mental lexicon and some of the factors that can influence this lexical access. But language is not only about words; it is also about putting words together to form sentences. File 5.4 introduced many of the rules that we follow when we form sentences from words. However, for any given sentence that we utter or hear, most of us could not explain which rules we use to correctly form the sentence. That is, language users know how to build sentences in their native language(s), but they are not necessarily conscious of the rules that underlie these cognitive processes. Recall that these rules are part of our linguistic **competence** (see File 1.2).

Just as the word recognition process starts as soon as we hear the first sound of a word (see File 9.4), there is good evidence that we start building a syntactic structure as soon as possible. Let's take the sentence *The jealous woman went away* as an example. As soon as we hear the word *the,* which can only be a determiner, we expect the next word to be a noun (e.g., *the woman, the platypus,* etc.), an adjective (e.g., *the jealous woman, the angry platypus,* etc.), or an adverb (e.g., *the incredibly jealous woman, the vaguely angry platypus,* etc.). We also expect *the* to be part of a noun phrase and part of the subject of the sentence. Upon hearing *jealous,* we update our expectations concerning what comes next (e.g., *the jealous woman, the jealous platypus,* etc.). That is, as we hear a sentence unfold, we assign words to lexical categories (File 5.3) and build a syntactic structure that is updated as a new word comes in. This reconstruction of the syntactic structure of a sentence that is heard or read is called **syntactic parsing.**

9.6.2 Structural Ambiguity

If all words were unambiguous and had only one possible lexical category, processing sentences would be a relatively easy task. However, this is not the case. Both **lexical ambiguity** (when a word has more than one meaning; see Section 9.5.6) and **structural ambiguity** are constantly present during sentence processing. Structural ambiguity, first introduced in File 5.4, occurs when a sentence has two different possible parses resulting from different possible syntactic structures. Below, we consider several different types of structural ambiguity and the problems they cause for sentence processing.

a. Temporary Ambiguity. Let's have a closer look at the example above. We said that the word *the* is unambiguously a determiner. So far, so good. But as soon as we hit the word *jealous,* we encounter our first ambiguity. *Jealous* can be an adjective, as in *the jealous woman,* or a noun, as in *the jealous are troublesome.* Thus, the ambiguity is due to the ambiguous category of *jealous.* Once we hear *woman,* we can easily determine that *jealous* is used here as an adjective. However, now *woman* is ambiguous between being a noun, as in *The jealous woman went away,* or an adjective, as in *The jealous woman plumber went away.* The ambiguity is finally resolved once we hear *went:* we now know that *woman* is a noun, and *went* is not ambiguous. This means that the sentence *The jealous woman went away* is ambiguous only temporarily, namely, up until we hear the word *went*. **Temporary ambiguity** is constantly present in everyday conversations. For example, the vast majority of sentences that start with *the* followed by an adjective or noun (e.g., *the good, the tea, the bad, the dream, the small, the dog, the educated, the paper, the slow, the party, the old,* etc.) are temporarily ambiguous in English because most adjectives can also be nouns and most English nouns can also be adjectives.

b. The Garden Path Effect. As listeners comprehend temporarily ambiguous sentences, they sometimes momentarily recover a meaning that was not intended by the speaker. These mistakes in syntactic parsing are called **garden path effects** because the syntax of the sentence has led the comprehender "down the garden path" (to a spot where they can go no further and must retrace their steps). Garden path sentences are temporarily ambiguous and initially interpreted to have a different syntactic structure than they turn out to have. Let's look at an easy example, given in (1).

(1) a. While Mary was mending the sock fell off her lap.
  b. While Mary was mending the sock it fell off her lap.

When we first read *the sock* in (1a), we are likely to interpret it as the direct object of the verb *mending*. That is, we interpret the fragment to mean that Mary was mending the sock. However, at the verb *fell,* we notice that this parse could not have been correct (because then *fell* would have no subject), and we have to go back and reanalyze the sentence. In this case, we come to the conclusion that *the sock* is not the direct object of *mending,* but the subject of the main clause *the sock fell off her lap*. Such garden path sentences fool us into temporarily interpreting the wrong syntactic structure.

But why are we led down the garden path? The explanation depends on both the syntactic structure of the sentence and the particular lexical items it contains. In (1) we are led down the garden path because the verb *mending* can be (and often is) transitive; that is, it takes a direct object: we can mend something, as in (1b), where Mary is actually mending the sock. The problem, however, is that *mending* can also be intransitive (stand alone without a direct object), as in the sentence *Mary fell asleep while she was mending*. In (1a), the initial assumption is that *mending* is transitive and that *the sock* is its direct object; it is only when we come to the verb *fell* that we realize this initial assumption is wrong. This sentence is then a garden path sentence because of this particular property of the verb *mending*. Compare this to the sentence *While Mary was sneezing the sock fell off her lap*. Even though the sentence has the same structure as (1a), we are not led down the garden path because people usually don't sneeze socks; rather, people just sneeze.

Not all garden path sentences are as easy to recover from as the one above. In fact, for some sentences it can take quite a long time to figure out another structure if our first choice turns out to be incorrect. Some difficult garden path sentences remain **unparsable** for some people. That is, some garden path sentences are so difficult that some people never figure out the correct interpretation. To them the sentence remains ungrammatical. A famous example of a difficult garden path sentence is given in (2).

(2) The horse raced past the barn fell.

If we interpret the sentence as being about the horse racing past the barn, the sentence seems ungrammatical. This kind of sentence contains a reduced relative clause (that is, a relative clause that lacks the word *that* together with a form of the verb *to be*), in this case, *that was*. Thus, *raced* is not the action of the main clause but the verb of the reduced relative clause, and the sentence means *The horse* ***that was*** *raced past the barn fell.* Notice that this nonreduced version of (2) is easier to parse. However, both sentences are grammatical and convey the idea of a horse falling while someone was racing it past a barn. To help you understand that *The horse raced past the barn fell* is indeed grammatical, consider the sentence in (3).

(3) The woman driven to the hospital fainted.

This sentence also contains a reduced relative clause and has exactly the same syntactic structure as our difficult garden path sentence. However, people have no trouble identifying (3) as grammatical. If both (2) and (3) have the same syntactic structure and if (3) is grammatical, then our garden path sentence must also be grammatical. Then why is (2) so much harder to parse than (3)? The answer again lies in the lexical items, in this case the words *raced* and *driven*. First, notice that as a stand-alone sentence, *the horse raced* by itself is fine, but *the woman driven* is ungrammatical. Now consider the simple past and the past participle forms of *race* and *drive* given in (4).

(4)

Infinitive	*Simple Past Tense*	*Past Participle*
drive	drove	driven
race	raced	raced

Notice that the simple past and past participle forms of *race* are identical. This conspires with the fact that a sentence-initial noun phrase such as *the horse* is much more likely to be followed by the verb of the main clause (e.g., *raced* or *drove*) than by a reduced relative clause (e.g., *raced* or *driven*). In conjunction, these facts mean that in the case of (2), we interpret *raced* as the verb of the main clause because this is not only a possible parse but also the more frequently encountered option. In the case of (3), however, we cannot interpret *driven* as the verb of the main clause. This would be ungrammatical and explains why we have little trouble interpreting *driven* as the verb of a reduced relative clause instead: unlike *raced*, it cannot be the verb of the main clause.

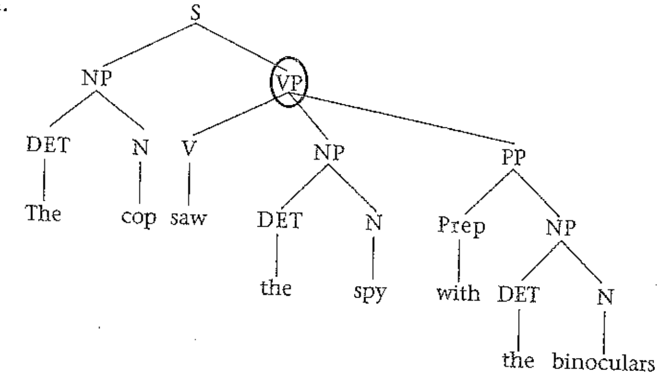
c. Globally Ambiguous Sentences. Not all ambiguous sentences are only temporarily ambiguous. There are also sentences that are **globally ambiguous,** that is, sentences in which the ambiguity is not resolved by the end of the sentence. Without additional context (such as intonation or preceding/following sentences), there is no way to determine which structure and meaning the sentence has. A typical example of a globally ambiguous sentence is given in (5).

(5) The cop saw the spy with the binoculars.

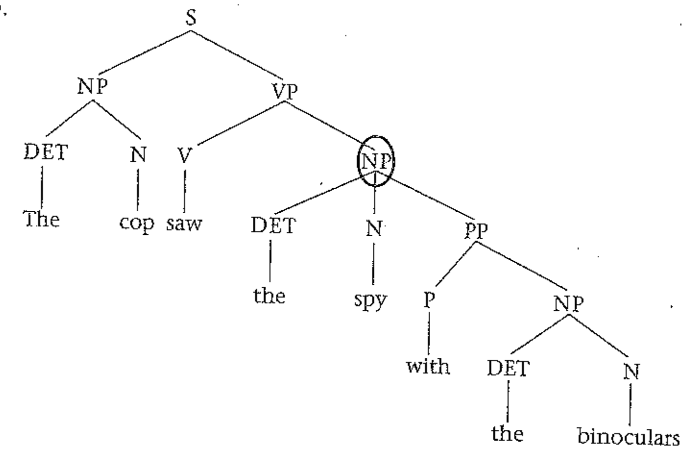
The ambiguity lies in how the prepositional phrase *with the binoculars* fits into the rest of the sentence. It could modify the verb phrase *saw the spy,* in which case the sentence means that the cop used binoculars in order to see the spy. This interpretation corresponds to the syntactic structure given in (6a). Intuitively, this structure fits well with the interpretation that the binoculars are used to see the spy. Alternatively, *with the binoculars* could modify the noun phrase *the spy,* in which case it specifies that the spy has binoculars. This is shown in (6b). Sentences that are globally ambiguous always have two or more possible syntactic structures, one corresponding to each interpretation.

(6) Syntactic representation of the two meanings of the same sentence

a.



b.



An important question in sentence processing is how people decide which structure a globally ambiguous sentence has. As with lexical ambiguity, people could consider all possibilities and then decide which one is best, or they could use some strategy to decide which interpretation to consider first and then reconsider if that interpretation does not work out.

The garden path phenomenon introduced above suggests that people try one analysis first and consider other possibilities only when the initial analysis does not work out. If people initially considered all the possibilities, they would not be led down the garden path. But what strategies could people use to decide which structure to consider first? One influential syntactic parsing strategy that has been proposed is called **late closure.** It suggests that incoming material is incorporated into the clause or phrase currently being processed if possible. In other words, people attach material to the closest clause or phrase, as long as such attachment is possible. In our example, *The cop saw the spy with the binoculars,* this means that *with the binoculars* should be preferentially attached to modify *the spy* rather than *saw.* If you look at the syntactic trees in (6), you can see that the noun phrase *the spy* is "closer" to the prepositional phrase *with the binoculars* than is the verb phrase *saw the spy.* What late closure suggests then is that, all else being equal, people would interpret the sentence in (5) to mean that the spy has the binoculars.

However, many other factors contribute to how we interpret globally ambiguous sentences. Obvious factors are the choice of lexical items and the preceding context. If we change the sentence in (5) to *The cop saw the squirrel with the binoculars,* we would be more likely to interpret it to mean that the cop used the binoculars to see the squirrel than that the squirrel had the binoculars, given that squirrels usually don't have binoculars. On the other hand, if the sentence was preceded by the context in (7), we would probably interpret the sentence to mean that the squirrel had the binoculars, despite that fact that this is an unlikely occurrence.

(7) An unusually large squirrel stole a movie star's million-dollar binoculars. The star called the police to report the incident, and a cop was assigned to look for the stolen item. After an hour, the cop saw the squirrel with the binoculars.

Both the choice of lexical items and the preceding context can so strongly favor one interpretation over the other that we may not even notice that a sentence is ambiguous. In fact, naturally occurring conversation is full of ambiguities that are never detected.

Another factor that influences ambiguity resolution is **intonation.** Many spoken ambiguous sentences can be disambiguated through intonation (and punctuation can be used to disambiguate an otherwise ambiguous written sentence). The sentence in (8a), for example, is ambiguous. It can mean either that Jack and Paul will be invited or else that Mary will be. Alternatively, it can mean that Jack will be invited, and so will either Paul or Mary, but not both of them. Depending on how the sentence is pronounced, listeners will favor one interpretation over the other. In particular, the sentence can be said with a **prosodic break** (see File 2.5) after *Paul,* as illustrated in (8b). To see how this works, say the sentence aloud and pause after the word *Paul.* This intonation pattern corresponds to the first interpretation listed above. If people hear the sentence produced this way, they are likely to interpret it to mean that either Jack and Paul, or Mary, will be invited. On the other hand, if the sentence is produced with a prosodic break after Jack, as illustrated in (8c), listeners are more likely to interpret the sentence to mean that Jack and either Paul or Mary will be invited.

(8) a. I will invite Jack and Paul or Mary.
 b. [I will invite Jack and Paul] [or Mary.]
 c. [I will invite Jack] [and Paul or Mary.]

The influence of intonation on ambiguity resolution helps explain why we rarely notice ambiguities even though they occur all the time in conversations. This is not only the case for global ambiguities. Sentences that might be garden path sentences if they were written do not frequently lead people down the garden path when they are spoken, because a speaker's intonation influences the listener's syntactic parsing process, determining the interpretation that will be chosen before he or she can be misled. For example, if a speaker said the sentence *While Mary was mending the sock fell off her lap* with a prosodic break after *mending,* as in *[While Mary was mending] [the sock fell off her lap],* the listener would choose the correct parse to begin with and would not be led down the garden path.

However, it should be mentioned that not all ambiguous sentences can be disambiguated through intonation. For example, there are no consistent intonation patterns corresponding to the two interpretations of the sentence *Flying planes can be dangerous,* which can mean 'Planes that are flying can be dangerous' or 'The action of flying a plane can be dangerous.'

In this file we saw that one of the major issues arising in sentence processing is structural ambiguity. Temporary structural ambiguity is constantly present in everyday discourse. Yet we deal with it effortlessly and usually don't even notice the ambiguity. Even when we are led down the garden path, we can usually recover the correct parse of a sentence rather easily. Globally ambiguous sentences aren't any different: we frequently don't notice the ambiguity and are able to decide on a syntactic parse seemingly effortlessly. This is possible because the context of a sentence, common sense, and intonation can help determine the correct parse of an ambiguous sentence.

FILE 9.7

Experimental Methods in Psycholinguistics

9.7.1 Some General Issues

Other than phonetics, psycholinguistics is probably the area in linguistics that is most experimentally oriented. Files 2.2, 2.3, and 2.6 introduced some of the methods used in experimental phonetics. In this file we introduce selected experimental methods used in the area of psycholinguistics. In particular, we describe some common techniques that are used to investigate particular linguistic phenomena. Both fMRI and ERP, introduced below in Section 9.7.2, directly measure brain activity. The methods introduced in later sections are less direct: they allow us to draw conclusions about processing activity by studying participants' behavior: measuring their response times, response types, and so on.

Before talking about methods, however, we should discuss some general issues that arise in experimental research. First, an experiment needs to be well thought through: a researcher needs to find a task or experimental protocol that will actually address her research question. After selecting a task, the researcher needs to assemble appropriate materials, which, in psycholinguistics, usually consist of words, sentences, and/or pictures presented to subjects either visually or auditorily. In many experiments the materials are designed to trigger some sort of linguistic response in the participants, or the participants are asked to perform a particular task upon being presented with the materials. Two kinds of materials are required. In addition to the experimental stimuli (those words, sentences, and/or pictures that the researcher is interested in), filler items are needed. These are other materials interspersed with the experimental stimuli, used to prevent participants from guessing which aspect of the stimuli the researcher is interested in. Finally, in order to be able to generalize findings, an experimenter should have gathered a large enough number of participants to generate statistically significant data—it's impossible, for example, to draw conclusions about "what speakers of English do" if only three people have been tested.

9.7.2 Measuring Activity in the Brain: ERP and fMRI

Two common methods used to study activity in the brain are **ERP** (event-related potentials) and **fMRI** (functional magnetic resonance imaging).

fMRI is a technique for determining which physical sensations or activities activate which parts of the brain. Brain activity is investigated by scanning the brain every 1 to 5 seconds. The scan maps areas of increased blood flow in the brain, which can be related to increased brain activity because active nerve cells consume oxygen, in turn increasing the blood flow to that region. Participants in an fMRI experiment cannot move, so the tasks of an fMRI study are somewhat restricted. However, participants can be played auditory stimuli, shown visual stimuli, or told to think about something. For example, bilingual participants can be told to think in one of their languages and then in the other to determine whether the same areas of the brain are used for both languages.

ERP is a technique that uses electrodes placed on the scalp to detect electrical signals in the brain. Unlike fMRI, ERP can be used to study the time course of an event, because it detects changes in electrical activity in the brain at the millisecond level. ERP analysis refers to certain patterns of electrical activity, usually a positive or a negative peak. For example, many studies have found a negative peak around 400 milliseconds after the presentation of an unexpected linguistic stimulus. This is called an N400. Since it occurs after sentences containing unexpected words, it is interpreted as the participant trying to integrate the unexpected word into the sentence context.

9.7.3 Offline Tasks

An **offline task** measures the final result of a process, rather than what happens during the process. The following paragraphs introduce some offline tasks that are often used to study language processing.

a. Lexical Processing. Some common tasks used in the study of lexical processing are **lexical decision tasks** and naming tasks. In lexical decision experiments, a participant is asked to identify stimuli as words or nonwords, and the time that it takes the participant to make a decision is measured. Lexical decision experiments have found, for example, that more-frequent words are recognized faster than are less-frequent words. **Naming tasks** are similar to lexical decision tasks, but instead of deciding whether a stimulus is a word or not, the participant responds by saying the stimulus aloud. A frequency effect is also found in naming tasks: more-frequent words are produced more quickly than are less-frequent words.

Both tasks are often combined with techniques such as **priming.** In priming tasks, participants are presented with a stimulus, the **prime,** right before the stimulus of interest, the **target.** Priming is often used to study the structure of the mental lexicon. For example, studies have shown that participants are faster to confirm that a stimulus is a word when the prime is semantically related to the target. This means that participants will be faster at confirming that *nurse* is a word when the prime is the semantically related word *doctor* than when the prime is the unrelated word *butter.* From this we can infer that the mental lexicon is partially organized by semantic relatedness. The idea then is that the prime *doctor* partially activated words semantically related to it, such that *nurse* was already partially activated when the target word appeared.

b. Sentence Processing. One common task used in the study of sentence processing is an end-of-sentence comprehension task, used to study globally ambiguous sentences. The procedure is very easy: participants read ambiguous sentences and answer a comprehension question after they have read the sentence. For example, to address a question that we addressed in Section 9.6.2, participants might be asked to read the sentence *The cop saw the spy with the binoculars* and answer the comprehension question *Who had the binoculars?* The answer to this question can tell the researcher how participants interpreted the ambiguous sentence.

9.7.4 Online Tasks

An **online task** is designed to uncover what happens during a process and when during the process it happens. Both fMRI and ERP, introduced above, are considered online tasks. The following paragraphs introduce some online tasks commonly used to study language processing indirectly.

a. Lexical Processing. One recent development in the study of online lexical processing is **eye-tracking.** In eye-tracking experiments, an infrared light is reflected off the participants' eyes and used to record all eye movements the participants make during an experiment: by tracking participants' eye movements, the researcher is able to determine what they were looking at at any point during the experiment. So, for example, the participants could be looking at an array of pictures including a caterpillar, a ham, a hamster, a hamburger, a bone, and a bobcat. Participants then hear an auditory stimulus saying, "Now look at the ______," for example, "Now look at the hamster." The eye-tracking device records when during the auditory stimulus the subject looks to the correct object, providing clues about how quickly words are processed as they come into their brains. For example, how quickly can the participant distinguish *ham* from *hamster* from *hamburger*? Do they have to wait until the end of the word, or can they begin to decide what the word must be from phonetic cues that appear much earlier in the word?

b. Sentence Processing. The end-of-sentence comprehension task described above gives us information only about how a person ends up interpreting an ambiguous sentence. But participants are often given as much time as they need to answer a comprehension question. As a result, while conscious decision making may be involved in answering the question, that approach cannot tell us what happened during the actual reading and processing of the sentence. Recall our earlier example sentence: *The cop saw the spy with the binoculars.* If, in response to the question *Who had the binoculars?*, a participant answered that the cop had the binoculars, we would not know whether this was the participant's initial interpretation or whether he had originally considered more options before settling on this meaning. It's possible that a participant initially used parsing strategies such as late closure while reading, but later decided that cops are more likely to have binoculars and therefore changed his interpretation. To find out what happened during reading itself, a task called **self-paced reading** can be used. In self-paced reading, participants read a sentence in small chunks, usually one word at a time. Whenever they have read and understood a word, they push a button to move on to the next word.

For this task, temporarily ambiguous sentences are used. Compare the sentences in (1).

(1) a. Someone shot the servants of the actress who was standing on the balcony.
    b. Someone shot the servants of the actress who were standing on the balcony.

In (1a), it is the actress who is standing on the balcony, while in (1b), the servants are standing on the balcony. For these two sentences, there is temporary ambiguity up until the point at which participants read *who*, because whatever follows *who* could modify the servants (i.e., *the servants who did something*) or the actress (i.e., *the actress who did something*). The choice of *was* or *were* as the next word disambiguates the sentence: it tells us what *who* modifies because rules about verb agreement say that *was* must go with *actress* and *were* must go with *servants*. What we are interested in is how long it takes participants to read the word *was* versus the word *were*, because this will tell us something about how much processing is required in order to get whichever interpretation the participant is presented with. To get this information, we measure the time it takes a participant to read each word (the time between button pushes). If participants thought that *who* modified the **servants** while they were reading the word *who*, they should take longer to read a following *was* than a following *were*. On the other hand, if participants thought that *who* modified the **actress** while they were reading the word *who*, they should take longer to read a following *were* than a following *was*. The reason is that participants would have to change their analysis if their initial interpretation of what *who* modifies turned out to be incorrect, and changing the analysis takes time. Thus, unlike the end-of-sentence comprehension task, self-paced reading allows us to see what happens during the reading of a sentence. Clearly, however, both types of task are needed to give us a more complete picture of sentence processing.