7

Word Recognition

In Chapter 5, we explored the ways in which babies and young children learn to pluck word chunks out of the streams of speech that they hear, pair these chunks up with meanings, and add to a growing collection of words that will eventually tally up to tens of thousands, if not more. But learning words and their meanings is only part of the story. Once we've added these thousands of words to our memory's stash of vocabulary units, we also need to be able to quickly and efficiently retrieve them while producing and understanding language.

Both speaking and comprehending language involve matching up meanings with their corresponding sound sequences at breakneck speed while at the same time juggling other complicated information. During a normal conversation, speech unfurls at a rate of about 150 words per minute; master debaters and auctioneers can reach speeds of 400–500 words per minute. This means that while listening to your conversational partner speaking at a typical leisurely rate, you have less than half a second to match up the sounds coming out of her mouth and rummage around in your voluminous vocabulary to find exactly the right word before the next one follows hot on its heels, all while working out how that word figures in the syntactic structure and meaning of the unfolding sentence. Most of the time you're doing all this while also planning your own response or interjection.

Perhaps it's not entirely surprising that, as a species that depends on transmitting information via language, we've evolved to have an efficient set of routines that allow us to pack and unpack large volumes of spoken information in very small spans of time. Presumably, we've had many generations—likely more than 100,000 years' worth—over which to develop the brains for it, and a knack for speedy language comprehension would seem to offer certain evolutionary advantages. On the other hand, *written* language is a fairly recent human innovation, with the oldest known writing system dating back a mere 5,000 years, to about 3200 B.C. Even today, many (hearing) societies in the world get by without a written language, though no society is speechless. Nevertheless, our ability to strip meaning from symbols on a page is at least as fast as our ability to do so with spoken sounds. The pleasure reading of skilled readers proceeds at a clip of about 200–400 words per minute—roughly 3–6 words per *second*.

The recognition of spoken words and the reading of written language offer quite different challenges from the standpoint of information processing. In speech, sounds have to be correctly identified, word boundaries have to be accurately located, and a string of sounds has to be mapped onto the correct word, despite the fact that many other words might offer *almost* as good a match as the right one. In writing, a jumble of arbitrary symbols has to be mapped onto words that have usually been learned in terms of their sound, not in terms of their visual symbols. The system of visual symbols represents a whole new set of symbolic units and mappings that's been artificially grafted onto the "natural" system of spoken language. Skilled reading relies on smoothly integrating this artificial system with the "natural" one during normal language processing—perhaps it's a little bit like the linguistic equivalent of learning to use a prosthetic limb.

In the previous paragraphs, I've been a bit preoccupied with the *time* it takes to recognize words. As it happens, time is the central obsession of the researcher who studies word recognition or, more generally, language processing. This preoccupation with time goes far beyond mere trivia, or even the desire to explain how it is that we process language as quickly as we do. Specific theories of language processing stand or fall based on the predictions they make about the relative timing of certain processing events. Time also serves as one of the most important methodological tools for studying how language processing works. As you'll see, researchers have come to learn a great deal about *how* linguistic information is processed by looking in minute detail at *how long* people take to process it, and by making careful comparisons of the time it takes to process stimuli that differ along specific dimensions. We owe much of what we know about word recognition to clever and meticulous forms of timekeeping. In this chapter, you can get a small taste of what it's like to do research on word recognition by taking part in the many Web Activities that are sprinkled throughout. A number of these activities will focus your attention on the precision and attention to detail that's needed in order to construct experiments where the results often hinge on finding differences of less than a tenth of a second of processing time.

7.1 A Connected Lexicon

Word webs

What happens in your mind when you hear or read a word? In Chapter 1, I emphasized the fact that our subjective intuitions about language often miss some critical information, and as you'll see, that's certainly the case when it comes to understanding something as basic as word recognition. Let's start by trying to describe the very simplest case of recognizing a word in isolation, outside of the context of a sentence. Based on how the experience *feels*, we might describe the process as a bit like this: Retrieving words from memory is like getting words out of a vending machine—let's think of word representations as the snacks you're trying to buy. Specific sequences of sounds you hear or letters you read are like the sequences of letters and numbers you have to punch into the machine to get the right product to come out. Just as a vending machine has a program set up to link a sequence of button presses with a specific location that delivers your chosen snack to you, letters or sounds are programmed in your mind to activate a specific word representation. In both cases, what you want conveniently drops down for your use in the form of a single, packaged unit. When it's described this way, there doesn't seem to be much to the process of retrieving words; it's simply a matter of punching in the right input in terms of sounds or letters, which is then linked in a one-to-one fashion with the right word representation. Why *shouldn't* this happen very quickly in real time?

The truth is that this subjective impression of word recognition is deeply wrong. In actual fact, recognition of either spoken or written words is quite a bit messier than this. It seems that words aren't organized in our minds independently of one another, but rather, are connected together in complex webs. When we try to retrieve one word, we end up pulling a string that has the actual matching word, but also has a bunch of connected words dangling from it as well. But since we have the *impression* that we've pulled a single word out of our mind, what is it that has led language researchers to the conclusion that words are actually highly interconnected? To get there, they've had to find ways to probe aspects of the word recognition process that may not be completely accessible to conscious intuition.

Evidence for partial retrieval of related words

One of the strongest (and earliest) sources of evidence for the interconnections among words has come from the phenomenon of **semantic priming**, which suggests that when you hear or read a word, you also partially activate other words that are related in meaning. The relevant experiments (the first of which were performed by David Meyer and Roger Schvaneveldt, 1971) typically use a method known as the **lexical decision task**, in which participants read strings of letters on a screen that are either actual words (for example, *doctor*) or nonsense words (*domter*) and then press one of two buttons—one button if they think they've seen a real word, and another to signal that the letters formed a nonsense word. The speed with which subjects press each button is recorded.

Semantic priming experiments have shown that participants are faster to recognize a real word if it follows hard on the heels of a word that's related in meaning. For example, responses to the test word *doctor* would be speedier if that word occurred in the experiment just after the word *nurse* than if it followed an unrelated word like *butter*. This suggests that the word *nurse* wasn't accessed in isolation. If you think of word recognition as "lighting up" a word in your mind, then some of the light spills over onto neighboring words in the semantic space. This partial lighting up of *doctor* following *nurse* makes it easier to recognize *doctor* when you see that word on the screen.

The lexical decision task shows that this happens with written words. But spoken language also seems to activate related words in much the same way. One way to see quite vividly what's happening as people process spoken words is to track their eye movements to a visual scene in real time as people hear instructions related to that scene. For example, imagine seeing a screen showing pictures of, say, a hammer, a nail, a cricket, and a box of tissues, and then hearing an instruction to "click on the hammer." Seems like a trivial enough task. But a record of subjects' eye movements (see **Figure 7.1**; Yee & Sedivy, 2006) shows that as they're hearing *hammer,*

WEB ACTIVITY 7.1

Interconnected words In this activity, you'll explore how to ascertain which words might be linked together in memory, through a pen-and-paper word association test in which people are asked to list the first words that come to mind when they encounter a word.

semantic priming The phenomenon by which hearing or reading a word partially activates other words that are related in meaning to that word, making the related words easier to recognize in subsequent encounters.

lexical decision task An experimental task in which participants read strings of letters on a screen that might either be actual words (*doctor*) or nonsense words (*domter*). Subjects press one button if they think they've seen a real word, or a different button to signal that the letters formed a nonsense word. Response times for real words are taken as a general measure of the ease of recognizing those words under specific experimental conditions.

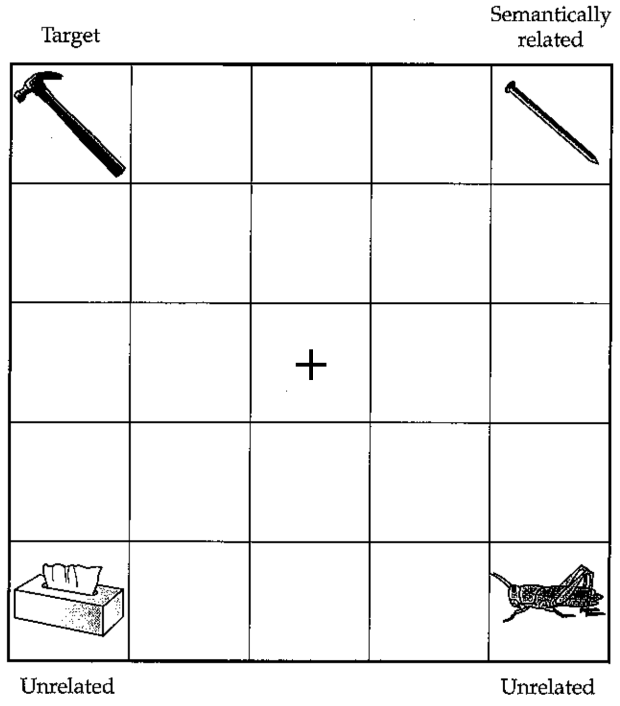


Figure 7.1 A sample display from an experiment by Yee and Sedivy (2006). Shortly after hearing the word *hammer*, subjects were more likely to look briefly at the nail than at unrelated items such as the cricket. This suggests that hearing a word results in the activation of semantically related words as well as the target word.

they might be briefly lured by a picture of a related word—that is, they're more likely to look at the nail than at the unrelated cricket. This aligns nicely with the evidence from lexical decision tasks, suggesting that words that are semantically related to the target word get an extra burst of activation, relative to other words in the lexicon.

Competition from partially activated words

The lexical decision experiments show evidence of **facilitation**: a word is made easier to recognize by having it occur after a semantically related one. But as you might imagine, if the process of retrieving a word partly lights up several other words, this could sometimes make it more *difficult* to pick out the correct word from among the other, partially lit-up ones. That is, related words that are partially activated could get in the way of recognizing the right one, a result that is referred to as **inhibition**. The experimental record shows evidence of this too, especially when it comes to words that are similar in *form*, rather than meaning. For example, if "primed" with the word *stiff*, people might be slower to subsequently recognize *still* as a word (Slowiaczek & Hamburger, 1992). In this case, the prime word seems to *compete* with the target word, tripping up the recognition process. Again, the data from the priming paradigm is corroborated by eye-tracking experiments; if the visual array includes a beetle as well as a beaker, for instance, recognition of the spoken word *beaker* tends to be slowed down, so it takes people longer to locate the image of the beaker in the display (Allopenna et al., 1998).

Competition among lexical items in memory can be seen easily when the form of a target word is similar to other words that are relevant in the immediate context; for example, in the lexical decision priming paradigm, a word might inhibit the access of a similar one immediately following it, or in the eye-tracking paradigm, locating the target referent for a word may be slowed down because the eye is drawn to an object corresponding to a similar-sounding word. But these are somewhat contrived experimental scenarios in which similar words have been deliberately planted. Most of the time when we're retrieving words from memory in real-life conversation, we don't have to cope with such blatant decoys—for example, how often would we be in a context where both *beetle* and *beaker* would be uttered? So, to what extent is lexical competition likely to really play a role in our routine language comprehension?

Quite a bit, it would seem. Competition among words happens even when there are no similar words in the immediate context—it's enough merely for there to be similar words in your own personal mental storehouse. The evidence comes from comparing how quickly people retrieve words that have either many or few sound-alikes in the general lexicon, even when those sound-alike words aren't prominent in the context. As it happens, some words that you know are relatively unique when it comes to their sound structure. For instance, take the word *stench*: try to come up with as many other words as you can that differ from that one by only one sound. After *staunch* and the more uncommon word *stanch* (as in: *stanch the flow of blood*), I pretty much draw a blank myself. But I can reel off quite a few examples that differ by only one sound from the word *sling*: *sting, fling, bling, cling, slung, slang, slim* (remember, *-ng* counts as one *sound*), *slit, slip, slid, slick*. Psycholinguists have invoked a real-estate metaphor to describe this difference between *stench* and *sling*; we say that *stench* is found in a "sparse neighborhood," with very few sound-based neighbors, while *sling* resides in a "dense neighborhood," with its neighbors crowding up on all sides.

facilitation Processes that make it easier for word recognition to be completed.

inhibition Processes that result in word recognition becoming more difficult.

Again, psycholinguists have used response times in lexical decision tasks to probe for **neighborhood density effects**. All things being equal, after the length and frequency of words has been controlled for (since these factors can also affect how long it takes to recognize a word), people take longer to recognize words that come from dense neighborhoods than words that come from sparse neighborhoods (e.g., Goldinger et al., 1989). This effect can be seen even though no similar words have occurred before the test word, suggesting that lexical neighbors can compete for recognition even when they are not present in the immediate context. It's enough that they're simply part of the vocabulary.

WEB ACTIVITY 7.2

Neighborhood density In this activity, you'll generate words that belong to dense and sparse neighborhoods, and you'll consider other variables that threaten to mask the effect of neighborhood density if not properly controlled.

neighborhood density effects Experimental results demonstrating that it is more difficult and time-consuming to retrieve a word from memory if the word bears a strong phonological resemblance to many other words in the vocabulary than if resembles only a few other words.

Building a model of word recognition

Vague metaphors involving spotlights or neighborhoods are useful for getting an intuitive sense of how words in the mental lexicon affect each other in the process of word recognition. But it's also worth moving toward more precise theories of what lies beneath these effects. Making explicit models can advance a field very quickly, whether they're communicated as simple diagrams or actually implemented as working computer programs that simulate real language processes. A model not only serves as a way of explaining current experimental results, but also tends to force researchers to grapple with other perfectly reasonable ways to account for the same data. When you have to make a decision about the details of your model, you become aware of the things you don't yet know. So the process of building models is really useful for throwing light onto as-yet-unanswered questions that researchers might otherwise not think of. One of the main benefits of a model is that it makes new predictions that can be tested experimentally.

There's probably no other psycholinguistic process that's been modeled in as much detail and by so many different rivals as word recognition. These modeling efforts have led to an enormous volume of published experimental papers dealing with word recognition. There isn't room in this chapter to give a thorough overview of all of the models that have been proposed, or the experimental evidence that has been generated. Instead, my goal will be to bring out some of the key findings and big questions about how words are processed in the mind, and to relate these to the choices that go into building a good working model.

Let's start with one simple way to model how related words affect each other in the process of word recognition, as sketched out in **Figure 7.2**. Here, rather than just being listed as isolated units, words are represented as belonging to a complex network. Words that are related in meaning are linked together in the network—for example, *lion* is linked to *mane* and *tiger*, and *tiger* is linked to *stripes*. (Links could also be established between words that frequently occur together in speech, whether or not their meanings are similar—for example, between *birthday* and *cake*, as illustrated by the word-association task in Web Activity 7.1). If a match is made between the sounds (or letters) being perceived and a particular word, that word gets a surge of ener-

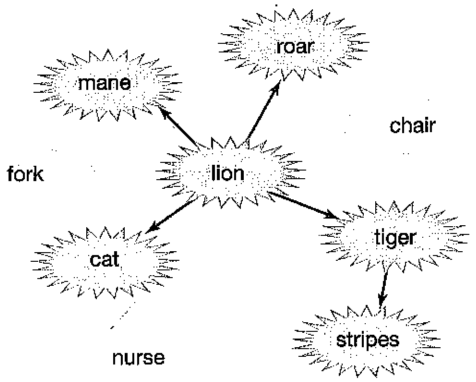


Figure 7.2 A simple spreading-activation network. Hearing the word *lion* activates the mental representation of that word and, in turn, the representations of related words such as *mane*, *cat*, *tiger*, and eventually *stripes* (via *tiger*).

mediated semantic priming The process by which a prime word (e.g., *lion*) speeds up responses to a target word (e.g., *stripes*) not because of a direct connection between *lion* and *stripes*, but due to an indirect connection via some other intervening word (e.g., *tiger*).

gy or activation. Since the words in the lexicon aren't isolated from one another but share links with quite a few other words, activation "flows" throughout the network along these links. For example, hearing or reading the word *lion* will also excite the connected words *mane* and *tiger*. This has behavioral consequences, such as the priming effects we've just seen.

Setting words up in a network like this (we've just created a mini-model!) results in a new prediction: Since each word is connected to a number of other words, and energy spreads along links in the network, we should see signs of activation traveling along paths of interconnected words. That is, *lion* should activate not just *tiger*, but also the word *stripes*. The latter word isn't related to *lion*, but it *is* related to *tiger*, which is itself linked to the perceived word. In fact, there is evidence for just this kind of **mediated semantic priming**, in which a prime word like *lion* speeds up responses to a target word like *stripes*, even though the prime and target are only related via some other intervening word.

Our simple spreading-activation story looks promising, but it's going to need some adjustments and refinements. So far, there's nothing to prevent activation from continuing to spread throughout the network from link to link to link in such a way that it eventually activates just about every word in the network. For example: *lion* → *tiger* → *stripes* → *paisley* → *shirt* → *tie* → *neck* → *head* → *hair*, and so on. The model needs to have a way to prevent such overwhelming buzzing within the lexical network.

One way to achieve this is by building in a *decay function* by which activation levels gradually die down over time. This means that the activation of *lion* would surge at first, but then fade. Since it would take some time for activation to spread throughout the network, the decay function would limit how much energy is passed on to remotely connected words or concepts. So, the activation of *lion* would have dwindled somewhat by the time it spread to *tiger*, meaning that *tiger* would be less activated than *lion* was initially. As a result, *stripes* would receive less activation than *tiger*, and any activation that *stripes* passed on to *paisley* might well be negligible by that point. With this modification, it should now be possible to quantitatively tinker with the specific rates at which activation spreads and decays so that the model will closely simulate the patterns of results from experiments with humans. For example, the model should now also be able to capture the fact that the degree to which a related prime speeds up responses to a target word depends not only on how closely related the two words are, but also on how much time has elapsed between the presentation of the prime and target words.

So far, we have a model that does a nice job of capturing *facilitatory* priming effects and their limits. But what about the cases in which the presence of related words *slows down* or impedes the recognition of the target? As it stands, the model doesn't predict these competition effects, so we need to either scrap it or add something else. Given that we already have some mileage from our little model in accounting for priming effects, the latter strategy seems like a good place to start.

Competition effects are seen most clearly among words that are related to each other in form rather than meaning, so we might focus on the process of mapping sounds or letters to word representations. This is depicted in **Figure 7.3A**, where phonemic (or orthographic) units are connected to word representations. When a sound (or letter) is identified, it becomes activated, and by virtue of the connections it has to words, activation flows to those words that contain it. So far, this predicts that words that contain overlapping sounds (for example, *beam* and *beat*) should both become active. But there's nothing in the model yet to explain why hearing *beam* should make it harder to subsequently recognize *beat*—in fact, as it stands, the model predicts the opposite, since connections allow activation to spread. We can get the right result if we propose that two kinds of connections can exist between representations that share a link: **excitatory connections**, which pass activation from one unit to another, and **inhibitory connections**, which have exactly the opposite effect so that the more active a unit is, the more it *suppresses* the activation of a unit it is linked to. You can see this in the revised model in **Figure 7.3B**; in this version of the model, once the last sound or letter of *beam* is perceived, it will excite the word *beam*. But *beat* will be inhibited: as the activation of the unit *m* rises, it will suppress the word *beat* through the inhibitory link between *m* and *beat*; and *beat* will also become inhibited by virtue of the inhibitory link from the word *beam* to *beat*.

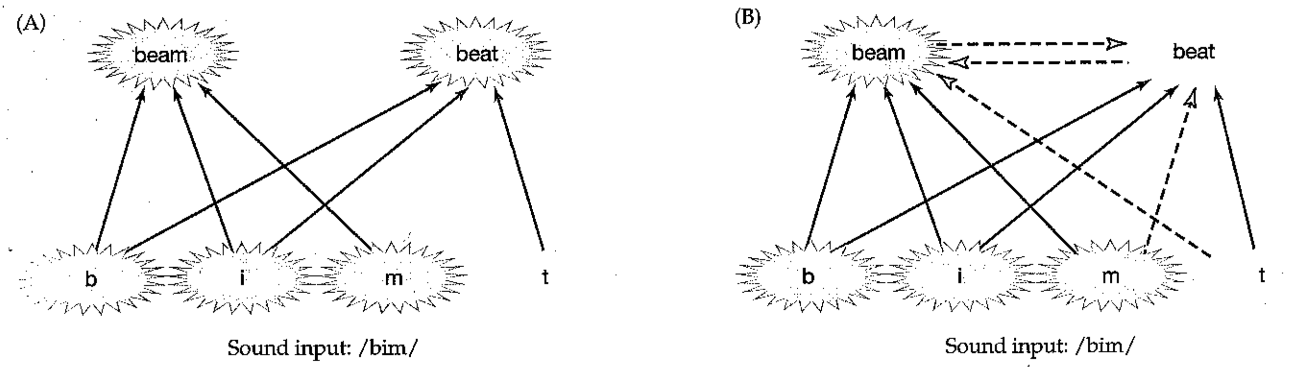


Figure 7.3 (A) A simple model showing only excitatory connections (solid lines and arrows) from the phonemic units in the words *beam* and *beat* (/bim/ and /bit/ respectively). With only excitatory connections, there's nothing to explain why hearing the word *beam* should make it harder to subsequently recognize *beat*. (B) A model with inhibitory connections (red dashed lines) in addition to excitatory links (solid lines and arrows). As the activation of a phonemic or word unit rises, activation is decreased for units that are connected to it via inhibitory links. For example, the rise in activation for *beam* results in the suppression of activation of *beat*.

excitatory connections Connections along which activation is passed from one unit to another, so that the more active a unit becomes, the more it increases the activation of a unit it is linked to.

inhibitory connections Connections that lower the activation of connected units, so that the more active a unit becomes, the more it *suppresses* the activation of a unit it is linked to.

As you've seen, a great deal of the evidence for the degree of activation of word representations comes from either eye movement studies or experiments using the lexical decision task. In many cases, researchers strive to test predictions that involve very subtle differences in activation between types of stimuli, or changes in activation levels for the same stimuli over time. Hence, the experimental methods that they use need to be deployed with great precision. **Method 7.1** lays out some of the challenges that come up in using the lexical decision task, and the techniques that researchers use to get the most out of this simple task.

WEB ACTIVITY 7.3

Tinkering with models and making predictions In this activity, you'll try your hand at making empirical predictions based on subtle variations in the details of the model you've just seen.

Probing the model's assumptions

Our model-building exercise so far has focused on two aspects of the model: (1) the existence of links that connect representations to one another, and (2) what happens along these links (that is, do they spread activation or suppress it, and how do these effects dissipate over time?). But our little model also makes certain implicit commitments that we haven't yet defended empirically.

We've assumed that the input to word recognition is a set of discrete letter or sound units, represented as integral units rather than as bundles that

METHOD 7.1

Using the lexical decision task

There's a simple logic behind the semantic priming techniques that are used in lexical decision tasks. You create an experiment in which you record and compare the response times to test words, or targets, that have been preceded by primes: words that are either *related* or *unrelated* to the target. For example, response times to the target *doctor* preceded by the prime *nurse* would be compared with results for *doctor* preceded by *chair*. If responses to the former are significantly faster than to the latter, you've found a priming effect, suggesting that seeing the word *nurse* facilitated the subject's recognition of the word *doctor* through spreading activation. But actually running an experiment using priming techniques involves making many small technical decisions, each of which could conceivably have some impact on your results or how you interpret them.

First, you'll need to deal with the possibility that participants might try to anticipate patterns within the experiment in order to respond strategically. If this happens, their response times might say less about what people typically do in daily conversation than about what students do when trying to puzzle out a psychology experiment. The goal is to minimize patterns wherever possible. For example, you can easily eliminate the expectation that the correct response to the target will be to press the button for "Yes, it's a word" by making sure that you put in plenty of fillers, and balance the experiment so that the likelihood of the target being a real word is exactly 50% over the course of the experiment.

But some patterns are impossible to eliminate outright, such as the relationship that exists between some of the primes and targets. If, after a while, your participants begin to catch on to the fact that the target is often related to the prime, they might approach the task less as a word recognition task, and more as a word association task: once they see the prime *nurse*, for example, they might start actively thinking of related words. In that case, it would be unsurprising to find that there's a close relationship between word association tests and priming patterns (see Web Activity 7.1); but you'd be less confident in concluding that related words are *spontaneously* activated during routine word recognition out in the wild.

You can rely on two approaches to reduce the possibility that participants will consciously notice the relationship between primes and targets. The first is to simply include a great number of filler items in which the prime and target aren't related in any way, making it harder to detect the pattern. For example, it would be harder to notice the relationships between words in this set of stimuli:

FREEDOM—METAL
WRENCH—BOOK
HANDLE—SHOES
NURSE—DOCTOR
FLOWER—SCREEN
PAPER—ROOF

than in this one:

FREEDOM—METAL
WRENCH—HAMMER
HANDLE—DOOR
NURSE—DOCTOR
FLOWER—VASE
PAPER—ROOF

METHOD 7.1 *(continued)*

The other way to reduce strategic responding is to provide your participants with as little time as possible to anticipate words that are related to the prime. You can do this by shrinking the time between the presentation of the prime and the presentation of the target. In fact, you can even exploit an intriguing quirk of the human perceptual system, namely, the fact that there's a time lag between conscious and unconscious perceptual processing, so at very rapid speeds, it's possible for our unconscious minds to have processed a stimulus while our conscious minds don't "know" that we've done so.

Researchers have used subliminal presentation of the prime word—in a paradigm known as **masked priming**—as a tool to eliminate the possibility of conscious, strategic responding. At the beginning of a trial, participants see a row of dashes or # symbols "masking" the prime. The prime is then briefly flashed and quickly covered up again before the presentation of the target stimulus. For example:

mask	#####	(1,000 milliseconds)
prime	NURSE	(50 ms)
mask	#####	(500 ms)
target	DOCTOR	(response required—is it a word?)

As we've seen from our discussion of modeling, the interconnected lexicon is in a constant state of flux, reflecting the spreading and dampening of activation over time along various links. This means that using the priming technique—which probes for responses at *one* specific time point—is a bit like relying on a snapshot to capture an object that is in motion. Depending on when you press the shutter, you may get a very different picture of the object's path of motion. For instance, if you're interested in finding out whether target words that have some particular relationship to their primes become activated above baseline levels (that is, relative to targets with unrelated primes), whether or not you find a priming effect may depend upon when you present the target. Probe too soon, and activation may not have spread to the target yet. Probe too late, and activation may have dissipated or become suppressed.

For this reason, researchers often sample at different time points, by varying either how long the prime is presented or the **interstimulus interval (ISI)**, which refers to the amount of time between the offset of the prime and the onset of the target.

Another issue with using a snapshot technique to capture a temporally dynamic process is that response times will be affected by how deeply people process the target words before making a decision. If they press the button after very shallow processing, they may not yet have had time to fully access the word's semantic representation. This can happen if the non-word targets are very distinct from any possible words (for example, *bgltx*, *aoitvb*), so decisions about the target's status can be made very quickly and on the basis of very superficial characteristics. On the other hand, if the non-words look like possible though non-existent words (for example, *blacket*, *snord*), then participants will have to process them more deeply before deciding whether the targets are words or non-words.

masked priming A priming task in which the prime word is presented subliminally, that is, too quickly to be consciously recognized.

interstimulus interval (ISI) The amount of time between the offset of the prime and the onset of the target.

contain other parts or properties. This doesn't seem like an unreasonable assumption for recognizing written words—after all, the letter *p* looks pretty much the same, regardless of which other letters it's next to (at least if it's typed or printed). So it doesn't seem odd to think of it as a basic unit. Even so, by treating letters as integral units, the model predicts—without justification—that similarities among letters should be irrelevant. For example, it predicts that the word pair *bin/din* should be no more confusable than *bin/sin*, which is probably wrong.

But things get especially complicated in dealing with the perception of *spoken* language. Individual sounds vary a great deal, depending not only on where in the word they appear, but also on the particular speaker's age, gender, regional dialect, and individual characteristics of the vocal tract. What's more, sounds often smear together when pronounced, making it tricky to carve up the speech stream into separate units. We'll take up these issues in Section 7.4.

Another built-in assumption that we've made so far is that word representations themselves are discrete units. Our model captures these as individual nodes that become activated or inhibited as entire units. I've been extremely vague about what's *in* a word representation. Let's assume that a word node is really just a container for information about a word's meaning and sound, as well as information specifying how the word combines syntactically with other words. But by representing words as nodes, our model implies that all of this information is available simultaneously and all to the same degree when the word node is activated. In fact, there's an ongoing debate about whether words

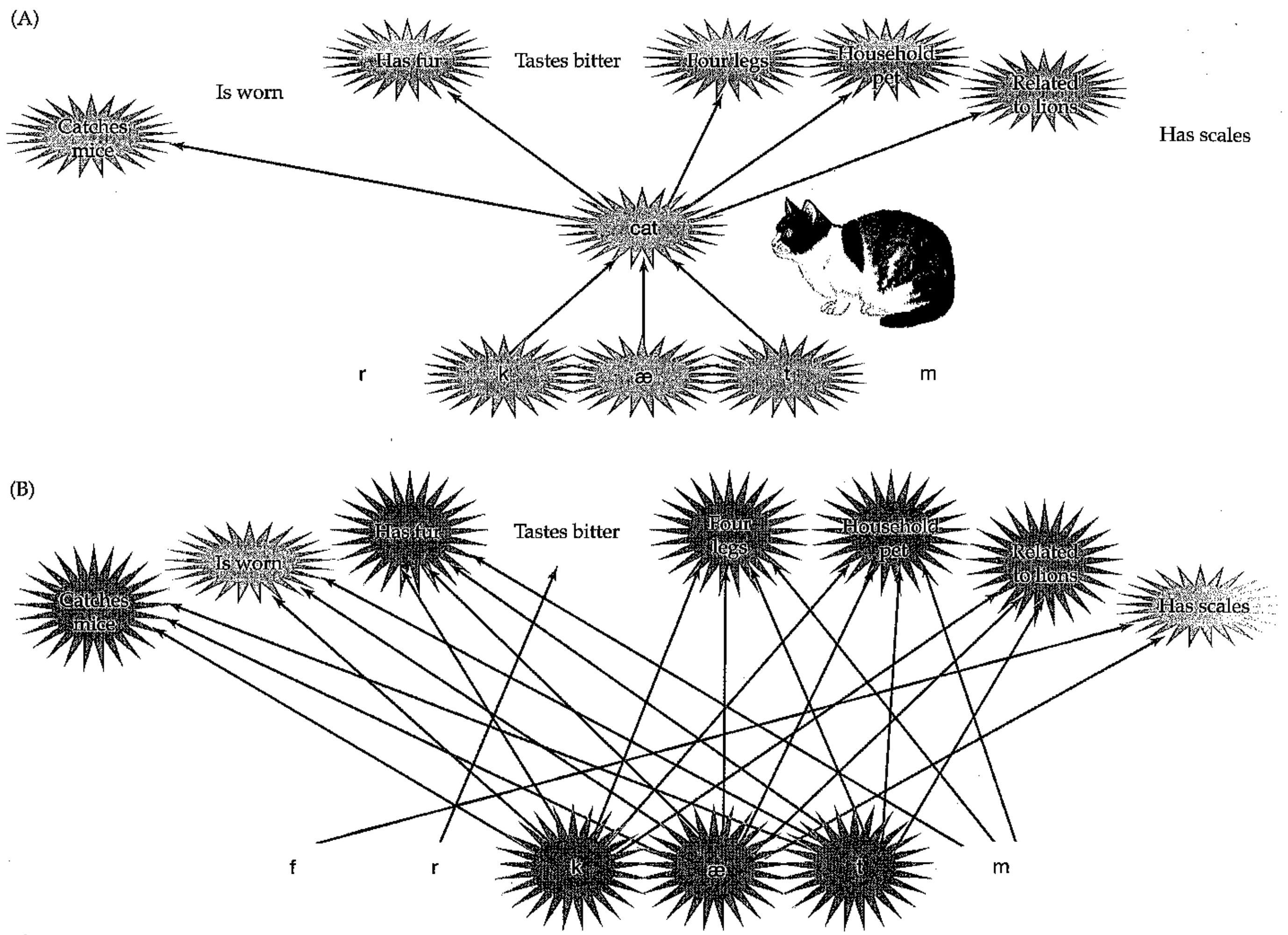


Figure 7.4 Localist versus distributed representations. (A) Localist word representations are shown by this example of a model in which the semantic features are connected to a single word unit (only excitatory connections are shown here). The activation of phonemic units results in the activation of the word unit and, in turn, the activation of that word's semantic features. (B) Distributed representations are shown in this example of a model in which phonemic units are directly linked to various semantic features without any intervening word units.

should be represented as discrete units (or containers) at all. An alternative view is that words should be captured as bundles of features instead. In dealing with the meanings of words, for example, the difference between *localist* word representations (with discrete nodes) and *distributed* representations (just bundles of features) is shown in **Figure 7.4**.

In the localist representation (Figure 7.4A), sound units connect to word nodes, which in turn connect to semantic features that become turned on when the word node is activated. In the distributed representation (Figure 7.4B), bundles of sound units connect directly to bundles of semantic features, without any intervening word nodes. The advantages and disadvantages of each approach, as far as generating correct predictions goes, are quite subtle, and I won't elaborate on them. But just to give you a taste for where they might differ, I'll point out that one argument that's been made for distributed representations is that they more accurately reflect some of the losses that might happen with brain damage. For example, when people suffer from aphasia due to a stroke, access to word representations is often impaired. But the damage doesn't seem to selectively wipe out swaths of a person's vocabulary, as might be expected if words were represented as integral units. Rather, the damage seems to lead to an across-the-board dysfunction in access, as if some *parts* of all word representations were destroyed, rather than some subset of word units in a person's vocabulary. (See **Box 7.1** for additional discussion on how to think about word meanings.)

BOX 7.1
Words: All in the mind, or in the body too?

So far, we've been talking about meanings of words as clusters of semantic features or properties that are linked to word representations and become "turned on" when a word is heard or read. For example, the word *dog* might be represented with properties such as [has fur], [barks], [has four legs], [is a mammal], and so on. The assumption is that we've learned and mentally stored some properties about the things in the world that are called *dog*. It's often thought that these properties are abstract—that is, the property [has fur] comes to be stored in the mind in the same way regardless of how we came to know this fact about dogs, whether through visual observation of the fur, by feeling it, or by being told that dogs have fur. In other words, we don't think about meanings in terms of pictures or tactile experiences; we think about them in *thoughts*. The perceptual experiences are just the delivery device for getting to the meanings; they're not part of the meaning representation itself. To take a slightly more provocative example: the idea is that even though the color blue is closer to green on the color spectrum than it is to red, the *meaning* of the word *blue* in our minds would be no closer to the meaning of *green* than to that of *red*, since word meanings are just abstractions.

But many researchers argue that links between words and our perceptual experiences are preserved, and some even go so far as to say that meaning isn't really about pulling abstract properties from those experiences so much as it is about encoding the perceptual memories and linking them to words and to each other.

A number of studies suggest that at the very least, there is cross talk between word representations and bodily information. For example, people are usually faster to respond to a word like *pen* or *knife* when their hands are positioned as they would be if they were actually using these objects (Klatzky et al., 1989). They are faster to respond to sentences like *He closed the drawer* if the response requires them to make a movement that's consistent with simulating the action—for example, if subjects have to move the hand away from the body to push a button rather than starting at a faraway resting point and moving the hand toward the body to push a button (Glenberg & Kaschak, 2002). Semantic priming shows that reading a word like *typewriter* speeds up responses to a word like *piano*, where the main similarity between the words is in how the corresponding objects are physically manipulated, rather than some more abstract property that they share (Myung et al., 2006). And people seem to take less time to read words when they can easily imagine ways in which they might physically interact with their corresponding objects, such as *cat* versus *sun* (Phillips et al., 2012).

But it's one thing to show that words set off resonances with corresponding perceptual memories, and quite another to show that knowing and retrieving the meanings of words somehow *depends* on accessing those perceptual memories, which is what more fervent proponents of embodied meanings want to claim. It would be useful to see what happens to word processing when there is damage to sensory or motor systems in the brain. One study by Véronique Boulenger and colleagues (2008) looked at people with Parkinson's disease, a condition that dampens the brain activity responsible for planning physical movement. Would such subjects show a selective impairment in accessing the meanings of words involving action? They did; compared with other subjects, those with Parkinson's showed reduced priming for action words (suggesting that these words were being weakly activated) while showing normal priming for words that don't evoke actions. But when these subjects were treated with medication that improves the brain functioning in motor areas, they showed normal priming for action words as well. This suggests that perhaps word meanings are not entirely abstract, and that some are woven tight with bodily action. Other classes of words seem to be more tightly linked with specific sensory domains. For instance, damage to the visual region in the brain often leads to difficulty processing words for things that we normally experience visually, such as *birds* (Warrington & Shallice, 1984).

Results like these are hard to explain with a story that says word meanings are made out of pure thought rather than out of pictures or patterns of movement. They're driving a rethinking about how to talk about the meanings of words. Of course, building meanings out of body memories poses some challenges of its own, the greatest of these being how to account for more abstract words like *freedom* and *hypothesis*, let alone *if*, *but*, or *not*.

WEB ACTIVITY 7.4

Bottom-up versus top-down links In this activity, you'll work through the implications of drawing links going only from the bottom up, from the sound level to the word level, in contrast to a model that allows a bidirectional flow of information between levels.

Finally, you might have noticed that in Figure 7.3, the links between the sound units and word units aren't bidirectional; they go in one direction only, from the lower to the higher level. It certainly makes sense to have activation flow from the sound level to the word level—after all, we recognize words mostly on the basis of their sounds. But we might question whether this is the *only* direction in which information can flow. In theory, our model could have been developed otherwise, with facilitation and inhibition going from the top down as well.

I'll take this issue up in later sections, after you've had a chance to think about the empirical consequences of each move, with the help of prompts in Web Activity 7.4.

7.2 Ambiguity

A multiplicity of meanings

In the preceding section, you learned that retrieving a word from memory is not a simple matter of pulling out a single word from its designated slot in the mind's vending machine. Instead, multiple word representations are simultaneously activated, resulting in competition among lexical representations, sometimes to a degree that causes a discernible delay in the time it takes to retrieve a word. Competition is most intense when words overlap a great deal in terms of their sounds or orthography (so that *bean,* for example, might interfere with the retrieval of *beat*). Which might lead us to ask: What about when the sound overlap between words is not partial, but *complete*? As it happens, English is riddled with **homophones,** words that mean completely different things and may be spelled differently, but that sound exactly the same. Here's just a small sample:

bred, bread
none, nun
bridal, bridle
sun, son
team, teem
made, maid
blew, blue
waste, waist
stare, stair
pea, pee
side, sighed
missed, mist
know, no
seen, scene
hour, our
flea, flee
main, mane
in, inn
fair, fare
retch, wretch

The problem of ambiguity gets worse. English is also rife with **homographs,** words that share the same spelling but have different meanings (and may or may not sound the same). Consider, for example:

The performer took a deep **bow.**
It's difficult to hunt with a **bow** and arrow.

Jerry is headed **down** the wrong road.
I've really been glad to have my **down** parka this winter.

Silvia is **content** with her lot in life.
The **content** of this course is difficult.

To round the list off, consider the many words that are **polysemous,** conveying a constellation of related but different meanings. Consider some the many possible uses of the word *run:*

She's got a **run** in her stockings.

There was a **run** on the banks this week.

homophones Two or more words that have separate, non-overlapping meanings but sound exactly the same (even though they may be spelled differently).

homographs Words that are spelled exactly the same but have separate, non-overlapping meanings (and may or may not sound the same).

polysemous words Words that can convey a constellation of related, but different meanings, such as the various related meanings of *paper,* which can, among other meanings, refer to a specific material, or a news outlet.

Sam went out for an early morning **run.**

I'd like to **run** my fingers through your hair.

Let's **run** through the various options.

He's had a **run** of bad luck.

Can you **run** this over to the post office?

You can try generating a similar list for words like *paper* or *dish*—you may find yourself startled at how many different meanings or uses you can come up with.

Ambiguity is so rampant in language (and not just in English) that you might begin to wonder whether it's a serious design flaw common to many languages. It's hard to imagine that the presence of ambiguity does anything useful to promote effective communication between people (but see **Box 7.2**).

BOX 7.2

Why do languages tolerate ambiguity?

If the goal of language is for a speaker to plant his intended meaning firmly and decisively into the mind of the hearer, ambiguity appears to be a serious flaw. By definition, an ambiguous word or phrase is compatible with multiple interpretations, not just the meaning the speaker intended. You'd think that languages would strive to avoid ambiguity. Yet all known languages seem to be rife with it, despite the fact that lexical ambiguity could very easily be avoided. In the words of blogger Geoff Pullum (2012):

> *Let me make a numerical point to begin with. The number of [possible letter sequences] with length not more than 10 over the roman letters a to z plus the apostrophe is* 27^{10} *= 205,891,132,094,649—about 200 trillion. The total number of words in the workaday word list is about 25,000. What I'm saying is that English could easily have a distinct letter sequence for every different meaning, using letter sequences much shorter than the present ones. It doesn't because the language in general shows no signs of being the slightest bit interested in that. English uses the same two-word phrase for denigrating, ceasing to hold, making notes, and euthanasia. [The phrase is* ***put down.****] It wantonly employs a single three-letter word for meanings relating to understanding, judging, experiencing, finding out, dating, visiting, ensuring, escorting, and saying farewell. [The word is* ***see****; see if you can create a sentence for each of these uses.] Nobody who thinks about English for a few seconds could possibly believe it shuns ambiguity. It doesn't give a monkey's fart about avoiding ambiguity.*

Steve Piantadosi and his colleagues (2012) have gone even further and argued that not only do languages not "care" about avoiding ambiguity, they actively seek it out because ambiguity actually makes a language more effective. The logic goes like this: Ambiguity rarely creates serious impediments to understanding—yes, processing ambiguous words comes with a small cost for the hearer (as you'll see in the rest of this chapter), and yes, occasionally, communication may rupture as a result. But the vast majority of the time, hearers are quite competent at relying on context to navigate through the various meanings offered up by a single string of sounds or letters. The benefits of ambiguity come from considering the costs of *producing* language. Speakers can minimize their effort by re-using bits of language that are common, short, and easy to pronounce, rather than resorting to longer words with unusual combinations of sounds. The idea is that languages tend to strike a balance between comprehensibility and ease of production. If ambiguity is managed fairly easily by the hearer, the speaker may as well take advantage of it to reduce his own cognitive workload.

Throughout this chapter, you'll see how the word recognition mechanism is set up to avoid ambiguity sinkholes, suggesting that Piantadosi and his colleagues are right that ambiguity doesn't do much damage to understanding. In support of their second point—that ambiguity makes the task of speaking easier—Piantadosi and his colleagues presented evidence from several languages showing that words that are easier to produce are exactly the ones that are most likely to be re-used for new meanings. That is, ambiguous words in those languages were generally shorter, more common, and composed of fewer unusual combinations of sounds than unambiguous words.

The English writer Virginia Woolf had an interesting perspective on ambiguity and the usefulness of language. In her 1937 essay on writing titled "Craftsmanship," Woolf argued that if we think of a useful statement as one that can mean only one thing—that is, a statement that unambiguously communicates a very specific idea—then it should be apparent "how very little natural gift words have for being useful. … They have so often proved that they hate being useful, that it is in their natures not to express one simple statement but a thousand possibilities." To make her point, Woolf suggested the reader imagine what's going on inside the mind upon hearing a simple and seemingly utilitarian phrase:

> Take the simple sentence "Passing Russell Square." That proved useless because besides the surface meaning, it contained so many sunken meanings. The word "passing" suggested the transiency of things, the passing of time, and the changes of human life. Then the word "Russell" suggested the rustling of leaves and the skirt on a polished floor also the ducal house of Bedford and half the history of England. Finally, the word "Square" brings in the sight, the shape of an actual square combined with some visual suggestion of the stark angularity of stucco. Thus one sentence of the simplest kind rouses the imagination, the memory, the eye and the ear—all combine in reading it.

For Woolf, the point of language is just that—to rouse the imagination, rather than communicate a specific idea. Poets (and maybe even advertising copywriters) would likely agree with that, and I myself would admit that it's certainly the aim of great *writing* to rouse the imagination. But I strongly doubt that whoever announced "Passing Russell Square" on the train did so with the intent of evoking the rustling of long skirts on polished floors. And I also suspect that your average passenger understood the announcer to be communicating a specific idea (even if thoughts of rustling skirts *did* enter his mind).

But Woolf's discussion of language, fanciful though it may be, is in some ways a plausible psycholinguistic hypothesis about word recognition. There is good reason to think that during the course of recognizing a single word, a plethora of meanings presents itself.

Look back at the model we built in Figure 7.3, and notice the connections between the units of sound and word representation units. So far, we've been assuming that word units are activated to the extent that the sound units they are connected to become activated; perfect matches will be activated the most, but highly similar words will also light up. But as Virginia Woolf notes, often a set of sounds will match up *exactly* with a number of different word meanings. For example, the sounds in the name *Russell* match up with the name of whoever Russell Square happens to be honoring, but also with many different Russells, as well as the homophonous word *rustle*. Presumably, the mental lexicon includes connections from this set of sounds to *all* of these word representations, so activation of the sound units should spread to all of these word representations and their respective semantic features. If that's the case—and if, as suggested by Virginia Woolf, words commonly have a multiplicity of meanings—then the real puzzle is: How is it that words manage to make themselves useful after all? In other words, how is it that we ultimately arrive at a single interpretation, despite the numerous possible meanings?

The easy answer to this question is that we undoubtedly use context to disambiguate meanings. But since we've been building detailed models, let's be a bit more precise: *How* do we use context? We might build up our model using two different approaches.

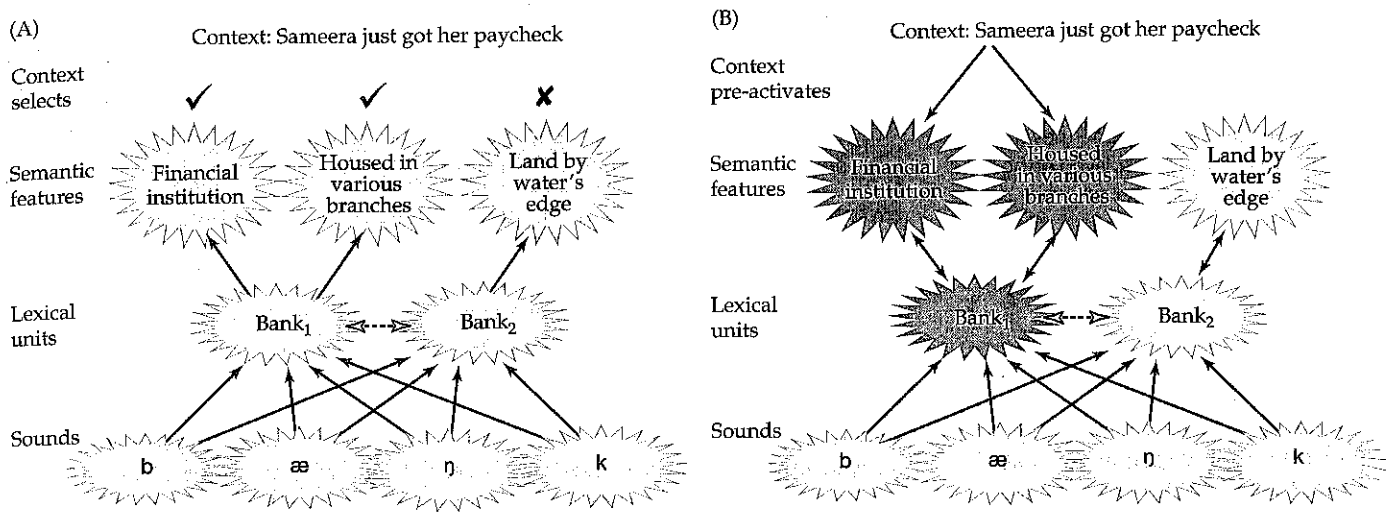


Figure 7.5 Two ways in which context can help word recognition. (A) Activation flows from the bottom up, from phonemic units to words and in turn to semantic features. Both meanings of the word *bank* are equally active until contextual information is recruited to select the most appropriate meaning. (B) Context can generate more expectation for some meanings than others by "pre-activating" some semantic features. Hence, by the time the word *bank* rolls off the tongue of the speaker, one of the meanings of that word is already more active than the other. Inhibitory connections appear in red.

The first is illustrated in **Figure 7.5A**, using two meanings of the word *bank*. Here, sound units are activated as these sounds are heard. Activation flows from the sound level to the lexical level, activating both word representations and, in turn, their associated semantic features. The flow of activation is in one direction only, from the lower level of sound to the higher level of meaning. (Lateral inhibitory links have also been drawn in between word representations, but since each word representation receives equal activation, these links would have no impact on the relative activation of the two possible meanings.) Once words and their meanings are activated, a separate decision mechanism is triggered to select the most contextually appropriate word, based on a good match between the semantic features of that word and the semantic expectations that have been set up by the context. To relate this to Virginia Woolf's observations, this would mean that multiple meanings do in fact routinely flare up in the mind, even if we ultimately have a way of picking out the most "useful" one.

A second approach is shown in **Figure 7.5B**. The crucial difference here is that activation can move from the top semantic level down to lower levels. Context activates certain semantic features, which in turn activate associated word representations. This means that even before there's any sound input, one word representation may be more active than the other. Once the word itself is uttered, activation moves from the sound units to each of the competing meanings, but since the contextually favored word is already more strongly activated, it inhibits the less-favored reading. As a result, the activation level of the competing meaning may remain very weak, perhaps even negligible. Our response to Virginia Woolf might be: even though language is rife with ambiguity, the mind is very efficiently set up to promote the most "useful" meanings of words at the expense of the less useful ones. In that case, what do we make of Woolf's poetic meditations about rustling skirts? Well, perhaps such alternative meanings come to mind when you think about language in a more deliberate way, with a purposeful focus on the connections between words. But they're unlikely to spontaneously arise in the normal course of language comprehension.

Evidence for the simultaneous activation of word meanings

In 1979, David Swinney published a seminal study that tested whether competing meanings of ambiguous words become simultaneously activated even when plenty of contextual information makes it easy to home in on the most useful meaning. For example, the word *bug* could refer to either a small crawl-

crossmodal priming task An experimental task involving both spoken and written modalities; participants typically hear prime words, which are often embedded within full sentences, and they must respond to test words displayed orthographically on a computer screen.

ing creature or a surveillance device for eavesdropping, but its meaning should be clear in the following context:

> Rumor had it that, for years, the government building had been plagued with problems. The man was not surprised when he found several spiders, roaches, and other bugs in the corner of his room.

Swinney devised an elegant experiment to test the following predictions: Clearly, the relevant meaning of the word *bugs* must become activated in order for someone to understand the passage. Hence, its activation should spread to other words that are related in meaning—for example, the word *ant.* Now, if the irrelevant meaning *also* becomes activated, then its activation should spread as well to related words, such as *spy.* But if the activation of the irrelevant meaning remains low, then the activation flowing to *spy* should be no greater than to a word completely unrelated to either meaning (such as *sew*). And how to measure the spread of activation? Through the familiar semantic priming task.

Swinney used a **crossmodal** variant of the priming task in which subjects listened to passages like the one above and then responded to test words presented visually on a screen (see **Figure 7.6**). Subjects responded by pressing one of two buttons to indicate whether they thought they'd seen a word or a non-word (naturally, plenty of non-word filler items were included). Subjects saw the test word *ant, spy,* or *sew.* Swinney also varied whether subjects heard passages with helpful context or with context that didn't help to disambiguate between the meanings of *bugs.* To complete the comparison, he also varied whether subjects heard the ambiguous word *bugs* in the passage, or an unambiguous counterpart, *insects.*

Figure 7.6 The crossmodal priming task. (A) Sample experimental materials for the task as used by Swinney (1979). (B) A subject from one of the four experimental conditions listens to sentences while sitting in front of a computer screen. At some predetermined point in the sentence, a string of letters appears on the screen, and the subject must press a button to indicate whether the letter string is a real word or a non-word. (Adapted from Swinney, 1979.)

(A)

BIASING CONTEXT

Condition 1: Ambiguous prime
"Rumor had it that, for years, the government building had been plagued with problems. The man was not surprised when he found several spiders, roaches, and other **bugs** in the corner of his room."

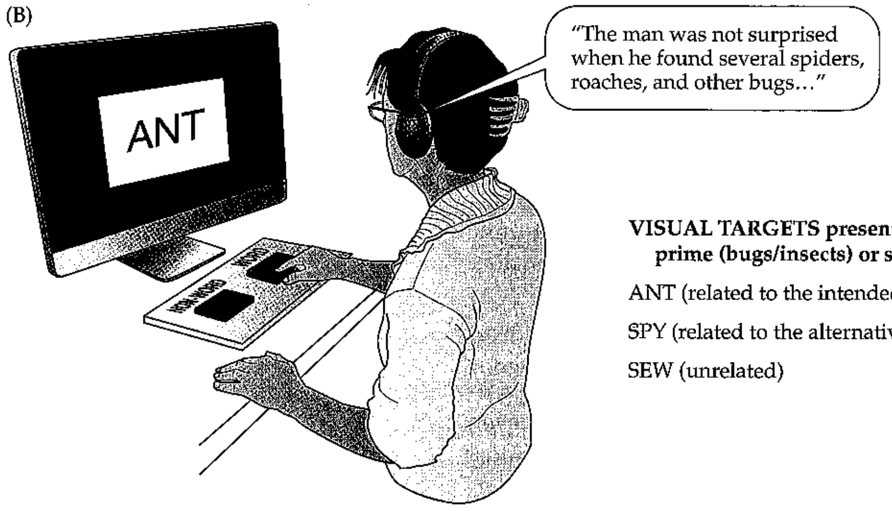
Condition 2: Unambiguous prime
"Rumor had it that, for years, the government building had been plagued with problems. The man was not surprised when he found several spiders, roaches, and other **insects** in the corner of his room."

NEUTRAL CONTEXT

Condition 3: Ambiguous prime
"Rumor had it that, for years, the government building had been plagued with problems. The man was not surprised when he found several **bugs** in the corner of his room."

Condition 4: Unambiguous prime
"Rumor had it that, for years, the government building had been plagued with problems. The man was not surprised when he found several **insects** in the corner of his room."

(B)



VISUAL TARGETS presented either immediately after the prime (bugs/insects) or several syllables downstream

ANT (related to the intended meaning of the ambiguous prime)

SPY (related to the alternative, unintended meaning)

SEW (unrelated)

When subjects had to respond to the target word right after hearing the ambiguous word *bugs,* Swinney saw clear evidence of the activation of *both* meanings of that word. That is, response times were faster to both *ant* and *spy* than they were to the control word *sew.* In contrast, the pattern was quite different when the unambiguous word *insects* was substituted for *bugs*: here, as would be expected, response times for *ant* were faster than for *sew,* but response times for *spy* were not. This shows that it was clearly the presence of the ambiguous word *bugs* that caused the activation of the word *spy.* So Virginia Woolf is at least partly right: even the useless "sunken meanings" become active in the mind along with the obviously intended "surface meanings."

But Swinney's experiment also showed that people quickly converge on the intended meaning. When the same experiment was repeated but with the test word appearing on the screen three syllables after subjects heard the word *bugs* (or *insects*), the results were quite different. In that case, responses to the word *ant,* but not to the word *spy,* were sped up relative to *sew.* It appears that the irrelevant meaning had been activated in parallel with the more pertinent one, but then was quickly suppressed.

These results support a model in which all lexical candidates that match the sound input become active, at least for a time. But later research suggests that the picture is actually a bit more complex, and that in some cases, contextually inappropriate meanings never reach discernible levels of activation. The full story, it turns out, needs to take into account the relative frequencies of the competing meanings.

To see how frequency of meanings might play a role, play along with the following exercise. Quick, now: for each of the words below, say aloud the first synonym or brief definition that comes to mind:

port	straw	chest	pitcher
yarn	cabinet	bark	mint

All of these words are ambiguous. But for some of them, one particular meaning may have sprung to mind immediately, while for others you may have been aware of a mental tug-of-war between two meanings. And if you compared your answers with your those of classmates, some words would have converged on one meaning, while others may have been split between them. This is because for some words (for example, *bark, pitcher, straw, chest*) both meanings occur with roughly equal frequency in English, while for others (for example, *port, cabinet, yarn, mint*) one meaning is dominant and the other is subordinate. A well-known study by Susan Duffy, Robin Morris, and Keith Rayner (1988) explored how this factor interacts with contextual expectations. Rather than using semantic priming as the basis for their experiment, they exploited another behavioral consequence of lexical activation: remember that when multiple words are activated at the same time and compete with each other, as in the studies of neighborhood density effects, people usually take longer to recognize the target word. As you might expect, ambiguous words generally take longer to read than unambiguous ones, providing some additional evidence that multiple meanings of these words are activated simultaneously and compete with each other. And, based on David Swinney's priming results, we might predict that such competition would show up regardless of whether the context favors one of these readings or not, since both meanings appear to be activated, at least for some time.

But Duffy and colleagues found something a bit more subtle. When the context favored the subordinate meanings of words like *mint* or *cabinet,* subjects did read these more slowly than unambiguous control words in the same sentence, indicating competition from the alternative meanings. (For example *mint* was read more slowly than *jail* in a sentence like *Although it was by far the largest building in town, the mint/jail was seldom mentioned.*) But when the words were equally biased in frequency between the two meanings, as for *pitcher* or *straw*, and when the context favored one of their meanings, people spent no more time reading these ambiguous words than unambiguous control words, suggesting that there was no discernible competition from the alternative meaning. This raises the possibility that both frequency of meanings *and* contextual expectations can affect the activation levels of word representations. When these conspire to boost the activation of the same meaning, this lexical representation becomes disproportionately activated, allowing it to very quickly inhibit any potential competitors; however, when the two sources of information conflict, with frequency boosting the activation of one and context favoring the other, the result is roughly equal activation of both meanings, leading to competition between them.

LANGUAGE AT LARGE 7.1

The persuasive power of word associations

What if hearing a word didn't set off resonances within an entire connected network of information? What if word recognition really did just work like a vending machine, with the sounds and letters of words merely acting as pointers or addresses to slots containing meanings or concepts?

If that were the case, marketers, advertisers, and politicians would probably be a lot less preoccupied with words than they are. For example, in a word-vending-machine world, a politician trying to persuade the public on a point of policy shouldn't really care what that policy is called. After all, different ways of saying the same thing would just be different ways to get to the same concept slot in the vending machine.

But politicians and their strategists do care about the names for policies, sometimes obsessively so. For example, Frank Luntz (2007), who has worked as a communications consultant for Republican party candidates in the U.S., has a list of suggestions, seen in the table, for the wording of various policies or initiatives to advance the interests of his clients.

Never say	Instead, say
Tax cuts	Tax relief
Drilling for oil	Energy exploration
Private health care	Free market health care
Wiretapping	Electronic intercepts
Estate tax	Death tax

If you feel insulted at the suggestion that you might feel differently about a plan to reduce taxes depending on the name used to refer to it, you might consider the results of a 2010 poll conducted by CBS and the *New York Times* to probe how Americans felt about having gay people serve in the U.S. military. When people were asked, "Do you favor or oppose gay men and lesbians serving in the military?" the results were as follows:

Strongly favor: 51%
Somewhat favor: 19%
Somewhat oppose: 7%
Strongly oppose: 12%

But when people were asked, "Do you favor or oppose homosexuals serving in the military?" they were less receptive to the idea:

Strongly favor: 34%
Somewhat favor: 25%
Somewhat oppose: 10%
Strongly oppose: 19%

This makes no sense if words are just pointers to concepts. But as you now know, an awful lot happens in the mind on the way to retrieving a word's meaning from memory. Different patterns of activation will resonate throughout the lexicon, depending on whether you hear a phrase like *drilling for oil* or *energy exploration*. (If you want to get a feeling for the differences, you might try a little experiment: give two groups of people a word association task like the one you did in Web Activity 7.1. Among the words on the list, include *energy*, *oil*, *exploration*, and *drilling*, and see what comes up.)

These additional results make it clear that both of the models in Figure 7.5 need to be adjusted to take into account the effects of frequency and the way it interacts with context. But *exactly* how these factors interact is still a matter of some debate. (For example, do they both exert an influence at exactly the same point in time during word recognition, or does one factor come into play earlier than the other?) As a result, researchers are still in the process of refining the models to greater and greater levels of precision and continue to test finer and finer predictions about the behaviors that should result. But there's general agreement that multiple sunken meanings *are* often aroused in the mind (even though the ambiguity doesn't generally impede the eventuality of getting to a single useful meaning), and that the extent to which this occurs depends jointly on the context and frequency of the alternative meanings.

Virginia Woolf's literary excursions raise some questions that don't seem to come up in the more scientific literature. For the most part, even when there is

LANGUAGE AT LARGE 7.1 *(continued)*

A growing body of evidence suggests that the activations that are set off during word recognition probably amount to more than just brief mental flickers that quickly dissipate without any consequences for behavior. Within social psychology, researchers have studied a phenomenon known as **implicit priming**, in which exposing people to certain stimuli increases the likelihood that they'll behave in ways that reflect stored associations, which are activated upon perceiving the stimuli. For example, in one classic study by John Bargh and colleagues (1996), undergraduate students formed sentences out of scrambled word lists, with some students receiving lists that contained words associated with the elderly (for example, *Florida*, *wrinkles*, *bingo*, *gray*) while other students got lists of neutral control words. After the students had finished the test, the experimenters measured and compared how quickly students from the two groups walked down the hall. Those who'd been exposed to the words associated with the elderly walked more slowly than those who'd been in the control condition.

Naturally, marketers are also highly intrigued by the possibility of meaningful links between word associations and behavior or attitudes, and such links are increasingly being tested in the lab by researchers who are interested in the psychology of consumer behavior. To give you just one example, Jonah Berger and Gráinne Fitzsimons (2008) constructed an experiment to see whether exposing people to photographs of dogs would make them more likely to give positive evaluations of sneakers carrying the brand name of Puma. In case the logic behind this study escapes you, it goes like this: Generally, the more familiar people are with an idea or concept, the more they're inclined to like it (which explains why you might have the same TV commercial inflicted upon you half a dozen times during a single program). Berger and Fitzsimons reasoned that because of the similarity between the concepts of dog and puma, pictures of the dog would activate the Puma brand name, making it feel more familiar. As a result, people should experience warmer feelings toward the Puma products, which, in fact, was what the researchers found.

implicit priming A psychological phenomenon in which exposing people to certain stimuli increases the likelihood that they'll exhibit behaviors that are associated with the stimuli. For example, exposing people to words associated with the elderly may trigger behaviors that are stereotypically associated with the elderly, such as walking slowly.

Naturally, just because you see an effect in the lab doesn't mean it will carry the day out in the real world. Real-world choices made by consumers or voters are complex, and likely to be affected by a wide range of different variables; I seriously doubt that you'd be convinced to buy a product you otherwise have no interest in, simply because of the associations set off by its name. Nevertheless, experimental research does lend some credibility to the notion that the Edsel automobile, one of the greatest marketing flops of the last century, wasn't helped in any way by its name. The car was named after one of its makers, Edsel Ford, who unfortunately bore a highly unpopular, old-fashioned, Germanic-sounding name—perhaps not the right one to attach to an American car a mere decade or so after the Second World War.

clear evidence of parallel access of competing meanings, the irrelevant meaning is quite fleeting, and quickly submerged. But do these active meanings, however fleeting, nevertheless manage to have an impact on our aesthetic or emotional experience of language? We don't really know—but recent findings and discussions about how alternative ways of saying the same thing may have different persuasive effects (see Language at Large 7.1) suggest that, perhaps, even brief flickers of activation from "useless" meanings or associated words may not be inconsequential.

WEB ACTIVITY 7.5

Product names and sunken meanings Many product names are deliberately chosen from among the inventory of existing English words (for example, the product names *Apple* and *Tide*). This creates a new lexical ambiguity, where the word can now refer either to its original meaning, or to the newly named product. In this activity, you'll explore some possible implications of this practice.

7.3 Recognizing Spoken Words in Real Time

The flow of spoken words

So far, most of what we've seen about word recognition could apply equally well to the spoken or written language modality, and in fact, the experimental methods that we've seen have relied on both spoken and written stimuli, and sometimes both, to explore the underlying psychological mechanisms. But spoken language offers some particular challenges for hearers, along with some specific puzzles for researchers. We now turn to these modality-specific issues.

One obvious difference between spoken and written words is that when you read a word on a screen or on the page, you can see the whole word at once, and in normal circumstances, you can stare at it for as long as it takes to recognize it. But spoken language unfolds one sound at a time, rather than being uttered all at once, and once it's been uttered, it's gone. As aptly described in a paper by James Magnuson and colleagues (2007), if reading were like listening to spoken language, it would be like this: "Imagine reading this page through a two-letter aperture as the text scrolled past, without spaces separating words, at a variable rate you could not control." It would feel deeply weird to read a word that appeared one or two letters at a time from left to right, and this intuition is confirmed by studies that look at where people focus their gaze while reading. Rather than scanning the word left to right, their gaze lands somewhere within the word, and they can usually read the entire word from that position (or, if the word is very long, they might move their eyes rightward once to take in the rest of the word).

The "scrolling by" nature of spoken word recognition raises a very interesting question: At what point do people initiate the process of matching a string of speech sounds to a stored word representation? In the previous sections, you saw how in the process of word recognition, activation flows from sounds to word representations that contain those sounds. For example, when you hear the sequence of sounds /k/, /ae/, and /t/, the word *cat* will be activated, and also, to a lesser extent, the words *cot* and *can*, among others. But when does the activation of possible word representations begin, given that there's a time lag between the first and last sounds of a word? Does the activation of word candidates start even before the end of the word is encountered in the speech stream, or is it delayed until all the sounds of the word have been uttered? And if people do wait until the entire word has been uttered before activating lexical candidates, how do they identify where the end of the word is anyway, given that usually no silences occur between words in running speech?

In principle, it should be perfectly possible to first locate likely word boundaries, and then activate all of the sounds that bundle together in one word so that they in turn can activate matching lexical candidates. Remember, after all, that we saw in Chapter 4 that even tiny babies were able to figure out where to break the speech apart into words on the basis of statistical information, probably months before they'd acquired much in the way of a working lexicon. So one could imagine that adult word recognition might rely on the same kind of statistically based word segmentation, which would serve as the very first step in spoken word recognition—in a way, we'd be mentally inserting "spaces" between the words before any lexical activation occurred. An analogy with text might be that you'd run a program, based on statistical probabilities, to put spaces between the spoken words before any actual "reading" of the words themselves began.

How can we test to see whether lexical activation begins only after both edges of the word are identified, or whether it's initiated before this point? We can readily recruit some of the methods we've already talked about, and simply tweak them a bit.

For example, remember that in his famous study, David Swinney used semantic priming as a tool to probe for the level of activation of competing word representations: evidence of priming (that is, speeded responses) for words like *spy* and *ant* meant that both meaning representations for *bugs* were activated. We can use a similar logic now, but instead of giving subjects entire words, we can present *partial* words and see whether there's any evidence of priming for words that are related to potential matches to partial words. For example, imagine recording a word like *conform*, and cutting off the sound file right in the middle of the sound /f/. Statistically, the sound sequence /nf/ is extremely unlikely to correspond to the end of a word, so hearers should be able to guess that the end of the word hasn't occurred yet, based solely on information about the sound patterns of English words. So, if lexical activation is delayed until the ends of words are identified, we wouldn't expect to see priming for any words that are related to *conform*—for example *copy* or *imitate*. On the other hand, if lexical activation *is* initiated, we'd expect to see priming not only for words related to *conform*, but also for words related to other possible continuations of this snippet, that is, words semantically related to *conflate, confabulate, confuse, confine, confide, conflicted*, and so on. Such words, with their overlapping onsets, are known as **cohort competitors**.

This latter scenario is exactly what William Marslen-Wilson (1987) predicted. In his **cohort model** of word recognition, he suggested that lexical activation begins right after the beginning of a word, with multiple cohort competitors becoming active. As more and more sound input comes in over time as the word unfolds, the set of possible matching candidates dwindles until the **uniqueness point**, at which there remains only one possible match with the sound input. **Table 7.1** illustrates how the set of cohort candidates becomes smaller and smaller with each incoming snippet of speech.

cohort competitors Words with overlapping onsets (e.g., *candle, candy, candid*, etc.).

cohort model A model of word recognition in which multiple cohort competitors become active immediately after the beginning of word is detected, and are gradually winnowed down to a single candidate as additional acoustic information is taken in.

uniqueness point The point at which there is enough information in the incoming speech stream to allow the hearer to differentiate a single word candidate from its cohort competitors.

TABLE 7.1 Winnowing down cohort candidates as a word unfolds in time

Initial sounds heard	Cohort candidates
/kæ /	*cat, cap, cast, can, cash, cad, camp, cab, cattle, capture, candidate, catholic, candelabra, captain, canteen, castrate, Canada, cancel, castle, canister, captive, candle, cantaloupe, castoff, candy, cannibal, cashew, cantankerous, California, castaway*, many others
/kæn/	*can, candidate, candelabra, canteen, Canada, cancel, canister, candle, cantaloupe, candy, cannibal, cantankerous*, others
/kænɪ/	*canister, cannibal*
/kænɪs/	*canister*

incremental language processing The processing of language in such a way that hearers begin to generate hypotheses about the meaning of the incoming speech on the basis of partial acoustic information, refining and revising these hypotheses on the fly rather than waiting until there is enough information in the speech stream for the hearer to be certain about what the speaker meant.

Evidence for the activation of multiple cohort candidates

A number of experiments by Marslen-Wilson and his colleagues (e.g., Marslen-Wilson, 1987; Zwitserlood, 1989) used crossmodal priming tasks and found evidence for the parallel activation of multiple cohort competitors based on partial words. That is, even in the middle of a word, there was evidence for the activation of words that were semantically related to cohort candidates. For example, while hearing the word fragment *cap-*, people were relatively fast in a lexical decision task to respond to words like *ship* or *jail*, presumably because these words were semantically related to the cohort candidates *captain* and *captive*.

These results were important and suggested that language processing is highly **incremental**—that is, hearers don't cautiously hang back during language processing and wait until there's enough information to be certain about what the speaker meant; rather, they eagerly generate hypotheses about the meaning of the unfolding speech, and refine and revise these hypotheses on the fly.

The incremental nature of word recognition can be demonstrated robustly with the use of another method we've discussed earlier—that of tracking hearers' eye movements continuously in response to spoken language. This method is especially useful for studying word recognition in real time. Unlike the priming method, which provides a set of snapshots of activation levels at different points in the speech stream, eye tracking provides something more like a movie of the word recognition process, showing how patterns of eye movements respond in lockstep with the unfolding speech stream.

Eye-tracking studies provide clear evidence that cohort competitors become activated. Paul Allopenna and colleagues (1998) were among the first to report extremely detailed eye movement data, providing a dynamic, moment-by-moment view of the word recognition process. They argued that the likelihood of looking at an object was directly related to the activation level for the word corresponding to that object, so by looking at the rising and falling patterns of eye movements for a group of subjects, you could get a fairly continuous look at activation levels for words over time. You can see a cohort effect in action in **Figure 7.7**.

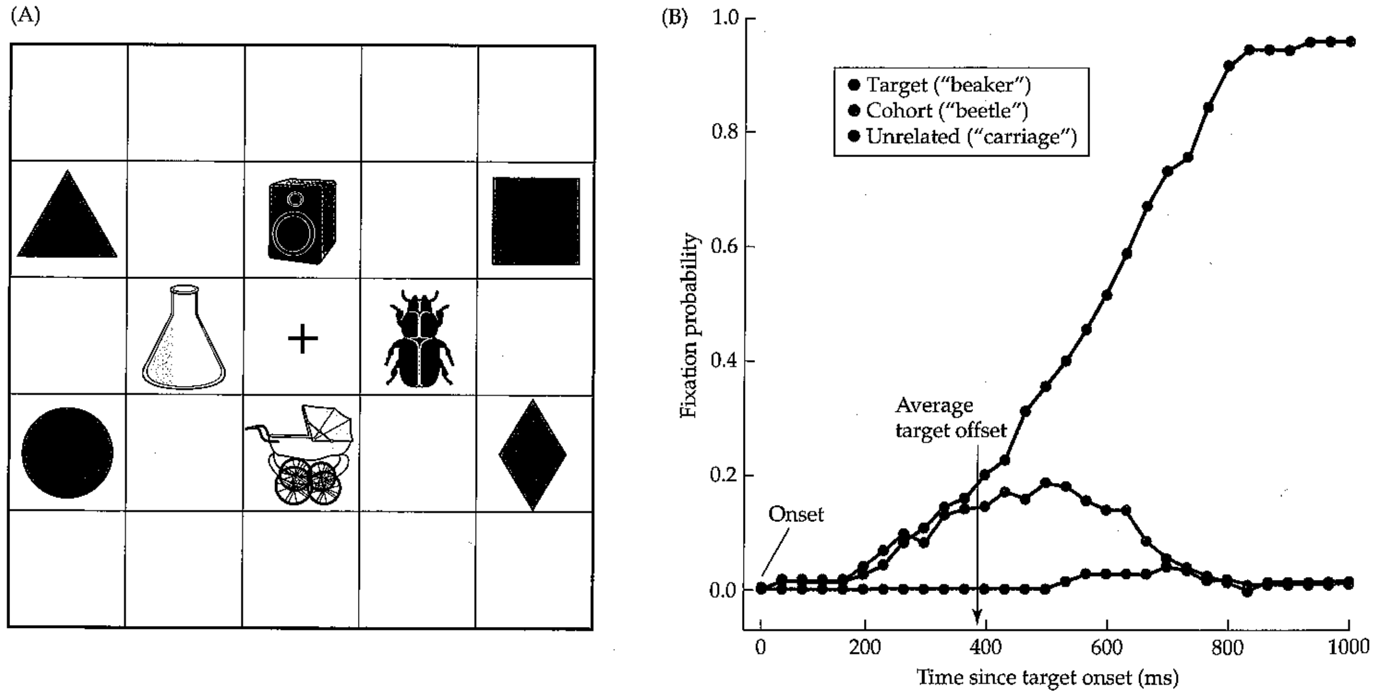


Figure 7.7 Sample visual displays and data from the eye-tracking experiment by Allopenna et al. (1998). (A) An example display, which was accompanied by instructions such as "Pick up the beaker. Now put it above the triangle." Each display contained a target referent (beaker) and a cohort competitor (beetle) as well as an unrelated item (carriage). (B) A graph showing the likelihood of looking at the target referent, the cohort competitor, and the unrelated item in the display. Notice that subjects were equally likely to look at the target referent and the cohort shortly after the onset of the word. (Adapted from Allopenna et al., 1998.)

Subjects heard spoken instructions, such as "Pick up the beaker," accompanied by visual displays in which the target object (i.e., a beaker) was always present. In some of the displays, they saw a cohort competitor (for example, a beetle), as well as other, unrelated objects. The results showed that people were much more likely to look at the cohort competitor objects than at the unrelated objects, and that these eye movements were often initiated extremely early in the word, on the basis of very little phonetic information. The time needed to say the average target word in this study was about 400 milliseconds (ms). At about 200 ms, subjects were beginning to launch more eye movements to the objects matching either the target or cohort words than to the unrelated objects, suggesting that the words *beetle* and *beaker* were already more activated than other words. Given that there may be as much as a 200 ms time lag between the point an eye movement is programmed and the point the eyes actually begin to start moving, this reveals that people were starting to zero in on the target and cohort as potential matches for the incoming word on the basis of the thinnest slice of phonetic evidence, in some cases merely a sound or two. Eye movements to the cohort began to drop off shortly after the end of the word (and hence were programmed roughly 200 ms earlier), in response to new sound input that ruled the cohort word as an incompatible match. This means that before the end of the word occurred, not only were the target and its competitor both activated, but people were already beginning to identify the target as the actual word that had been spoken. Data like these very neatly rule out the notion that word segmentation has to occur before any word activation takes place. Figure 7.7 reveals the truly incremental nature of word recognition.

Since these early studies, continuous tracking of eye movements has corroborated some of the findings about word recognition that have come from other methods. For example, eye movements show that not only cohort competitors, but words that are semantically related to the cohort words become activated shortly after the beginning of a spoken word, nicely paralleling Zwitserlood's (1989) semantic priming study with cohort competitors (for example, the word *hammock* triggers eye movements to a picture of a nail, via its semantic relationship to the cohort competitor *hammer*, as found by Yee & Sedivy, 2006). Effects of frequency and neighborhood density are also apparent (see Magnuson et al., 2007), lining up tidily with studies from lexical decision tasks. Researchers have also used eye movement studies to figure out whether people who are bilingual keep their two languages separate or experience crosstalk between them (see **Box 7.3**). And in the next section, you'll see how tracking eye movements has also been useful in addressing subtle theoretical questions about spoken word recognition.

WEB ACTIVITY 7.6

Identifying cohort candidates A simple task known as "gating," can provide a quick, accessible window on how multiple cohort candidates become activated on the basis of partial sound input from a word. This activity will give you a feel for the process of winnowing down a list of candidate matches as sound input accumulates.

How important are the left edges of words?

So far, I've presented a tidy pile of data to convince you that the activation of lexical candidates is triggered well before the end of the word, and that multiple words whose onsets overlap become activated all at the same time. This shows that you don't need to identify the right edges of words (that is, their ends) before generating possible matches for the sound input, as emphasized by the cohort model. But notice: the cohort model *does* require that you've identified the *left* edge of words, since the whole point is that cohort candidates are activated on the basis of whether they're consistent with the sound input of the word *so far*. Meaning, you have to know where the word's beginning is.

BOX 7.3

Do bilingual people keep their languages separate?

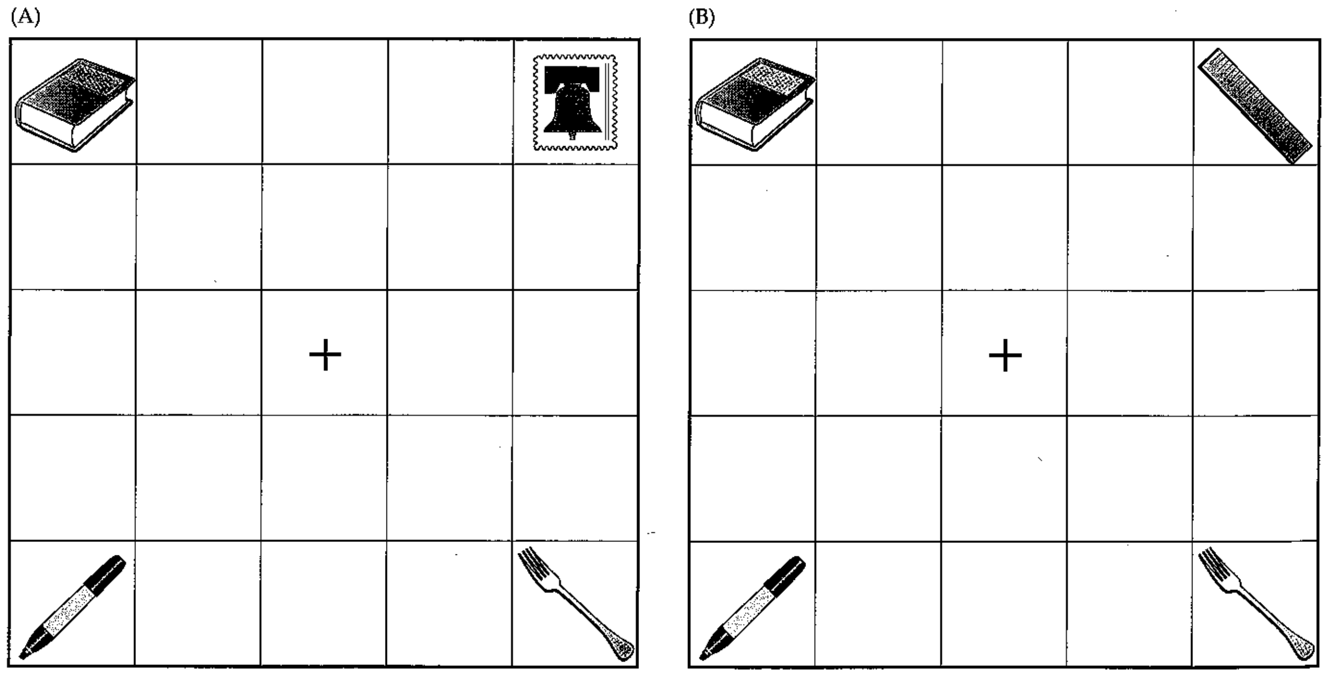
If you know more than one language, to what extent are you able to keep your two language systems separate in the frenzy of daily speech and comprehension? For example, when you hear someone speaking to you in one of your languages, do you limit access to the words in memory that belong to that language? Or do you also activate the words that belong to your other language?

To find out, we can draw on some of the methods from this chapter to see whether you experience competition from words that sound alike but belong to separate languages. In this section, you've seen evidence that within a single language, cohort competitors with overlapping sounds at the beginnings of words are simultaneously activated—so a spoken word like *beaker* leads to the activation of *beetle*. Suppose you're a Russian-English bilingual, and you're listening to someone saying, "Can you hand me the marker?" Would you *also* activate the Russian word for a stamp—pronounced as /marka/?

Michael Spivey and Viorica Marian (1999) designed a simple eye-tracking study to probe for competition effects from cross-language cohort competitors. They created visual displays in which some of the trials contained a cross-language cohort competitor for the target word (for example, if the target was the English word *marker*, the cross-language cohort competitor was a stamp). They compared the eye movement patterns for these "cohort" displays with "control" versions of the same displays in which none of the objects in the display had names in either language that overlapped with the target word (for example, a ruler—pronounced /liñejka/ in Russian). Specifically, they wanted to know whether, upon hearing the English word *marker*, their bilingual subjects would be more likely to look at the stamp (/marka/ in Russian) than at the ruler (/liñejka/ in Russian), which served as the unrelated "control" object (see **Figure 7.8**). The experiment was conducted in both English and Russian (when /marka/ was the target word in Russian, the marker served as the cross-language competitor object).

The eye movement data showed that when the subjects heard the experimental instructions in Russian, there was strong evidence of competition from the English cohort competitor. There wasn't a clear effect of cross-language competition in the other direction, though; evidently in some cases, cross-language competition can be muted or absent (see **Figure 7.9**). Further experimentation (e.g., Marian & Spivey, 2003) suggested that the degree of cross-language activation could be dialed up or down, depending on a number factors, including which language is the first or dominant language and whether the experimental setting is a bilingual one or a purely monolingual interaction.

Figure 7.8 A sample visual display from the eye-tracking experiment by Spivey and Marian (1999), accompanied by the English instruction, "Pick up the marker." (A) The "cohort" version of the display contains a postage stamp, whose Russian name (/marka/) is a cohort competitor for the English target word (*marker*). (B) The "control" version of the display, in which the cross-language cohort competitor (the stamp) has been replaced by an object (the ruler) whose Russian name does not overlap with the English target.



BOX 7.3 *(continued)*

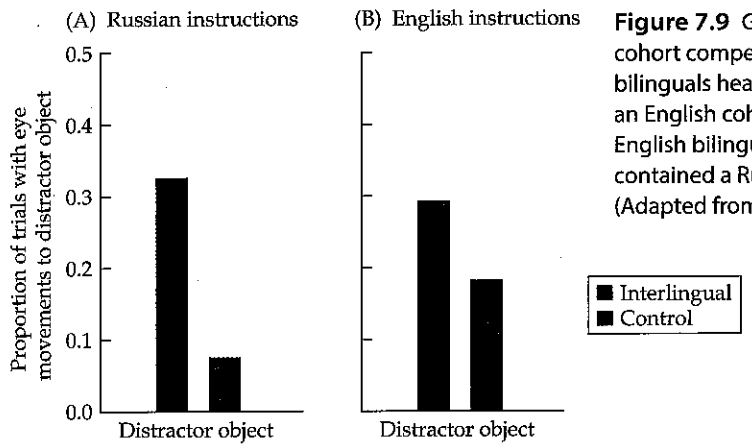


Figure 7.9 Graphs showing eye movements to the cross-language cohort competitor and to an unrelated control. (A) Russian-English bilinguals heard instructions in Russian, while the display contained an English cohort competitor to the Russian target word. (B) Russian-English bilinguals heard instructions in English, while the display contained a Russian cohort competitor to the English target word. (Adapted from Spivey & Marian, 1999.)

These studies (along with a number of others) show that languages aren't walled off from each other during daily use, and activation within one system can leak over into the other system. You might imagine that in some cases, cross talk between languages might be useful. For instance, languages that are related to each other or that have borrowed heavily from each other may have a number of cognate words, which share features of both sound and meaning, as in the following examples, all words that mean the same as English *night*:

Nacht (German)
nag (Afrikaans)
natt (Swedish, Norwegian)
nat (Danish)
nátt (Faroese)
notte (Italian)
noche (Spanish)

In the case of cognate words, activating the English word *night* upon hearing *Nacht* in German can help you access the word's meaning in German. But in many cases, overlapping sounds in words across languages will be completely coincidental (sometimes unfortunately so, as exemplified by the regrettable similarity of a certain English swear word and the French word for seal, which is *phoque*). In these cases, activation from the other language can only serve to make the task of word recognition harder, requiring extra time and effort to suppress irrelevant words. Bilingualism, then, seems to come with some added burdens for language processing. This isn't necessarily a bad thing; to foreshadow the upcoming chapter, the added strain of cross-language activation can have some surprising cognitive benefits in the long run.

This seems like a pretty reasonable assumption. After all, figuring out the left edge of a word in running speech shouldn't be that hard, for the simple reason that once you've recognized the previous word, it's quite obvious where the word break should go—at the end of *that* word! And as we've seen, the recognition of the previous word has, in all likelihood, already been a resounding success even before the last sounds of that word enter your ears, so there should be no ambiguity about where the word boundary should be. That's if all goes well, of course. But suppose that, for some reason, there's a slip-up in the recognition process, and the word boundaries are set in the wrong place. The effects of this initial error should be catastrophic, cascading down the speech stream. Or suppose that someone in the room coughs or a cell phone goes off right at the moment the initial sounds of a word are being uttered. This should be terribly disruptive to word recognition.

mondegreens "Slips of the ear" that result in errors of word segmentation.

"Slips of the ear" do happen for these and other reasons, resulting in mishearings that have become known as **mondegreens**. This term was first coined by American writer Sylvia Wright, who recalled how as a child, she'd heard the Scottish ballad "The Bonnie Earl of Murray," which went like this:

Ye Highlands and Ye Lowlands
Oh where hae ye been?
They hae slay the Earl of Murray
And Lady Mondegreen.

At least, that's how she remembered it, only to learn years later that the last line was actually "and laid him on the green." Other famous examples of mondegreens include hearing Bob Dylan's "The answer, my friend, is blowin' in the wind" as "Dead ants are my friend, they're blowin' in the wind" or hearing "the girl with kaleidoscope eyes" as "the girl with colitis goes by" in the Beatles' "Lucy in the Sky with Diamonds."

Perhaps it's not surprising that mistakes like this often happen with song lyrics, where the speech sounds are distorted and accompanied by music. But some researchers have argued that, even for normal speaking situations, the cohort model would lead us to expect many more such failures than actually occur, all because of the fact that the whole process rests on that crucial identification of the word's left edge. Jay McClelland and Jeff Elman (1986) proposed their TRACE model as a way to preserve the large-scale competition effects that are part and parcel of word recognition, while making the system more resilient against processing disruptions. In their model, streams of sound input are continuously fed into the word recognition system and activate the words that contain them, without any need for identification of the left edges of words. So, let's suppose you've just entered a conversation, and not having clearly heard all the sounds that are being spoken, you're in the midst of catching a short burst of sounds consisting of *astdan*. This ordered stream of sounds begins to activate possible matches, including words like *aster, astronaut*, or *Aztec*. But all of these possibilities quickly become mismatched with the sound input upon encountering the /d/ sound and so become deactivated. Here, the cohort and TRACE models predict very different outcomes. The cohort model would be at a loss to propose any remaining viable candidates, because there are simply no options left for words that begin with the sounds *astdan*. But TRACE isn't limited to activating words that make any assumptions about the left edges of words, so it can also activate possible matches, such as *fast Dan* or *last dance*, which may well turn out to be consistent with both the continuing sound input and the context of the sentence.

The TRACE model makes a clear prediction that is at odds with the cohort model, namely, that words whose sounds overlap with the target word in the middles or ends (and not just the onsets) should become activated as well. That is, a word like *beaker* should activate *speaker*, and not just *beetle*. Hence, using the semantic priming paradigm, TRACE would predict that the word *beaker* should speed up recognition times for a word like *music* (via *speaker*) as well as *insect* (via *beetle*) compared with some unrelated word (for example, *table*). It's turned out to be quite difficult to find effects of semantic priming via rhyme competitors. On the other hand, eye movement studies (for example, by Paul Allopenna and colleagues, 1998) have found evidence of activation of rhyme competitors, as illustrated in **Figure 7.10**. Nevertheless, overlap at the beginnings of words clearly results in greater competition than overlap at the ends of words.

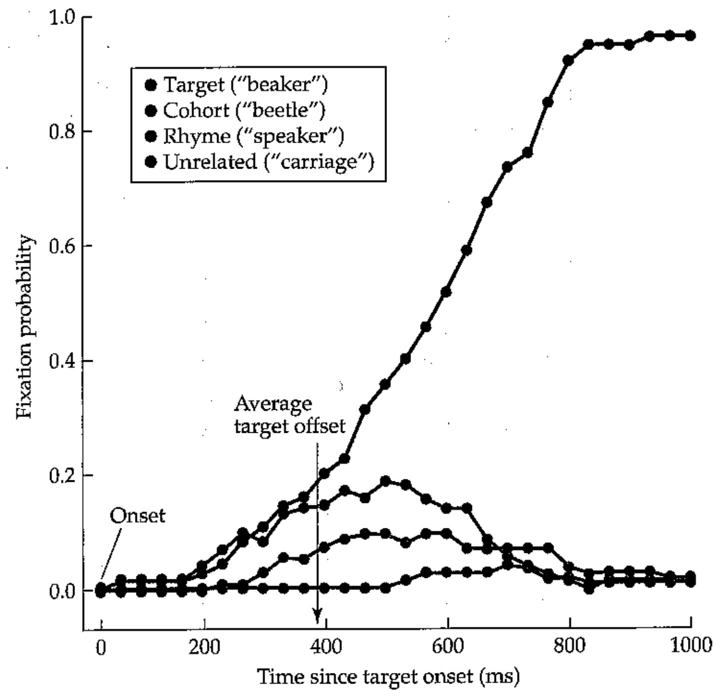


Figure 7.10 Eye movements reveal a "rhyme effect." The graph reflects the subjects' likelihood of looking at the target referent (beaker), its cohort competitor (beetle), its rhyme competitor (speaker), and an unrelated object (carriage). Note that the rhyme effect occurs later than the cohort effect (see Figure 7.7), and is more subtle. (Adapted from Allopenna et al. 1998.)

TRACE can explain this asymmetry by including inhibitory links between competing words and between the sound and word levels of representations. Rhyme competitors will be at a disadvantage because the activation of *speaker* will be pushed down relative to *beaker* or *beetle*, based solely on its mismatch with the first few sounds. Eventually as the word unfolds in time, there will be some overlap between the words *speaker* and *beaker*, so this will boost the activation of *speaker*, but this new surge of activation will have to overcome the initial dampening of that word based on the early mismatch of sounds.

The cohort and TRACE models differ on a number of dimensions, not just with respect to whether the possible lexical candidates have to be aligned with the left edge of the word. (For example, the models disagree on how top-down information can affect the activation of lower levels of representation, a theme that will be taken up in Digging Deeper at the end of this chapter.) Both models have undergone major renovations in response to new experimental data. To some extent, the changes to the models have made it more difficult to know which model—if either of them—is "right," as there's now a good deal of overlap in the results that the two can account for, even though they resort to quite different mechanisms to achieve them. This might seem a touch depressing to someone who believes that the ultimate goal of science is to declare a "winner" between competing theories or models. But ongoing tension between detailed models can really push a field forward, and it is scientific progress that becomes the ultimate winner in the game. Without competing models, it would be a lot harder to formulate the questions that drive the collection of data. For instance, the clashing of the predictions of the cohort and TRACE models is what motivated researchers to closely compare the activation levels of cohort versus rhyme competitors. And without these detailed models, what we know about word recognition would be more like a jumbled laundry list of experimental effects than a set of subtly contrasting mock-ups of what's going on inside our heads.

7.4 Coping with the Variability of Sounds

The problem of perceptual invariance

In the previous section, we considered some of the questions that come up as a result of the fact that words have to be uttered over a span of time. But an even more radical source of differences between spoken and written word recognition is the fact that spoken language is constrained by the shapes, gestures, and movements of the tongue and mouth.

So far, I've been talking about words as being made up of *sequences,* or *strings* of letters or sounds. Such language is fine for describing written words, which are indeed made up of separate letters strung together like beads on a necklace. But it's deeply misleading when it comes to spoken words. Far from resembling beads in a necklace, sounds combined in speech result in something like this, suggested the American linguist Charles Hockett (1955, p. 210):

> Imagine a row of Easter eggs carried along a moving belt; the eggs are of various sizes, and variously colored, but not boiled. At a certain point the belt carries the row of eggs between the two rollers of a wringer, which quite effectively smash them and rub them more or less into each other. The flow of eggs before the wringer represents the series of impulses from the phoneme source; the mess that emerges from the wringer represents the output of the speech transmitter.

Hence, the problem for the hearer who is trying to identify the component sounds is a bit like this:

> We have an inspector whose task it is to examine the passing mess and decide, on the basis of the broken and unbroken yolks, the variously spread out albumen, and the variously colored bits of shell, the nature of the flow of eggs which previously arrived at the wringer.

Unlike letters, which occupy their own spaces in an orderly way, sounds smear their properties all over their neighbors (though the result is perhaps not *quite* as messy as Hockett's description suggests). Notice what happens, for example, when you say the words *track, team,* and *twin.* The /t/ sounds are different, formed with quite different mouth shapes. In *track,* /t/ sounds almost like the first sound in *church;* your lips spread slightly when /t/ is pronounced in *team,* in anticipation of the following vowel; and in *twin,* the /t/ sound might be produced with rounded lips. In the same way, other sounds in these words influence their neighbors. For example, the vowels in *team* and *twin* have a nasalized twang, under the spell of the nasal consonant that follows. It's impossible to tell exactly where one sound begins and another one ends. This happens inevitably because of the mechanics involved in the act of speaking.

As an analogy, imagine a sort of signed language in which each "phoneme" corresponds to a gesture performed at some location on the body. For instance, /t/ might be a tap on the head, /i/ a closed fist bumping the left shoulder, and /n/ a tap on the right hip. Most of the time spent gesturing these phonemic units would be spent on the transitions between them, with no clear boundaries between units. For example, as soon as the hand left the head and aimed for the left shoulder, you'd be able to distinguish that version of /t/ from one that preceded, say, a tap on the chin. And you certainly wouldn't be able to cut up and splice a videotape, substituting a tap on the head preceding a tap on the left shoulder for a tap on the head preceding a tap on the chest. The end result would be a Frankenstein-like mash. (You've already encountered this problem earlier, in Chapter 4 and in Web Activity 4.4.)

The variability that comes from such coarticulation effects is hardly the only challenge for identifying specific sounds from a stream of speech. Add to this the fact that different talkers have different sizes and shapes to their mouths and vocal tracts, which leads to quite different ways of uttering the same phonemes, and yet we're somehow able to hear all of these as equivalent. And add to *that* the fact that different talkers might have subtly different accents—and again, unless the accent is very thick, this doesn't seem to prevent us from understanding each other. When it comes down to it, it's extremely hard to identify any particular acoustic properties that clearly map onto specific sounds. So if we assume that part of recognizing words involves recovering the individual sounds that make up those words, we're left with the problem of explaining the phenomenon known as **perceptual invariance**: how is it that such variable acoustic input can be consistently mapped onto stable phonemic units of representation?

The motor theory of speech perception

The problem of perceptual invariance has led some researchers to suggest that perhaps speech perception *doesn't* actually involve recovering the sounds of a word. Perhaps what we do instead is reconstruct the series of *gestures* that make up the word. The idea driving this account, called the **motor theory of speech perception**, is that there's a fairly direct link between the acoustic signals in speech and the gestures that produce it. So why not assume that the articulatory gestures form the backbone of the speech perception system, rather than the activation of abstract units of representation such as phonemes?

And if speech perception is really about reconstructing the gestures that make a word, then shouldn't *visual* input about how a word is formed—that is, lipreading—ultimately have an impact on how we "hear" a word? Strange as this hypothesis might seem, there's very strong evidence that this actually happens, as illustrated in Web Activity 7.7.

WEB ACTIVITY 7.7

The McGurk effect In this exercise, you'll see a demonstration of the McGurk effect, in which the perceptual system is forced to resolve conflicting cues coming from the auditory and visual streams.

As seems evident from the well-known auditory illusion in Web Activity 7.7, visual input does seem to be folded into the representation of sounds. When people see a video of a person uttering the syllable *ba,* but the video is accompanied by an audio recording of the syllable *ga,* there's a tendency to split the difference and perceive it as the syllable *da*—a sound that is produced somewhere between *ba,* which occurs at the front of the mouth, and *ga,* which is pronounced at the back, toward the throat. This finding, the **McGurk effect,** is a nice, sturdy experimental effect. It can be seen even when subjects know about and anticipate the effect, it occurs with either words or non-words (Dekle et al., 1992)—and it even occurs when blindfolded subjects *feel* a person's lips moving to say *ba* while hearing recordings of the syllable *ga* (Fowler & Dekle, 1991). The McGurk effect seems to support the motor theory quite nicely. Other support for this view comes from brain-imaging research showing that simply listening to speech sounds involves brain activity in motor areas of the brain—in fact, the same areas that are active when people imagine themselves actually articulating that sound (D'Ausilio et al., 2009).

But, while awareness of the gestures that make sounds clearly plays a role in speech perception, it doesn't seem to be true that such knowledge is *required.* If that were the case, we'd expect that brain damage to the areas that are responsible for speech production would inevitably disrupt speech perception as well.

perceptual invariance The phenomenon whereby acoustically different stimuli are perceived as examples of the same phoneme or word.

motor theory of speech perception A theory that the perception of speech sounds involves accessing representations of the articulatory gestures that are required to make those speech sounds.

McGurk effect An illusion in which a mismatch between auditory information and visual information pertaining to a sound's articulation results in altered perception of that sound; for example, when people hear an audio recording of a person uttering the syllable *ga* while viewing a video of the speaker uttering *ba,* they often perceive the syllable as *da.*

But it's not hard to find people who have had a stroke and suffered massive damage to the motor speech system, and have serious speech production problems as a result, but can still recognize words just fine. And there's other evidence of the dissociation between production and perception. For example, as we saw in Chapter 4, very young babies can distinguish many sounds just after birth, and even sort them into perceptual categories with clear boundaries. Between 6 and 8 months of age, they're able to track the statistical relationships among sounds, well before they can control their own mouths well enough to reliably produce those sequences of sounds. Even more damaging for the motor theory is the fact that animals like chinchillas can form sound categories in a way that is similar to humans, and even adjust their perception of sounds depending on neighboring sounds. Since it's a bit of a stretch to argue that chinchillas have mental templates for how human sounds are pronounced, we may be stuck after all with explaining how acoustic properties map onto representations of sounds and, along with it, grappling with the invariance problem.

Context effects in speech perception

Another way to explain how we might get perceptual invariance from variable acoustic data goes something like this: Perception itself involves much more than just mapping acoustic signals to corresponding sounds. It also involves using contextual cues to *infer* those sounds. In doing that, you work backward and apply your knowledge of how similar sounds "shape-shift" in the presence of their neighbors, to figure out which sounds you're actually hearing. It turns out that a similar story is needed to account for perceptual problems other than speech. For example, you can recognize bananas as being yellow under dramatically different lighting conditions, even though more orange or green might actually reach your eyes depending on whether you're seeing it outdoors on a misty day or inside by candlelight (see **Figure 7.11**). Based on your previous experiences with color under different lighting conditions, you perceive the banana as having a constant color, rather than changing chameleon-like in response to the variable lighting. Without your ability to do this, color would be pretty useless to you as a cue in navigating your physical environment. In the same way, knowing about how neighboring sounds influence each other might impact your perception of what you're hearing. The sound that you end up "hearing" is the end result of combining information from the acoustic signal with information about the sound's surrounding context. (Notice that much the same explanation can be provided for the McGurk effect: in that case, what you "hear" is the result of folding together auditory and visual information.)

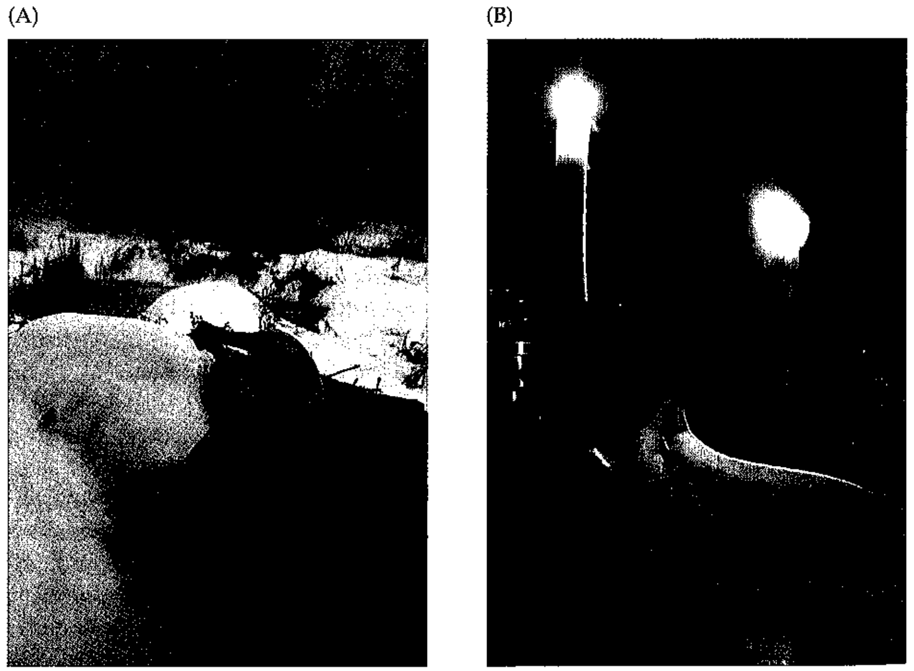

(A) (B)

Figure 7.11 Color constancy under different lighting. We subjectively perceive the color of these bananas to be the same under different lighting conditions, discounting the effects of illumination. Similar mechanisms are needed in order to achieve a stable perception of variable speech sounds.

Another way in which context might help you to identify individual sounds is knowing which *word* those sounds are part of. Since your knowledge of words includes knowledge of their sounds, if you have an idea of what word someone is trying to say, this should help you to infer the specific sounds you're hearing. A classic study by William Ganong (1980) illustrates just this effect, which has since become immortalized as the **Ganong effect.** In his experiment, subjects listened to a list of words and non-words and wrote down whether they'd heard a /d/ or a /t/ sound at the beginning of each item. The experimental items of interest contained examples of sounds that were acoustically ambiguous, between a /d/ and /t/ sound, and that appeared in word frames set up so that a sound formed a word under either the /d/ or /t/ interpretation, but not both. So, for example, the subjects might hear an ambiguous /d-t/ sound at the beginning of __ask, which makes a real word if the sound is heard as a /t/ but not as a /d/; conversely, the same /d-t/sound would then also appear in a frame like __ash, which makes a real word with /d/ but not with /t/. What Ganong found was that people interpreted the ambiguous sound with a clear bias in favor of the real word, even though they knew the list contained many instances of non-words. That is, they reported hearing the *same* sound as a /d/ in __ash, but a /t/ in __ask.

Ganong effect An experimental result demonstrating that the identity of a word can affect the perception of individual sounds within that word. When people hear a sound that is acoustically ambiguous between two sounds, their identification of that sound can be shifted in one direction or another depending on which of the possible sounds results in an actual word.

The Ganong effect helps to explain how it is that we're not bothered by the inconsistency of sounds in the context of word recognition. It also has some important implications for our general program of model-building, as laid out in Section 7.1. If you remember, we ended that section by leaving open the possibility that activation from a higher level of representation—the word level—sends activation down to the lower level of sound representation. A model like this would predict that the same sound should be perceived differently depending on the word it's embedded in—precisely the effect discovered by William Ganong (see **Figure 7.12**). But the Ganong effect also shows limits to the influence of top-down information. The word frame only had an effect on sounds that straddled the category boundary between a /t/ and /d/. If the sounds were good, clear examples of one acoustic category or the other, subjects correctly perceived the sound on the basis of its acoustic properties, and did not report mishearing *dask* as *task*, for example. This shows that when a sound is strongly activated on the basis of very clear bottom-up evidence from the acous-

WEB ACTIVITY 7.8

The phoneme restoration effect In this example, you'll hear an audio clip illustrating the phoneme restoration effect, in which knowledge of a word allows the hearer to "fill in" a missing sound in the speech stream.

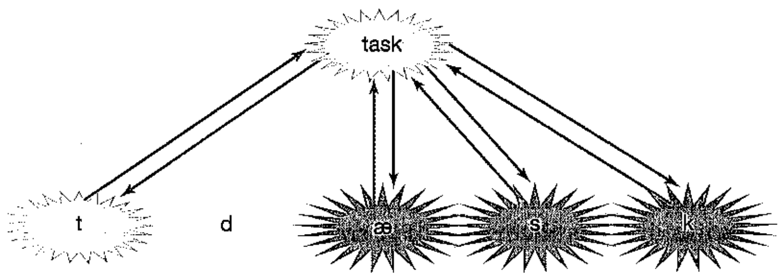


Figure 7.12 Top-down activation from words to sounds. In this model, the activation of the sound /t/ is initially weak, as shown, because the acoustic signal provides a poor example of this sound. However, over time, activation feeds up to the word level, activating the word *task*, and then bounces back down to the individual sounds contained in that word.

LANGUAGE AT LARGE 7.2

How does ventriloquism work?

Good entertainment often bends reality. That's true in spades for skillful acts of ventriloquism: we know that the words are uttered by the human handler, but we can't stop ourselves from hearing the dummy "speak" them. So how does this work?

Ventriloquism cashes in on several different kinds of illusions. The first of these has to do with perceiving the location of a sound. As is apparent from the McGurk effect, what we "hear" is usually the end result of combining visual and auditory information, whether the cues are about articulation or about the location of a sound in space. In the natural world, visual and auditory cues almost always line up, but since human beings are curious enough to wonder what happens if they don't, we now know what happens when there's a disconnect between the two. In a ventriloquist's act, the lips of the human appear to be still, while the dummy's mouth flaps in time with the words that are spoken. What happens if your ears are telling you one thing about the location of a sound, and your eyes are telling you another? It turns out that you believe your eyes. You don't have to see a ventriloquist to experience this illusion. Have you ever thought about what happens in a movie theater? The sound is coming from speakers placed around the room, but it *feels* as if spoken words are coming straight out of the mouths of the actors on-screen.

The reason you put more trust in what your eyes are telling you is simply that most of the time, visual information is a more reliable source of information about an object's location than auditory information is. But that can change in a situation where the visual cues become fuzzy or indistinct—in that case, you'd likely depend more on sound to figure out an object's location in space. It's also possible that the location illusion is affected somewhat by where you're directing your visual attention. If you look at YouTube clips of ventriloquist acts, you might notice that the dummy is usually designed to draw more visual attention than the human, with brighter clothing and humorous facial features, and is very animated in its movements, while the human tends to blend into the background, wear nondescript clothing, and have very limited movements of both the face and body while the dummy is "speaking."

The second kind of illusion deals with the actual sounds of speech and how they're interpreted. Ventriloquists manage to speak without moving their lips, so how is it that they're able to produce the full range of speech sounds? The answer is that they don't; to some extent, they can rely on us as hearers to hear what we expect to hear rather than what's really there.

Ventriloquists speak with their lips still and partly open, which nevertheless allows them to make many speech sounds by controlling the shape and movement of the tongue behind the lips in a way that's not visible.

tic signal, top-down expectations from the word level aren't strong enough to cause us to "hallucinate" a different sound. However, when the acoustic evidence is murky, word-level expectations *can* lead to pretty flagrant auditory illusions (you can experience one yourself in Web Activity 7.8).

The **phoneme restoration effect**, first discovered by Richard Warren (1970), offers a dramatic example in which the acoustic input *is* misheard. In these examples, a speech sound is spliced out—for example, the /s/ in *legislature* is removed—and a non-speech sound, such as a cough, is pasted in its place. The resulting illusion causes people to "hear" the /s/ sound as if it had never been taken out, along with the coughing sound.

Again, this suggests that word-level expectations can activate individual sounds by means of top-down connections. This particular illusion seems to become stronger as the activation for the word itself increases; the effect is stronger for longer words (which have fewer lexical competitors) than for shorter words. It's also more robust for words that make sense in the context of the sentence than for words that don't fit with the sentential context. These results make sense if words themselves can receive activation from an even higher level of representation.

phoneme restoration effect An auditory illusion showing that when a speech sound within a word is replaced by a non-speech sound, people often report hearing both the speech and non-speech sounds.

LANGUAGE AT LARGE 7.2 *(continued)*

But sounds made by moving the lips—like /b/, /p/, /w/, /m/, /v/, and /f/—are problematic. To some extent, the ventriloquist can write a script that avoids words containing these sounds, but when absolutely necessary, these labial sounds can be replaced by other similar-sounding phonemes. For example, /f/ might be replaced by /θ/ (the first sound in *think*), and /m/ by the similar-sounding /ŋ/in *sing*. We've already seen from the Ganong effect and phoneme restoration effect that the sounds we think we hear depend to some extent on the sounds that we expect to hear in a particular word. Ventriloquism takes full advantage of this.

Based on some of the results that we've talked about with these particular auditory illusions, we should also be in a position to offer some scientifically grounded advice to an aspiring ventriloquist. For example, remember the Ganong effect, in which a sound was perceived in such a way as to be consistent with the rest of the word. If you want your hearer to perceptually morph a *d* into a *b* sound, best to use the sound in a word where it wouldn't make sense as a *d*. (For example, in *dest*, but not in *dust* when it's intended to be heard as *bust*.) Remember also that the effect of the surrounding word was strongest when the sound itself was ambiguous—clear examples of sounds that made no sense in the words they were embedded in (for example, *dest*) were just perceived as mispronunciations. This suggests that you can't simply throw in a clear-sounding *d* and expect it to be heard as a *b* sound, even if the word makes no sense with a *d* sound. For this reason, ventriloquist manuals usually suggest altering the substituted sounds in some way. For instance, one way to make a *d* sound slightly more like a *b* sound is to pronounce it by placing the tongue against the teeth rather than in its usual position on the ridge behind the teeth. It takes some practice to distort certain sounds so that they become more acoustically similar to the intended labial sounds.

The phoneme restoration effect showed that even quite garbled sounds can be reinterpreted as specific phonemes, and that the illusion is especially strong with long words that have few competitors, and with words that are generally quite predictable in context. In other words, the more strongly activated the word is, the stronger an impact it's likely to have on how people perceive the sounds within it. It seems to make sense, then, that words that contain substituted sounds should be ones that are likely to be strongly activated in the context.

Once you start seeing how this genre of entertainment fits neatly with a model of normal, everyday speech perception, it might be hard to stop making the connections. In fact, you might even be tempted to make specific predictions about best ventriloquist practices, and put them to rigorous empirical tests by adapting some of the experimental approaches you've learned about here.

7.5 Reading Written Words

Word recognition in spoken and written modalities

As far as we know, spoken language emerged spontaneously as a result of humans interacting with each other, rather than being deliberately engineered by some unusually enterprising human. But that's not the way written language came about. A writing system *does* need to be consciously designed and usually taught in a much more formal, deliberate way than spoken (or signed) language. It's rather remarkable, then, that many people learn to recognize written words as quickly and automatically as they do. For expert readers, the consciously acquired skill of reading words has become fused to a large extent with the cognitive system for recognizing spoken words.

Many of the mechanisms for recognizing words are the same, whether they involve speech or written symbols. As you saw in Sections 7.1 and 7.2, both spoken and written language show similar effects of semantic priming and competition. Hence, the models for both spoken and written word recognition need to have similar ingredients such as spreading activation and inhibition. Moreover, parallels exist between the two modalities when it comes to asking whether there is top-down flow of information within the system, with activation spreading from words to sounds or letters. In the previous section, we saw that top-down activation could be very useful in solving the problem of perceptual invariance. While the problem of perceptual invariance is especially acute for *spoken* language, in cases where the shape of *written* letters is highly ambiguous, top-down connections may also help with the recognition of written words (see **Box 7.4**)

Just as there are specific challenges that arise for the understanding of spoken words, the same is true for words in their written form. In spoken language, the word recognition system has to confront issues that arise from the fleeting nature of speech in time. Recognizing written words, on the other hand, introduces some specific problems that involve mapping visual symbols onto spoken linguistic units. In this section, we'll explore some of the cognitive implications of recognizing words through the medium of visual symbols rather than sounds (or gestures).

BOX 7.4
Reading chicken scratch

In Section 7.4, I emphasized the ways in which spoken language is quite different from written language—sounds "smear" together, and different people make the same sounds very differently. Written text obviously doesn't present these problems. That is, unless you're trying to read someone's cursive handwriting. In that case, letters do change shape somewhat depending on their neighbors, and of course, the handwriting of one person can be very distinct from the handwriting of another. Writing that's produced by the human hand begins to look a bit more like speaking that's produced by the human mouth—the individual letters, like individual sounds, are not always easy to identify with confidence (see **Figure 7.13**).

The perceptual problem of reading chicken scratch is a lot like trying to recognize a spoken word whose sounds are indistinct, so it's possible that it could be solved by recruiting the same mechanisms that boost the understanding of blurry spoken words. For example, we saw through the Ganong effect and phoneme restoration effect that top-down information flowing from the word level to the sound level can help stabilize the perception of a sound that's on shaky ground acoustically. So, sounds that seem to fall in between a good /t/ and /d/ might be heard differently depending on the words they're embedded in. When reading *printed* text, the flow of bottom-up information would normally be highly efficient because the text doesn't vary that much visually, so you might think that there wouldn't be much need for links going top-down from words to letters. But some models of word recognition, such as TRACE, adopt the same architecture for both written and spoken language for just those cases where the visual identification of letters could be challenging—you can imagine this might happen at a distance, for example, or if a letter is partially obscured.

Top-down processing could be especially useful when people are faced with the perceptual puzzle of deciphering cursive writing. Anthony Barnhart and Stephen Goldinger (2010) designed an experiment to test whether semantic factors of a word would impact how quickly people recognized and named it. It's often been found, for example, that words whose meanings are easy to imagine visually (like *desk*) are recognized faster than ones whose meanings are less visually grounded (like *pain*). This suggests that when the meaning of a word is really easy to access, it becomes active more quickly, with that activation possibly feeding down to the lower level. Barnhart and Goldinger replicated this effect, but found that it was quite a bit stronger when people read words in handwritten cursive script than when they read the same words in a computer-generated font. The story seems to be that when the bottom-up flow of information from the letters to the word level was made more difficult, people relied more heavily on their knowledge about the words.

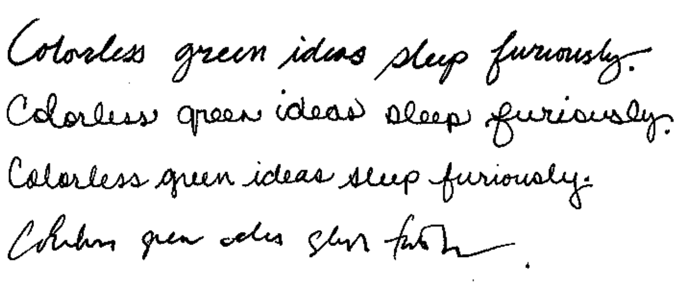


Figure 7.13 Handwriting samples by different writers, showing different renditions of the sentence *Colorless green ideas sleep furiously.*

Diversity in writing systems

If you were going about inventing a new written language to piggyback on the spoken one you already have, how would you do it? Obviously, the general idea is to create a set of symbols that match up to spoken linguistic units. But which units? Words? Syllables? Phonemes? All of these are reasonable strategies, each with certain advantages and disadvantages.

Words—or perhaps those smaller units of meaning, **morphemes**—might seem like a good starting point, but they have one obvious drawback: there are a *lot* of them. This means that learning to read would entail having to memorize possibly tens of thousands of symbols. This is not as impossible as it may sound. In fact, it's a feat that's in some ways less remarkable than what any human child has to do with a spoken language. If children can store in memory tens of thousands of sound patterns, *along* with discovering their meanings, surely they are capable of mastering the same number of visual patterns and mapping these onto already known meanings. Indeed, this type of **logographic writing system**, in which symbols are mapped to units of meaning, was adopted for some languages, such as Chinese. And when you use numeric symbols such as *1* or *7*, you're essentially using a logographic system, in which the symbol maps directly to a concept rather than being determined by how the word sounds. This can come in very handy in interactions between speakers of different languages. For example, even if you can't count in Italian, you as an English speaker can agree on the price of fish in a Roman market by resorting to the logographs that you and the vendor have in common in your writing systems.

Nevertheless, dropping down to a smaller size of linguistic unit can dramatically reduce the number of symbols you need, and this brings you into the realm of mapping symbols onto sound-based units. Some languages, like Japanese, have instituted a **syllabic writing system**, in which characters represent different syllables. This means that syllables like *ka* or *ki* might be captured by entirely different characters, ignoring the fact that the two units have the same first sound. A syllabic system works especially well for languages whose phonotactic constraints severely limit the shape of allowable syllables, and hence their number. For example, Japanese, which is limited to syllables that consist of one consonant followed by one vowel, can get away with fewer than 60 characters, whereas English, which allows varied pile-ups of consonant clusters at either end of a syllable, would need many more.

Many writing systems, including English, are based on an **alphabetic inventory** in which the goal is to map characters onto individual sounds or phonemes. This approach uses a conveniently small handful of written characters. But it's possible to go down to even smaller linguistic units—individual sounds, after all, are really clusters of articulatory features (see Section 4.3), so a sound like /k/ can be described as a velar voiceless stop (*velar*: produced at the back of the mouth; *voiceless*: produced without vibration of the vocal folds; *stop*: produced by completely obstructing the airflow). Notice that in our own alphabetic system, we totally ignore the very real sound-based similarity between /k/ and /g/, both of which are velar stops and differ only with respect to voicing. In principle, a featural writing system could use different characters for each feature to convey this information. Some elements of this approach can be found in the Korean alphabet, in which the strokes that make up letters have some systematic relationship to phonetic features.

Depending on your own language background, an alphabetic system may seem to be the most "natural" choice for a writing system to you. But it's not without its own artificial aspects and disadvantages. For example, it's not very natural at all for us to think (at least to *consciously* think) about speech in terms

morphemes The smallest bundles of sound that can be related to some systematic meaning.

logographic writing system Writing system in which symbols are mapped to units of meaning such as morphemes or words rather than to units of sound.

syllabic writing system Writing system in which characters represent different syllables.

alphabetic inventory A collection of orthographic symbols that map onto individual sounds or phonemes.

phonemic awareness The conscious recognition of phonemes as distinct units, usually only solidly acquired by individuals who are literate in an alphabetic writing system.

onset The material in a syllable that precedes the vowel.

rime The material in a syllable that includes the vowel and anything that follows.

of individual sounds. **Phonemic awareness**, or overt recognition of phonemes as distinct units, isn't something that just spontaneously happens at some point in development. We usually have to be forced to think that way as part of learning an alphabetic writing system. By contrast, conscious awareness of words, or even syllables, or even parts of syllables, comes more easily.

Until they're literate (in an alphabetic language), even adults do badly on phonemic awareness tests in which they have to pull words apart into their individual sounds. These tests might include questions such as, "How many sounds are there in the word *bed*?" and "What are you left with if you take the first sound away from the word *spring*?" and "What sounds do the words *bag* and *bed* have in common?" Preliterate children and adults find these quite challenging, and have an easier time with questions that probe knowledge about syllables and their internal structure, such as, "How many syllables in the word *bicycle*?" or "What sounds do the words *fling* and *spring* have in common?" (This last question tests for the separation of a syllable into an **onset**—the material in a syllable that precedes the vowel—and a **rime**—the material that includes the vowel and anything that follows. Evidently, syllable onsets and rimes are more cognitively accessible than individual phonemes.) Hence, one of the ironies of devising a writing system is that the smaller the linguistic unit you choose to map characters onto, the less cognitively accessible these units tend to be. Of course, the *larger* the linguistic unit, the more of them there are to be memorized! **Box 7.5** explores some of the implications of the different cognitive demands imposed by different writing systems.

In practice, almost no language adopts just one of the above writing strategies in a "pure" way. For an assortment of reasons—cognitive, practical, political, and historical—writing systems tend to emerge as hybrids of these different mapping approaches. For example, Chinese leans heavily on a logographic system, in part because of China's imperial history, in which a unified written language was created that could be understood among all the different dialects spoken across that vast country; a single writing system for multiple dialects or languages is only possible if it maps onto meaning rather than onto the finer-grained aspects of sound. This feature made the Chinese writing system a viable cultural export into other linguistic groups. Japanese writers incorporate logographs of Chinese

BOX 7.5
Do different writing systems engage the brain differently?

If you compare alphabetic languages like English with logographic ones like Chinese, it's hard to think of "reading" as involving the same set of skills in both cases. Learning to read in each of these languages introduces a new set of cognitive problems that don't apply to the spoken languages themselves. The writing system for English forces learners to consciously decompose words into phonemic units, a skill that's not necessary for speaking. On the other hand, the Chinese writing system doesn't require much phonemic awareness of its readers, but it does require them to be able to differentiate and memorize a very large number of minutely differing visual symbols, or *radicals* (see **Figure 7.14**).

The different demands of the two systems are evident in the way that reading is taught to children. Chances are that a child's first steps in learning to read English involved reciting the alphabet and learning to recognize individual letters and their corresponding sounds. And as it happens, the child's level of phonemic awareness is a strong predictor of her reading ability. In contrast, Chinese children spend many hours learning to *produce* the intricate symbols of their written language, and children's reading ability in Chinese is more strongly related to character copying than it is to phonemic awareness (Tan et al., 2005).

These differences lead straight to the question of whether reading involves the same patterns of activity in the brain for the two languages. For example, perhaps Chinese readers don't really activate the sounds of words, since in many cases, it's possible to get straight to the word's meaning without "sounding out" the symbols. Charles Perfetti and his colleagues (2010) reviewed a number of fMRI studies of reading in Chinese and in English and drew several conclusions. First, there are some similarities in the brain regions that are involved in reading both languages. Both languages rely on a reading network in the brain that connects visual areas (in posterior parts of the brain) with phonological areas (in temporal-parietal and anterior areas) and areas for meaning (the inferior frontal gyrus). That is, even for a writing system like Chinese, the phonological system is not bypassed during reading—the visual symbols are still well connected to information about words' sounds as well as their meanings. Nevertheless, there are some interesting differences, as shown in **Figure 7.15**. Specifically, while reading is very strongly lateralized in the left hemisphere for English readers, there's more bilateral activity in the visual areas for Chinese readers. This may be related to the added visual burden for learning Chinese characters. In addition, Chinese readers showed activation over a larger frontal area than English readers, but *reduced* activity in temporal areas that play a role in matching graphemes to phonemes.

报 打 搬 拉 把 捕

Figure 7.14 In addition to making fine-grained distinctions among characters, Chinese readers need to be able to recognize how basic symbols, known as radicals, can be combined and elaborated. The Chinese examples shown here all incorporate the radical for *hand* (the first symbol in each character).

BOX 7.5 *(continued)*

What happens when a reader, who's developed an efficient network for reading Chinese, learns to read English, or vice versa? Does a person's first-language reading network become recruited for the task of reading the second language? And if so, does this affect reading performance in the second language?

A study led by Jessica Nelson (2009) suggests that it depends on which writing system you start out with. Evidence from fMRI studies showed that Chinese subjects who learned English as a second language were able to use the same reading networks for both English and Chinese. But the reverse wasn't true; English speakers who learned to read Chinese showed different patterns of brain activity for the two languages. The authors suggest that it may be possible to read English as if it were Chinese—that is, by recognizing whole words and matching them to word meanings, rather than by sounding out the individual phonemes of a word. However, since using a "sounding out" strategy for reading Chinese is not really viable, English learners of Chinese would have been forced to develop new skills for reading.

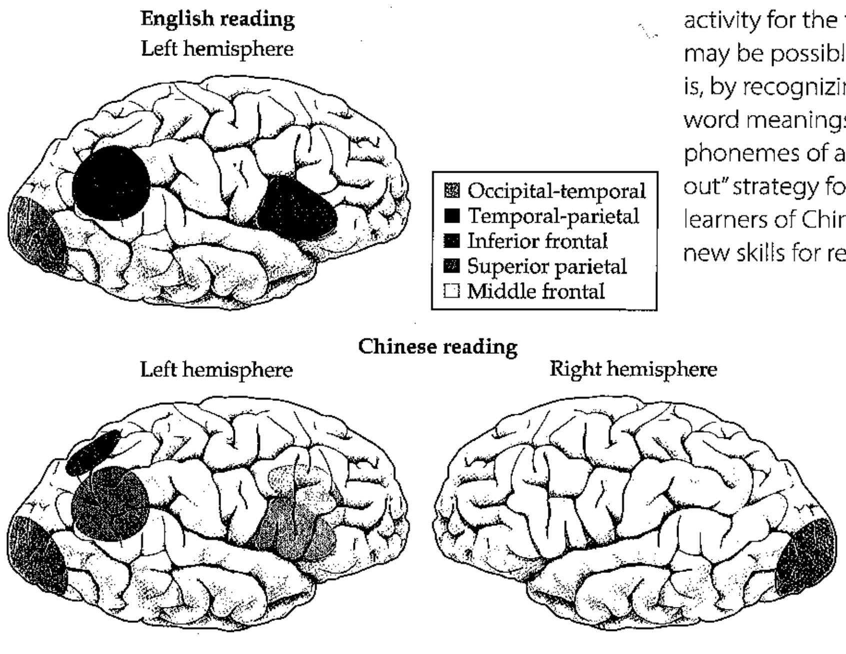


Figure 7.15 A rough diagram of English and Chinese reading networks, as identified by Perfetti et al. (2010). Chinese readers show more bilateral activity in occipital-temporal regions (green) and activity in the left middle frontal gyrus (blue). However, Chinese reading involves reduced activity in inferior frontal areas (purple) and in temporal-parietal regions (orange).

WEB ACTIVITY 7.9

Inventing a writing system Your own literacy training has a big impact on how easily you learn new writing systems. In this activity, you'll try your hand at inventing and decoding different types of writing systems.

origin into their orthographic toolkit, along with their own, more sound-based, syllabic script. Nevertheless, because of how hard it is to memorize these symbols, Chinese writing also uses some systematic markings that represent the pronunciation of the characters, to help reinforce the mapping between characters and words.

You might have noticed while reading about these different writing systems that, unlike spoken language, written language doesn't always exhibit the property of *duality of patterning*. That is, in spoken languages, sounds that have no intrinsic meaning are combined to form larger, meaningful units like morphemes and words. But in logographic writing systems, the smallest units of combination *do* have intrinsic meaning. Given that, as far as we know, duality of patterning is universal in spoken languages (see Box 2.1), this gives rise to an intriguing contrast between spoken and written languages. What could explain the difference? Is it that humans have certain innate expectations about the fundamental nature of spoken languages, but lack such expectations about the design of writing systems? Or is the difference due to the fact that different communicative pressures and constraints exist for spoken and written languages? At the moment, these questions are wide open.

Alphabetic irregularities

English, like all European languages, uses an alphabetic writing system. Of course, try telling someone who's grown up writing a *real* alphabetic language (like Finnish or Spanish) that English maps individual symbols onto sounds. For instance, how is the *sound* /k/ represented? Well, it depends on where it falls in a syllable, and which letter comes before or after it. For example: *cat, cot, cut* but *kiss, kept;* however, *pick, back,* but *meek.* Judging by these examples, it seems that it would sometimes be more accurate to say that *sequences* of letters map onto *sequences* of sounds rather than that individual letters map onto individual sounds. To make matters worse, sometimes how the sound is spelled depends on the individual word, possibly reflecting a long-ago borrowing from another language—you can't figure out how it's spelled from any rules, you just have to *know*: as in *click* versus *clique; tick* versus *tic; like* versus *psych* (incidentally, *which* sound is the letter *p* representing there?).

The irregularities of English spelling have drawn much comment over the years, at times humorous, or vitriolic, or simply resigned. The writer George Bernard Shaw, who was an advocate of reforming English spelling to make it more alphabetically transparent (see Language at Large 7.3), remarked that there was nothing to prevent the word *fish* from being spelled *ghoti* (*gh* as in *enough, o* as in *women, ti* as in *nation*). So how do people cope with really idiosyncratic words like *beau, have* (in contrast with *gave, save, rave*), *thorough, sign, colonel,* or *pint* (in contrast with *mint, lint, glint*)? These examples just don't line up symbols to sounds in any reasonable way. If you think about how these words are read, it's probably not that different from how people represent Chinese logographic symbols—the entire pattern of letters has to be matched up with individual words and their meanings. This means that English, a so-called alphabetic language, also relies on mappings of symbols to larger linguistic units than sounds, at least for some of its words (and you just have to know which ones).

graphemes Written symbols, analogous to phonemes in spoken language; individual graphemes may or may not correspond to individual phonemes (for example, two graphemes are used to represent the sound /k/ in *sick*).

Two systems for reading, or one?

The messiness of the English orthographic system has led researchers to suggest that the ability to link written symbols, or **graphemes**, with meaning—in other words, the ability to read English—can't be accomplished by any single cognitive route. One of the most influential models of reading, the **dual route model**, was developed by Max Coltheart and colleagues (1993, 2001) and proposes that there are at least two pathways that link graphemes with meaning. One is the **direct route**, in which a series of orthographic symbols is directly connected with the meaning of a word. The other pathway is called the **assembled phonology route**, in which graphemes are "sounded out" against their corresponding sounds, beginning at the left edge of the word. It's easy to see how having both routes available for reading would be extremely useful. The assembled phonology route allows us to read words that we know from spoken language but haven't yet mastered in print. And the direct route allows us to handle the messy exceptions that arise over time in a mongrel language like English. It might also be more efficient, once we've seen the same words in print repeatedly, to recognize words directly from their orthography without needing to fire up our phonological correspondence rules. In fact, it does seem to be the case that as people become more expert at reading, they rely less on the phonological route.

dual route model a theory of reading which proposes that there are two distinct pathways—the direct route and the assembled phonology route—that link written symbols (graphemes) with meaning.

direct route According to the dual route theory of reading, the means by which a series of orthographic symbols is directly connected with the meaning of a word, without involving sound-symbol correspondences.

assembled phonology route According to the dual route theory, the means by which graphemes are "sounded out" against their corresponding sounds, beginning at the left edge of the word.

But even highly skilled adults don't move entirely away from the phonological route, since there's some evidence that the sounds of words still have an impact even in a silent reading task when the focus is on recovering the meaning of the word. In one series of experiments by Guy Van Orden and colleagues (1987, 1988), subjects had to decide whether words that they saw on a screen were part of a larger category (for example, given the category of clothes, respond by pressing a "Yes" button for the word *shirt*, or a "No" button for the word *pear*). Sometimes, non-words flashed on the screen, and occasionally these non-words served as fake homophones for relevant words—for example, given the category flowers, the non-word *roze* might come up. If people are automatically sounding out the target stimuli, we might expect that when these targets sound just like actual words that fall into the flower category (*rose*), there might be some confusion, causing people to be slower to respond "No" to the targets. This was true for fake homophones (*roze*)—and it was also true for real-word homophones (*rows*), suggesting that even familiar words were being converted into their sound patterns and causing confusion when they were pronounced the same as the word *rose*. However, orthographically similar words and non-words (*robs, rone*) did not slow down the response times, indicating that the source of the confusion really was at the level of sound.

What happens when the two routes clash? Sometimes, the grapheme-to-phoneme correspondence rules conflict with the right way to pronounce an exceptional word—as in the word *pint*, which is a black sheep among the words *lint, mint, hint, glint,* and so on. In that case, the pronunciation by regular rule can interfere with the correct but irregular sound pattern, causing people to make mistakes when reading the word aloud, or to read it more slowly. The interference is more striking for less frequent words like *pint* than it is for very frequent words like *have*, which suggests that the two routes compete with each other to produce the right output, but that a heavily practiced word can win decisively along the direct route.

The dual route model accounts quite nicely for the split-personality nature of the English writing system. But it's not the only model that's been proposed to account for it. You may remember the great words-versus-rules debate at the end of Chapter 5. In that debate, some researchers claimed that we have two very distinct systems for forming plural or past-tense forms such as *cats* or *walked* in contrast to irregular forms like *ate* or *children*. According to this two-systems view, the regular forms are produced by a general rule, whereas the irregular forms just have to be memorized as individual words. Sound familiar?

LANGUAGE AT LARGE 7.3

Should English spelling be reformed?

Not all writing systems are equally user-friendly. English, for one, has an unsettling amount of spelling irregularity, compared with most other alphabetic writing systems. Children who are lucky enough to learn to read in more transparent alphabetic languages (such as Greek, Italian, or Finnish) usually breeze through basic reading lessons more quickly than their English-speaking peers. How did English spelling get to be this unwieldy, and isn't it time something was done about it?

Remember that spelling, unlike speaking, is the result of conscious decisions to capture spoken language with a set of symbols. It would be nice if spelling had been decreed for English by people with a sophisticated knowledge of English phonology and the cognitive interests of readers at the forefront of their minds. But it wasn't. The Anglo-Saxons, linguistic ancestors of English speakers everywhere, simply adopted the Roman alphabet to correspond with the sounds of their own language—where the Anglo-Saxon language had extra sounds, sequences of letters sometimes came to be composed to represent single sounds (for instance, *th*, *sh*, and *gh*, which was then a fricative sound produced at the back of the throat). Early spelling wasn't particularly consistent, in part because writing skills were still very unsettled (think of the unstable spellings of a first grader who's just learned to write), but also because different dialects pronounced the same words somewhat differently, and there was no central spelling authority to insist that it all be done in the same way.

The Norman conquest of England added chaos to the already somewhat slaphappy approach to orthography. More scribes were trained in French and not especially concerned with preserving whatever Anglo-Saxon spelling conventions there might have been. They corrupted the alphabetic system with French conventions of spelling that were intended to honor the Latin origins of words rather than their pronunciation. For example, the word *doubt* has the letter *b* in it simply because it is borrowed from the Latin word *dubitare*, and this is also how we ended up with the letter *c* capturing the /s/ sound for certain words of Latin origin (for example, *city*).

To make matters worse, consider that spoken language changes over time. Written language, on the other hand, tends to be much more conservative. This makes sense. You can hardly speak to someone who isn't alive and present in the current day, so there's not much reason to continue to talk the way people used to talk decades or centuries ago. But a good bit of literate behavior involves reading texts by people who are long dead. When the 1500s saw a massive shift in the pronunciation of English words, the spellings of those words ended up being stuck in the past. For example, what we now call "long" and "short" vowel sounds, as in *beet* and *bet*, are really just different vowels. But they used to be the same, distinguished only by length, until a language change known as the Great Vowel Shift rearranged the entire vowel inventory of English.

So much for how we got here. Should we get out of this orthographic morass? There have been many passionate advocates for spelling reform throughout the ages, including George Bernard Shaw, who suggested starting from scratch with a brand new alphabet, and the *Chicago Tribune*, which, more modestly, took it upon itself in 1934 to begin using more transparent spellings of certain words such as *thru*, *agast*, *iland*, and *telegraf*. The changes didn't stick.

There are monstrous practical obstacles against spelling reform. Spelling reform can happen, even on a large scale,

LANGUAGE AT LARGE 7.3 *(continued)*

but there has to be some single accepted authority that drives the change. For example, when Mustafa Kemal Atatürk, the first leader of the Turkish Republic, decided to shift the entire Turkish language away from using Arabic script to the Roman alphabet used by Western nations, he had the concentration of power to do so. (And to his credit, he consulted linguists, educators, and writers in the process.) It was hard enough to accomplish this in a single nation. But the success of English as a global language spanning numerous countries, many of which have a psychic allergy to centralized control of any sort, makes it hard to imagine how significant spelling reform could ever be achieved.

The global nature of English raises another set of issues when it comes to spelling reform: whose variety of English should become the basis for the written language? Even within North America alone, there is a great deal of variation, with differences among regional dialects apparently growing over time, rather than shrinking. For instance, the dialect region of the Inland North, clustered around the Great Lakes area, is currently undergoing a dramatic shift in the pronunciation of its vowels, not unlike the Great Vowel Shift of long ago. A speaker from Chicago or Detroit might pronounce the words *busses* and *socks* much like speakers from other regions might say *bosses* and *sacks*. Should spelling reflect pronunciation in Chicago, or Boston or London or Australia or Hong Kong? Whichever we choose as the official standard, some language groups are going to get a spelling system that is more transparent than others. On the other hand, if we democratically allow every English variety to change spellings to make them maximally transparent, communicating in writing between members of different dialect groups could get tricky. (Raise your hand if you've ever turned on the subtitles while watching a movie in which the actors speak a strong English dialect you're unfamiliar with. If everyone were to adopt their own transparent writing system, this would cease to be a viable strategy.)

When a language is spoken by many people who also speak other languages, as English is around the world, this tends to accelerate the rate at which its sound patterns change. Even if English speakers everywhere did agree to settle on a particular dialect as the basis for their writing system, this dialect would inevitably change over time. Should the writing system keep pace with the changes in the spoken language, thereby maintaining its easy-to-learn transparency? Or should it stay more fixed, allowing readers to be able to read texts written by people who spoke an earlier version of that particular language?

In the end, pushing for reforms that create a dramatically more transparent writing system for all users of English might make it easier to learn to read in English. But it could ultimately cut readers off from being able to read the thoughts of English speakers from other places and other times.

WEB ACTIVITY 7.10

Predict the future of English orthography The rise of the Internet means that written text can now be broadly disseminated without the involvement of "gatekeeping" professional editors. In this exercise, you'll be prompted to look at some samples of Internet text that is not filtered through an editorial staff, and to notice common examples of non-standard spellings. Take a stab at making some predictions about what these examples might mean for the future of English orthography.

This seems a bit like the claim that you have one rule-based system for constructing the sounds of a word, and then another direct route for memorized links between sequences of letters and the words they represent. You might also remember that the claim of two separate systems was hotly contested by a competing connectionist approach. Well, that's the case here too, and the dual route model of reading has its own connectionist competitor.

For the connectionist model of reading (Seidenberg & McClelland, 1989) there's no either-or distinction between getting to a word through its sound-based rules or getting to it through memorized links between graphemes and meaning. These researchers have argued that even irregularly spelled words are not completely exempt from letter/sound pairings, as a pure dual route model would have you believe. For example, even for a word like *sign*, the letter *s* still matches up with the sound /s/ as does *n* with /n/. There really aren't, it turns out, any words like George Bernard Shaw's *ghoti* for *fish*. And you can find clusters of *similar* irregular words—*resign, benign, align*—much like you find clusters of similar irregular past-tense forms (for example, *sleep-slept, creep-crept, keep-kept*). So, smaller rules seem to be embedded among the exceptional forms. The claim among connectionists is that there is a continuum between the most idiosyncratic forms and the most regularly patterned, and that what look like rules are really just the very strongest patterns.

In this model, hearing a word activates a set of phonological units, which are indirectly linked to a set of meaning units. These connections have formed over time and reflect learned associations between sounds and meaning. Learn-

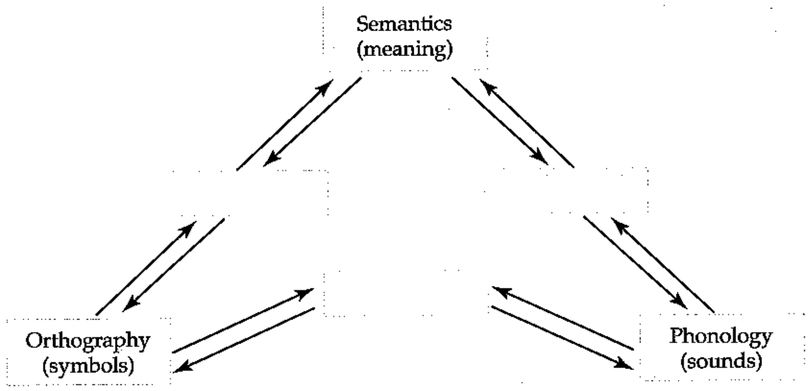


Figure 7.16 A single-system connectionist model of reading. Over the course of learning to read, connections between sounds (phonology) and meaning (semantics) are supplemented by a set of connections linking together orthography with phonology, and orthography directly with meaning. (Adapted from Seidenberg & McClelland, 1989.)

ing to read simply means grafting a new set of units and connections onto the existing sets for meaning and sound. As shown in **Figure 7.16**, a set of orthographic units has now been added to the system. These units are connected to both the meaning units and the phonological units, and the strength of these connections will depend on how often each set of units gets turned on.

For example, seeing the letters *p-i-g* while reading or hearing *pig* will turn on the phonological units /p/ /ɪ/ /g/. If units are turned on simultaneously very often, the system will establish a strong connection between them. For instance, reading the letters *w-i-g* will also turn on the /ɪ/ /g/ sound units. For very regular sound patterns, the same letters and sound units will be turned on very often, leading to more strongly weighted connections. In a connectionist system, very strongly weighted connections behave somewhat like rules. But the system can also form less regular, and hence slightly weaker, connections as well, for example, the correspondence between the letters *i-g-n* and the sounds in words like *sign* and *align*. For highly irregular words, the connections between orthographic units and *sound* units will be very weak. But there will still be connections between the orthographic units and the meaning units. These connections will be quite strong if the word is a very frequent one, like *have*, so the word can be recognized very efficiently on the basis of these robust links. The same system, then, can account for both the greater reliance on sound patterns for regular orthography, and the greater reliance on orthography-to-meaning links for irregular words. This is pretty much the same strategy taken by connectionist approaches to irregular morphology, and the claim is that it organically arises out of the way in which we generally learn patterns and associations.

The argument between these two models boils down to whether there are really two very distinct reading systems, or just one that makes use of connections between different subsystems. It's been a difficult argument to settle decisively with the usual set of lab tools that measure behavior, and it may be that careful brain-imaging work will ultimately help to settle the debate. But at the very least, the connectionist model suggests that it's possible to model many aspects of actual reading behavior without resorting to two very different reading routes.

GO TO
sites.sinauer.com/languageinmind
for **web activities, further readings, research updates, new essays,** and other features

DIGGING DEEPER

The great modular-versus-interactive debate

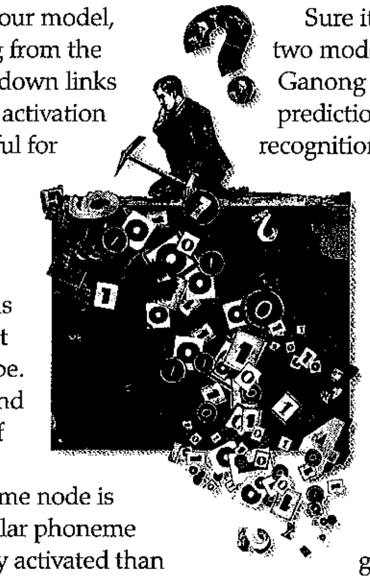


Figure 7.12 introduced an innovation to our model, adding top-down excitatory links going from the lexical level to the phoneme level. The top-down links allows activation of a specific word to send activation down to its component sounds. This is useful for capturing phenomena like the Ganong effect or the phoneme restoration effect, where the identity of the word alters the perception of individual sounds.

Or does it? Some researchers (e.g., Norris et al., 2000) have argued that these apparent top-down effects aren't what they seem to be. Think of it this way: the phoneme nodes send along information to the lexical level, sort of like reports being sent up by employees to higher management; the job of each phoneme node is to report on how likely it is that that particular phoneme was uttered, so nodes that are more strongly activated than others send up more convincing reports. Let's suppose that, going solely on the acoustic input, two phonemes, /t/ and /d/, send up equally convincing reports (are equally activated) based on a somewhat fuzzy example of a particular consonant sound—this is what happens in your typical Ganong-style experiment. Next, let's suppose that this fuzzy sound appears in the word frame *spi__*. Our current model with top-down links built into it would be the equivalent of this scenario: A higher-level manager gets conflicting reports, integrates these with his own knowledge, and comes to the conclusion that only one of these phonemes is consistent with all the information he has. So he talks to the employees at the lower level and convinces them that what they heard was wrong, and they alter their reports accordingly and send these back up. But, an alternative scenario that *doesn't* involve top-down nodes would go something like this: The manager gets conflicting reports, determines that only one of these fits with additional information that his employees didn't have and, as a result, makes an executive decision about which of these reports is right. He acts *as if* the reports from his employees had been altered to fit his own knowledge, incorporates that into his own report, and sends that up to yet a higher level of management. In this version of the model, the manager doesn't go back to try to convince his employees that one of them made an error; he simply decides that this must be the case.

Both accounts get the facts right. But one of them implicates top-down links, while the other has a strict bottom-up flow of information, with enhanced decision-making powers at the lexical level. Does it matter which one is right?

Sure it does, for the simple reason that although the two models align with the same facts about the basic Ganong effect, they might actually make different predictions about more nuanced aspects of word recognition and speech processing. And obviously, scientists want to get a really detailed and refined understanding of the processes they're studying. After all, this is one reason for making models in the first place—to push our understanding to greater and greater levels of detail. But the difference between these two models goes beyond general scientific diligence and attention to detail; it represents a quite profound disagreement about fundamental aspects of how the mind is designed, something researchers call **cognitive architecture**.

Researchers who propose top-down links are generally advocating an **interactive mind design** in which higher, more abstract levels of knowledge (usually what we think of as "more intelligent" levels of knowledge) can directly inform lower-level perception. Let's push this idea to an unlikely, extreme level. Imagine you're watching a horror movie in which spiders are crawling over the heroine's entire body. You begin to feel crawling sensations on your own skin, even though there is nothing making contact with it, and therefore nothing to be picked up by the sensory nerves in your skin. A top-down account would say that your suggestive higher-level thoughts are actually causing the nerve cells in your skin to send signals back up through your nervous system that are identical to the signals that would be sent in response to feeling spiders crawling over you.

This is wildly implausible, and if we recorded the activity of individual nerves, we'd find no such thing. Rather, your mind is creating its own interpretation based on other incoming information from the movie, the end result being that you have a subjective experience *as if* your nerves were being directly stimulated. But though your mind has convinced *you*, it hasn't convinced the nerve cells in your skin that you're being crawled upon. So there are obviously limits to how far down information can flow from higher levels.

cognitive architecture Fundamental characteristics of the mind's structure that specify how different cognitive components interact with each other.

interactive mind design A view of the mind's structure in which higher, more abstract levels of knowledge (usually what we think of as "more intelligent" levels of knowledge) can directly inform lower-level perception.

Some researchers have argued that in fact the mind is designed in such a way that the higher levels of processing *never* converse with the lower levels—that their job is simply to integrate information, interpret it, and pass that on to even higher levels, which in turn won't converse with them. Under this **modular mind design** view, employees at each level do their jobs without consultation or interference from above. One rationale for this particular cognitive architecture is that it's likely to be highly efficient. If lower levels of processing can do their work by restricting attention to only a very limited amount of information, they can prepare their reports quite quickly. But if they have to take into account the opinions and information from higher levels of processing as well, this might slow everything down. Proponents of a modular architecture (most prominently, the philosopher Jerry Fodor; see Fodor 1983) have argued that there's a striking difference between the type of information processing accomplished by lower-level modules, and the type of processing that does higher-level integration and makes decisions. The argument is that low-level perceptual processes are more like extremely fast reflexive instincts, while higher-level processes are more like slower, thoughtful judges.

The distinction seems intuitively appealing, in part because perception can be fairly stubborn—*reflex*-like—even in the face of compelling higher-level knowledge. For example, you've likely already seen the famous Müller-Lyer illusion involving lines of identical length that nevertheless look to be different in length (**Figure 7.17**). You probably knew already that the lines in this image were the same length, and if you didn't, you're invited to measure them now. But this knowledge has stunningly little impact on your perception of these lines. Your visual perceptual system, acting on a specific set of visual cues, simply cannot be reasoned with.

Such a bullheaded perceptual system might seem perverse in the case of illusions like the Müller-Lyer effect (after all, isn't it preventing you from seeing what's *really* there?), but the argument is that in fact, we'd be subject to a great many more hallucinations if our higher-level processes were allowed to readily interfere with the output of lower-level perception. In getting back to word recognition, for example, wouldn't top-down feedback render us incapable of noticing other people's mispronunciations and slips of the tongue? These are clearly inconsistent with our higher-level knowledge of the words they're attempting to pronounce, and usually we know exactly what they're trying to say, yet we persist in hearing the actually produced sounds.

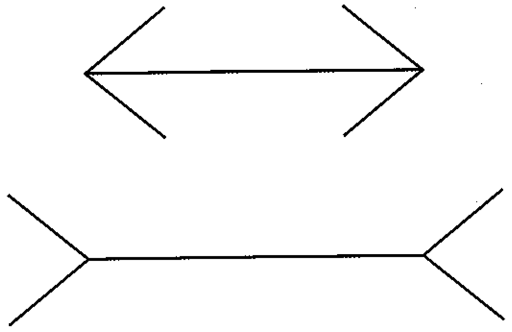


Figure 7.17 The Müller-Lyer illusion. No matter how much top-down certainty you have that the two red lines are actually the same length, lower-level visual processes still insist that they're different.

From a broader view of the field of psycholinguistics, then, the decision about how to model lexical effects on phoneme perception isn't just about making more precise predictions. The issue has become a test case for these two big, competing views of how the mind is designed. (You'll see in the next chapter that these competing views have far-ranging implications for aspects of language comprehension that go well beyond the realm of words.)

So, is there a way to decide which of these views is right? Experimental efforts in the domain of word recognition have focused on a particular phenomenon known as **compensation for coarticulation**. As we've seen, sounds are often affected by other sounds they're pressed up against. Hearers automatically adjust for this. For example, if they hear a sound halfway between /t/ and /k/, they'll be more likely to treat it as a /t/ if it follows the sound /ʃ/ (this symbol corresponds to the first sound in *ship*) and to treat it as a /k/ if it follows the sound /s/. This is because the shape of your tongue when you pronounce the second consonant is affected by where you place your tongue for the first consonant; a /t/ normally sounds a bit /k/-like following /ʃ/, and a /k/ normally sounds more /t/-like following /s/, so when hearing a sound that's in between, hearers automatically take into account the articulatory context, to decide which one is most likely. All of this is a low-level process, and affects the activation of sounds whether or not any words are involved at all—for example, even if the sounds are embedded in nonsense words. So it seems that compensation for coarticulation involves information that's readily available at the sound level, not information that's coming down from the word level. To return to our metaphor, in order to incorporate this information into their reports, the employees at the sound level talk to their neighboring sounds, rather than talking to their bosses at the word level.

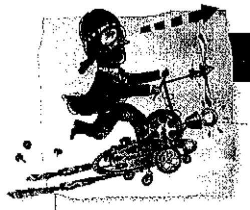
modular mind design View of the mind's structure in which higher levels of processing never directly influence the lower levels; instead, the higher levels integrate information based on lower-level processes, interpret it, and pass these interpretations on to even higher levels.

compensation for coarticulation Phenomenon in which the perception of speech automatically adjusts to take into account the tendency for sounds to be pronounced differently in different phonetic environments; thus the same ambiguous sound may be perceived differently, depending on the adjacent sounds.

But we can use this phenomenon to test whether word representations are able to alter the reports of their sound-employees from the top down (rather than simply making an executive decision based on their levels of activation). To do this, we can leverage the compensation for coarticulation effect together with the Ganong effect. In this more complex experimental scenario, research subjects hear words like *Christma?* or *fooli?* in which *?* is a sound halfway between /s/ and /ʃ/, followed by the word *?apes*, in which *?* is a sound halfway between /t/ and /k/. In other words, the second word could be interpreted as either *tapes* or *capes*. We know that compensation for coarticulation would make people more likely to hear *capes* after *Christmas* and *tapes* after *foolish*, and that this is a result of influence from neighboring sounds. But in this case, they're hearing exactly the *same* sound embedded in the word *Christmas* or *foolish*. So we would expect that people would report hearing *capes* just as often after *Christma?* as after *fooli?*—*unless*, that is, the word-level representation was able to change the activation levels of the sounds themselves so that these sounds could in turn influence their neighbors. In that case a more highly activated /ʃ/ sound in the frame *fooli?* should bias the hearer into thinking she is hearing *tapes* rather than *capes*, just as if the /ʃ/ sound had been activated on the basis of the acoustic input.

This subtle logic provides a way to wedge apart the predictions made by competing top-down and bottom-up explanations. Sure enough, results have been reported in favor of the top-down account, first by Jeff Elman and James McClelland (1988), proponents of the TRACE model of word recognition, in which top-down links figure quite prominently. But too much is at stake theoretically to let these results by without closer scrutiny. Mark Pitt and James McQueen (1998) came back with a challenge, arguing that there was still a way in which the results could have come about from a strictly bottom-up system. They pointed out that the identity of the ambiguous /s-ʃ/ sound could have been determined not by feedback from the word level, but simply because the last vowel in *Christmas* occurs much more often with /s/ than with /ʃ/ and, conversely, the last vowel in *foolish* occurs much more often with /ʃ/ than with /s/. So, the activation of the /s/ and /ʃ/ sounds could still have been determined solely by consulting neighboring sounds: the probability of certain sound sequences over others would bias the perception of either /s/ or /ʃ/, and this would in turn bias the perception of /t/ and /k/ in the following word, based on compensation for coarticulation. Hence, no interference was needed from the word level. (To revisit our office scenario, the employees' reports now are affected by discussions with their peers, rather than top-down instructions from their supervisor.)

Several papers have flown back and forth since then, either supporting or challenging Elman and McClelland's original results, and this particular test case for the modular-versus-interactive debate is still unresolved. But there's more than one way to skin a modular cat, and researchers have been attacking the question from a number of different angles. For example, they've looked at whether top-down processing can be found between other levels, such as whether higher-level syntactic or semantic expectations can drive up the activation of one lexical candidate over others, even when the others are equally consistent with sound-based information. In the next chapter we'll see whether cues from context and meaning can affect lower-level decisions about which syntactic structure to build during sentence comprehension. And discussions have occurred about whether a top-down architecture is consistent with what we know about how brains work. After all, the strongest version of the modular perspective—that higher levels of knowledge and information processing *never* feed back to influence lower ones—is a sweeping and interesting idea, generating predictions about a great many separate phenomena. Over the decades, the sheer bigness of the idea has inspired an impressive body of work, and has mobilized researchers from both camps to think of ever-more-compelling experiments to defend their claims.



PROJECT

Step into the modular/interactive battle arena! Generate a set of predictions that are made by a model that proposes the top-down flow of information. Flesh these predictions out by sketching an experimental design, describing in detail the experimental conditions, stimuli, and methods. Suppose that your results bear out the predictions made by the top-down model. Try to identify any way in which these results might be open to attack by proponents of modular, bottom-up models of word recognition.