Accessing the Mental Lexicon

Words, and how we (eventually) find them

The history of science is littered with examples of analogies that do not work. Often, they are simply inappropriate, simply wrong, or simply confusing. But even when inappropriate, they can prove useful. For instance, it is not unnatural to think of our knowledge about the words in our language as residing in some sort of dictionary. The Oxford English Dictionary (OED), all 20 volumes of it, is as good an example as any—its purpose is to provide, for each entry, a spelling, a pronunciation, one or more definitions, general knowledge about the word itself, and perhaps a quotation or two. Getting to this information is relatively efficient. You scan down the page, ignoring the definitions of the words you are uninterested in until you come to the word you want. Of course, if you are lucky enough to possess the dictionary on CD-ROM, or to have on-line access over the internet, you do not even need to scan down the page (let alone pull the appropriate volume off the shelf); just type in the word, and up pops everything you ever wanted to know about it. But so what? Why should anyone care about how the different incarnations of the OED work? In the last chapter we saw that the analogy between accessing a written dictionary and accessing the mental lexicon is at best fragile. So why carry on with it? The answer to this is simply that it provides a useful starting point from which to proceed, using a vocabulary that is easily understood to describe a process (accessing the mental lexicon) that is easily misunderstood.

Describing the mental lexicon as 'something like you've never imagined' is probably accurate, but certainly useless. At least a conventional dictionary can be imagined, and is therefore a useful place from which to start our exploration of the mental equivalent. Most importantly of all, the questions one can ask of a dictionary such as the *OED*, and the questions one can ask of the mental lexicon, are remarkably similar. The answers, though, can be surprisingly different.

Before delving into the mental lexicon and looking at how we retrieve words, we need first to address an important, if basic, question: what, exactly, are these things we call 'words'?

Words and what they contain

The purpose of language, and communication in general, is to convey meaning. In spoken language, the word is the smallest stand-alone thing that can do this. It is not, however, the smallest verbal gesture capable of expressing meaning. An 's' added onto the end of 'fact' also expresses meaning; namely that we are dealing with more than one fact. So words can generally be broken down into even finer units, called **morphemes**. This last sentence contains 13 words, but 21 morphemes — the word 'units', for instance, consists of the morpheme 'unit' and the morpheme 's', and the word 'morphemes' consists of 'morph'+'eme'+'s'. Figuring out which kinds of morpheme can be stuck onto which other kinds, and how this affects the meaning of the resulting word, has been studied within a branch of linguistics called *morphology*.

words The smallest units of sound which can both stand alone and convey meaning.

morphemes The smallest units of sound that can convey meaning ('words' is made up of the morpheme 'word' and the morpheme '-s', indicating more than one word).

Stems and affixes

There are different kinds of morphemes: **stems** (e.g. 'unit', 'word', or 'speak') and **affixes** (e.g. '-s', or '-ing'). Languages differ in terms of where they put their affixes — in English, the most common affix is a *suffix*, coming at the end of the word, but we also have *prefixes* which come at the beginnings of words (e.g. the 'pre-' in 'premature'). Some languages (for instance, Tagalog, the language of the Philippines) have *infixes* too; these are affixes that are inserted in the stem. Mark Aranoff, a morphologist, has an entire section in his book *Word formation in generative grammar* devoted to one of the few English infixes — 'fuckin' as in 'fan-fuckin-tastic' and 'Missi-fuckin-ssippi' (if you need a diversion, work out why neither 'fantas-fuckin-tic' nor 'Mississi-fuckin-ppi' sound right!). But affixes are not the only device we can use for modifying the meanings of words; the irregular past tense in English — 'run–ran', 'speak–spoke', and so on — is a remnant of a time when the past tense was produced not by adding the suffix '-ed' onto words, but by modifying a vowel in the stem. In a Semitic language like Hebrew, this is the rule, rather than the exception.

stem A form of a word to which affixes can be added to create new words.

affix A morphological component of a word other than the stem which contributes either to its grammatical form (inflectional affix) or to its core meaning (derivational affix). Affixes can occur as prefixes (at the beginning of a word), suffixes (at the end), or infixes (in the middle — although English is not considered an infixing language).

Inflectional and derivational morphology

To complicate matters further, different kinds of affix do very different things. Some affixes are called **inflectional**; these include the plurals ('-s'), and the various ways of inflecting verbs (e.g. for 'govern': 'governs', 'governing', 'governed'). Inflectional affixes do not change the meaning of the word, but convey additional information relevant to it. **Derivational** affixes, on the other hand, do change the meaning; they are used to derive new words, so from 'govern' we can get 'governor', 'government', 'governance', 'governable', 'ungovernable', 'governability', and so on. Although related in meaning, each of these words means something different, and in the case of 'governable' and 'ungovernable', they mean exactly the opposite. But not all derived words are related; 'casual'

inflection A change in the grammatical form of a word which does not affect its core meaning: the noun 'words' is an inflection of the noun stem 'word' and indicates the number of words; 'talking' and 'talked' are inflected forms of the verb stem 'talk' — all three are verbs but they differ in respect of when the talking happened. In neither case is the core meaning or the grammatical category changed (although 'talking' can also be a noun — 'their talking annoyed me' — but in this case it is a *derived* form of the verb 'talk' – see below). Not all inflections are formed by adding an affix: 'spoke' is an inflected form of 'speak'.

derivation A change in the form of a word which results in a new lexical meaning: 'speaker' (noun: a person) is a derived form of 'speak' (verb: an action), as is 'unspeakable' (adjective: a description).

and 'casualty' are unrelated in meaning, as are 'depress' and 'express' (although a glance at the *OED* will reveal their common historical ancestry). Another complication (there are several more) is that although the majority of stems can be free-standing, there are some inflected words which do not contain a free-standing stem, such as the verbs 'permit' and 'submit'. And whereas the meaning of 'ungovernable' can be deduced by stripping away the affixes and recovering the meaning of the stem 'govern', the meaning of 'permit' and 'submit' cannot be deduced from the meaning of 'mit'.

So words are complicated things. And knowing whether something should be called a word or a morpheme, an affix or a stem, a prefix or a suffix, an inflection or a derivation, matters far less than knowing that words have an internal structure. Somehow, as we hear or read a word, we have to strip off the excess (but important) baggage and reveal that word's core. And sometimes it looks as if we ought to do this, but in fact we should not (as in 'permit', 'report', and so on.). Linguistics has told us an enormous amount about how words are structured and how the meaning of a word is dependent on the meanings of the different morphemes it is composed of; it tells us which kinds of morpheme can be combined with which other kinds, and in which order. But that is just the periodic table again—it tells us what the result is, but it does not really tell us how the result comes about. It does not explain how the brain comes to acquire the conventions that tell us that 'un-' on the front of a word results in a meaning that is the contrary of whatever the word meant in the first place. It does not tell us where this knowledge is stored, or how it is stored, or how the brain takes a complex word like 'unspeakable' and breaks it down into its components, or even whether it does break it down at all. All these questions fall under the remit of psycholinguistics, from where some answers are beginning to emerge (we shall come across them shortly). Perhaps surprisingly, one starting point from where to consider these questions from a psycholinguistic perspective is to return, notwithstanding its inadequacies, to the dictionary metaphor once more: At the root of each word is a meaning, and representing this meaning, and the different forms that words can take, is precisely what a dictionary is for.

Of course, we knew all along that words convey meaning, and that the mental lexicon is a store of word meanings and (presumably) word forms. But so is the *OED*, and yet physically they could hardly be more different. So what, if any, are the consequences of the physical differences? The fact that they evidently are different does not mean that they are necessarily used any differently — for instance, the *OED* in book form could hardly be more different from the *OED* on CD-ROM, and yet there are aspects of their use which are common to both of them. Apart from the fact that the *OED* and the mental lexicon are physically different, what else is different about them?

WHY STEMS ARE NOT ROOTS

There is a subtle distinction between *stems* and *roots*. Whereas the stem is what you add affixes to, the root is what you get when you strip all the affixes off. In inflectional morphology the distinction is most usefully thought of in the context of languages such as French or Italian. The Italian for 'talk' (the verb) is 'parlare', with 'I talk' being 'parlo', 'you talk' 'parli', 'I will talk' 'parlerò', and 'you will talk' 'parlerai'. In this case, the root is 'parl-', and the stem for the purposes of the future tense is 'parler-' — it is onto this stem that the different inflections (for the 'I' and 'you' versions, for example) are added. So the stem can be a larger unit than the root, something that is relatively rare in English inflectional morphology (where they are often the same). In derivational morphology, the distinction is more easily captured in English: the prefix 'un-' can be added to the stem 'speakable' to create 'unspeakable', but this stem is itself composed of the root ('speak') to which the suffix '-able' has been added.

In a semitic language, the root of a word consists of the consonants between which different vowel patterns are inserted as infixes. Thus the root of a Hebrew word, like that of an Italian word, is not itself a word.

Accessing our dictionary

We already know that not all dictionaries are the same, and that depending on which dictionary we use, we can access the words (and narrow down the search) on the basis of how they are spelled, how they are pronounced, what they rhyme with, what they look like, how long they are (as in a crossword dictionary), or even how frequently they occur in the language at large (as in some specialized dictionaries used by psycholinguists). But crucially, however we do it (see Chapter 5 for some discussion of this), it is an inevitable consequence of accessing the dictionary that we will encounter, during the search, other words that share certain features with the word we are ultimately interested in finding, whether they share their spelling, pronunciation, rhyme, shape, length, or frequency. They may also share morphological structure ('unspeakable' shares its first four phonemes, its first four letters, and its first morpheme '-un', with 'unsparing', but it shares nothing in meaning other than that conveyed by that first morpheme). It is in respect of how we react to these brief encounters that our intuitions about what we do with a written dictionary are quite at odds with what we actually do with our own mental lexicon. For instance, although we do not burden our minds with the definitions of the other words that we pass as we scan down the page of a written dictionary, the same is not true of the process by which we access the mental lexicon. We do burden our minds with the contents of the neighbouring words we encounter as we narrow down the search. Our intuition that we do not is wrong, and our expectation on the basis of what appears to be a similar process (using the *OED*) is also wrong. The challenge, of course, is to *prove* that these intuitions are wrong, and to determine what should take their place.

It seems somewhat unreasonable to access the meanings of the words 'ram' and 'ramp' simply because they are encountered during the search for 'rampart'. It would be equally unreasonable for an Australian listening to the word 'acoustic' to access the entry for 'acubra' (a traditional Australian hat) just because they start off sounding the same, or for a naturalist to access the meaning of the word 'pichiciago' (a kind of armadillo) just because it starts off like 'pitch'. It would surely make sense only to look up the definition of the word being looked for, as we do with written dictionaries, and not to look up the definitions of all the other words that just happen to overlap in their first few sounds. So why, when searching the mental lexicon, do we access the meanings of these neighbouring words? And how can we, as psycholinguists, be so sure that this happens? As we shall see, it is unclear how things could possibly happen any other way.

The earliness of word recognition

During the 1970s, **William Marslen-Wilson** demonstrated that we can recognize a word even while it is still being heard (before, even, the speaker has fin-

Marslen-Wilson, 1975
Marslen-Wilson & Welsh, 1978

Marslen-Wilson, 1975 in ***Science*** People had to shadow (repeat) sentences played over headphones. Some were able to articulate words within 250 ms. of the words' onsets – well before the acoustic offset of the words they were listening to. Words which had been intentionally mispronounced on the original recording were repeated in their original (correct) form, but only when the sentence made sense (and it was not jumbled up — in the jumbled case, shadowing was considerably slower). An extended version of this study was reported later on by **Marslen-Wilson & Welsh, 1978** in ***Cognitive Psychology***, and it was here that some of the original assumptions made by Marslen-Wilson in his account of lexical access were most explicitly laid out. In this later study, all the sentences made sense (there was no jumbling), but the words containing the mispronunciations were either highly likely or relatively less likely. The fluent restoration of mispronounced words to their correct form occurred more often when the word was likely.

ished saying it). We therefore access the lexical entry of a word well before the corresponding physical stimulation has ceased (that is, before its *acoustic offset*). In one of the first demonstrations of this, people were asked to repeat aloud as quickly as they could what they heard over headphones (to *shadow* what they were listening to). Marslen-Wilson found that often they would start to vocalize a word before it had finished playing on the tape. This was not simply some blind repetition of the sounds they heard, because if the words were jumbled up so that they made no sense ('up words jumbled he they so no sense the made'), people could no longer shadow as fast — so they were clearly interpreting what they were listening to, and were therefore recognizing the individual words before repeating them. In other experiments, he asked people to press a button as soon as they heard a particular word on a tape (*word-monitoring*, similar in spirit to the syllable-monitoring task mentioned in Chapter 5). He found that once you took into account the time it takes to decide to press a button, and the time it takes to press it, people were responding so fast that they must have been initiating their response well before the end of the word.

Marslen-Wilson also found that the time it takes to recognize a word correlates very well with how much of the word has to be heard before it becomes uniquely distinguishable from all the other words in the language that share the same beginning. So 'slander' becomes uniquely distinguishable only when the /d/ is encountered (its **uniqueness point**). Before then, the input would be compatible with 'slant'.

Uniqueness point the point within a word at which there ceases to be any overlap with other words in the lexicon. The uniqueness point for 'captain' is at the /n/ – before that point, other words such as 'captive' overlap with 'captain', as does 'capsicum' when only 'cap' has been heard.

'Activation', not 'Access'

An important component of the account of lexical access developed by Marslen-Wilson is that the entries in the mental lexicon are not simply accessed, they are *activated*. The idea that information is activated has a long established history in psychology, although its application to word recognition became more wide

Marslen-Wilson & Tyler, 1980
Marslen-Wilson, 1984

Marslen-Wilson & Tyler, 1980 in ***Cognition*** One of many findings in this paper concerned the time it took to make a 'yes I hear it' response in a word-monitoring task. People responded, on average, before the acoustic offset of the target words, and it was estimated that the response must have begun within 200 ms. of the beginning of the word (corresponding to the first two or three phonemes). Response times were considerably slower when the sentence didn't make sense.

Marslen-Wilson, 1984 in ***Attention and Performance X*** People were asked to monitor for phonemes which occurred at different locations within a three-syllable word. The first finding was that phonemes towards the end of real words ('crucible'), but not nonsense words ('brucicle') were spotted faster than those towards the beginning. If people were responding purely on the basis of the sounds they heard, there would have been no difference in position or word-type — so this suggested that people waited to respond until they had recognised the word and recognised its component sound structure (why else would there be a difference between real words and nonsense words? — there could be no other reason). The second finding was that the distance between the position of the target phoneme (the one that had to be responded to) and the point at which the word became uniquely distinguishable from all other words (the uniqueness point) correlated highly with the time it took to spot the phoneme – the earlier the uniqueness point, the faster the monitoring time.

spread in the late 1960s following the work of John Morton, formerly Director of the Medical Research Council's Child Development Unit in London.

One way to think about this is to remember that ultimately, all the information in the mental lexicon is stored within the neural structures of the brain. When a pattern of light enters the eyes, or a sequence of sounds enters the ears, those stimuli do not access anything within the brain, even if they result in the recognition of, for instance, a politician speaking or a baby babbling (or both, if they are indistinguishable). Instead, the stimulation passes through the neural circuitry of the brain, being modified by, and in turn, stimulating (or activating) different parts of the circuit. Only certain kinds of stimulus will provide the right kind of stimulation for some particular part of the neural circuit — the stimulus is a key that can activate a part of the circuit, and depending on which part is activated, we experience 'seeing a politician' or 'hearing a baby'. There will be more of this later, but for now, the important point is that nothing is accessed; it is activated. And although we might just as well continue to refer to lexical entries, we shall return later to the idea that the mental lexicon is in fact a collection of highly complex neural circuits.

So what has this to do with why we access/activate the meaning of anything but the intended word? Why does this suddenly make it reasonable to suppose that we start to activate words and their meanings even before they become uniquely distinguishable from their neighbours? The answer has to do with the quite reasonable assumption that sounds entering the auditory system (i.e. the ear and beyond) stimulate the neural circuitry as they enter the system — a sequence of sounds is much like the combination to a safe; the tumblers in a combination lock fall into place as the correct sequence of rotations is performed, without waiting until the sequence is complete. Similarly, those neural circuits which require a particular sequence of sounds (before a particular word is 'experienced') will become activated as that sequence enters the system. So the neural circuits that encode what we think of as lexical entries could quite reasonably become activated on the basis of a developing (but not yet completely developed) sequence of sounds — /slan/ would activate the neural circuits associated with (and hence would activate the meanings of) both 'slander' and 'slant'. But so much for what is possible in principle. What actually happens? Where is the proof?

Multiple activation in lexical access

What we need is a way of establishing which meanings of a word have been activated, and when. The priming task (first mentioned in Chapter 1) does just this. The task here is to decide whether a word that has just appeared on a computer screen is a real word in their language (e.g. 'broom'), or a non-word (e.g. 'broam'). How long it takes people to make a response (a *lexical decision response*) depends on all sorts of things. First of all, non-words take longer to say 'no' to if they are similar to real words, and they take longer to say 'no' to than real words take to say 'yes' to (these are **non-word legality** and **lexical status** effects respectively). Real words that are used infrequently take longer to say 'yes' to than words that are used frequently (a **frequency effect**), and words which are similar to many other words take longer to say 'yes' to than words which are similar to few other words (a '**neighbourhood' effect**). But the recognition of a real word can also be faster if a related word has been seen beforehand— lexical decision times to 'broom' are faster following 'witch' than following the unrelated *control* word 'pitch'. This effect is called priming; 'witch' (the prime) can prime 'broom' (the target), 'doctor' can prime 'nurse', 'bug' can prime 'ant', and so on. Activating the prime causes the target to be activated faster. Conversely, if a target word is activated faster (primed), you can be sure that the priming word must have been activated.

Lexical status effect words are responded to faster in a lexical decision task than non-words (shown by **Rubenstein and colleagues, 1970** in *Journal of Verbal Learning and Verbal Behaviour*).

Non-word legality effect non-words that could not possibly be real words ('cmptcjn') are faster to respond to in a lexical decision task than real words ('captain'), which are faster to respond to than non-words that could in principle be real words ('coptein') —shown by **Stanners and Forbach, 1973** in *Journal of Experimental Psychology*.

Frequency and word recognition — A QUICK OVERVIEW

Words that are frequent in the language are recognized faster than words that are infrequent. **Forster and Chambers, 1973** in *Journal of Verbal Learning and Verbal Behaviour* found that high frequency words were named faster than low frequency ones, and Whaley, 1978 also in *Journal of Verbal Learning and Verbal Behaviour* found that high frequency words were responded to faster than low frequency ones in a lexical decision task. Although these studies explored effects of frequency on visual word recognition, **Marslen-Wilson, 1987** in *Cognition* demonstrated such effects on auditory word recognition. But recognition isn't just faster, it is better: **Savin, 1963** in *Journal of the Acoustical Society of America* found that high frequency words are better detected than low frequency words when embedded in a noisy signal.

Neighbourhood effects and word recognition A QUICK OVERVIEW

In a study of the recognition of spoken words, **Pisoni & colleagues, 1985** in *Speech Communication* showed that the number of similar words influences the time to recognise a target word — words with lots of neighbours are recognised *slower*, because of interference from their neighbours. This result proved controversial, with **Marslen-Wilson, 1990** in *Cognitive Models of Speech Processing* finding that competing neighbours do not influence recognition times. The controversy centred in part on the definition of a neighbour — Pisoni and colleagues defined a neighbour in terms of words differing only in one phoneme, no matter where that phoneme occurred (so 'walk' and 'talk' would be neighbours); Marslen-Wilson defined the 'neighbourhood' as changing as the word unfolds in time: a word will have many different neighbours early on, when little of it has been heard, but fewer neighbours later on (and none once its uniqueness point has been reached). The controversy also stemmed from the different methodologies that researchers use (Marslen-Wilson found that neighbourhood effects could come or go depending on the experimental technique). **Allopenna and colleagues, 1998** in *Journal of Memory and Language* monitored people's eyes as they looked at a board on which were placed various toy objects including, for example, a loudspeaker and a beaker. When told to 'pick up the beaker', people's eyes did (with some small but measurable probability) move towards the speaker — evidence of a (small but measurable) effect of the neighbour.

In the mid-1980s, a student of William Marslen-Wilson's, **Pienie Zwitserlood**, used a version of the priming task called *cross-modal priming* to explore when, during the sound sequence, words are activated. In cross-modal priming, the priming word is presented in the auditory modality, and the target is presented visually. Zwitserlood and Marslen-Wilson reasoned that if lexical entries are activated before the end of a word, and if this activation is all it takes to get priming to related words, it should be possible to find cross-modal priming effects when the visual target word is presented on the screen part way through the auditory presentation of the priming word. So people would hear only the first part of a word, and at the end of that part, a related word would be flashed up on the screen. The actual experiment was performed in Dutch in The Netherlands, but it translates very easily into English.

Zwitserlood used the Dutch equivalent of 'captain', and played people a recording of this word up to and including the /t/. At this point, the sound stopped, and a related word (e.g. 'ship') appeared on the screen. Sure enough, she found priming — the word 'ship' was responded to faster partway through 'captain' than partway through the control word 'wicket'. Of course, this simply shows that the lexical entry for 'captain' can be activated before the entire word has been heard. But the clever thing about this experiment was that 'captain' was not the only word compatible with the fragment played to people; 'captive' is just as good a continuation (the words can only be discriminated between on the basis of the final phoneme). And crucially, Zwitserlood also found priming to words related to these alternative continuations. In other words, the two alternatives that were compatible with the auditory input were both activated. And just to really prove the point, Zwitserlood demonstrated that if the visual targets were flashed up on the screen towards the end of 'captain' or 'captive' (by which point it was clear which word it was), then only the related target words were primed — there would be no priming to words related to 'captive' when presented towards the end of 'captain'.

Multiple activation and the frequency of the alternatives

So it looks pretty cut-and-dry; as the acoustic input enters the system, we activate all the lexical entries compatible with the input that we have heard so far. This is exactly what Marslen-Wilson's theory had predicted. And as the input becomes incompatible with certain alternatives, so those alternative entries (or **competitors**) begin to de-activate. But there were two further aspects to Zwitserlood's experiments that were important. First, recall that one of the determinants of response times in lexical decision experiments is the frequency of occurrence of that word in the language at large; the more common words are responded to faster than the less common words. And although the more common words tend also to be the shorter words, it has been shown that this is not simply a length effect — once length is held constant (something that most studies strive to do), it is still the case that more frequent words appear to be recognized faster

Competitor a word (or rather, its representation) which becomes activated on the basis of the acoustic input, but which is not the actual word being heard. Partway through 'captain', 'captive' and 'capsicum' are competitors. 'Capsicum' ceases to be a competitor at the /t/.

Lexical neighbourhood the set of competitors that are simultaneously active at any one point during a word. Competitors that are very similar to the actual word are often called *neighbours*.

Priming and word recognition A QUICK OVERVIEW

The first demonstration of priming was reported by **Meyer & Schvanaveldt, 1971** in *Journal of Experimental Psychology*. In fact, they each discovered the effect independently of one another — once they found out that the other had also discovered the effect, they decided to cooperate on the publication of a single paper. The basic effect is simply that lexical decision times are speeded up if the target word is preceded by a related word ('doctor' then 'nurse', and so on). **Neely, 1977** in *Journal of Experimental Psychology* systematically manipulated the delay between the first word (the prime) and the second (the target), and showed that the faciliitatory (advantageous) effects of a related prime occur even at short delays, whereas the inhibitory effects of an unrelated prime only begin to occur at later delays. This was just one finding that led Neely to conclude that priming is both fast and automatic (that is, not consciously controlled). These early studies used visual primes and visual targets. **Swinney, 1979** in *Journal of Verbal Learning and Verbal Behaviour* showed that priming could also be found if the prime was spoken and the target visual ('cross modal' priming).

Subsequent studies have attempted to find out what the relationship must be between prime and target for priming to be found: an important distinction here is between 'semantic' priming — between two words related in meaning but which do not necessarily co-occur ('hammer' and 'screwdriver'), and between 'associative' priming — between two words that are associated through co-occurrence but not necessarily through meaning ('couch' and 'potato' or 'fruit' and 'fly'). Many studies use prime-target pairs that are both semantically and associatively related ('doctor' and 'nurse'). In fact, **Fischler, 1977** in *Memory and Cognition* showed priming between semantically related, but not associated words, and **Seidenberg and colleagues, 1984** in *Memory and Cognition* showed priming between associated, but not semantically related words.

than less frequent words. The priming words used by Zwitserlood ('captain' and 'captive' being just one pair) did not have the same frequencies; either the actual word was more frequent than the alternative, or the alternative was more frequent than the actual word, or they were roughly the same. Zwitserlood found that if the alternative was more frequent than the actual word, then the target word related to that alternative would be responded to faster than the target word related to the actual word (but only early on, before the identity of the actual word became clear). And similarly for when the actual word was more frequent than the alternative — there was generally more priming to the target word related to the actual word in this case than to the alternative. It is as if the lexical entries corresponding to the more frequent words become more strongly activated on the basis of similar acoustic input than their less frequent neighbours. Again, this had been predicted by the theory.

Multiple activation and the contextual fit of the alternatives

A second important aspect of Zwitserlood's study is that the words 'captain' and 'captive' were embedded in two kinds of context: one which permitted either of these words at the end of the sentence ('the men stood around for a while and watched their captain/captive'), or one which permitted only the actual word that was used ('the men had spent many years serving under their captain'). If context has an effect on the activation of the alternatives, one would expect more priming to 'ship' after the biasing context ('the men had spent many years serving under their') than after the neutral context ('the men stood around for a while and watched their'). And sure enough, this is what Zwitserlood found.

Zwitserlood, 1989 in *Cognition*

The Zwitserlood study manipulated three factors of interest: the position within the spoken word (e.g. 'captain') at which the visual probe ('ship' or 'slave') was presented, the type of context that the spoken word was embedded in, and the frequency of the spoken word relative to the competing word ('captive') that shared its onset ('capt'). In the final probe position, when it was clear that the spoken word was 'captain', there was priming only to 'ship' (that is, lexical decisions were faster to 'ship' than when 'ship' was preceded by an unrelated sentence). In the earlier probe positions, there was priming to both 'ship' and 'slave'. In the very early probe positions (near the beginning of the word), there was a hint of a frequency effect, with more priming from whichever of 'captain' or 'captive' was most frequent. This effect was very weak, although it has been replicated in other studies (**Marslen-Wilson, 1987** in *Cognition*). An effect of context was found in the intermediary probe position (between the very early ones and the late one, but still at a point when it wasn't clear which of 'captain' or 'captive' was being heard) — there was more priming of 'ship' when the context was biased in favour of 'captain' than when it was compatible with either 'captain' or 'captive', and conversely, there was less priming of 'slave' when the context was biased in favour of 'captain'. The fact that this effect of context was found at the intermediary position and not at the early position suggests that context can affect the activation of the alternatives, but that it does not affect their initial activation which is based primarily on the acoustic input.

So the cross-modal priming studies show that we activate the entries for all possible words compatible with the acoustic input. But does this not mean that there is a real danger of system overload? How do we prevent an explosion of lexical possibilities? How do we choose which possibilities are the right ones? And how is it possible that we can activate the meanings of all these words without recognizing that we have done so? Apparently, activation does not imply recognition. But if it does not, what exactly is recognition? What does it mean to say that we recognize (or even hear) a word? And when does this recognition happen? Presumably, we first activate the possibilities, and recognition is what then happens when we (somehow) determine which of these possibilities is the right one. But how do we do that?

The effects of acoustic mismatch

According to Marslen-Wilson's theory, lexical access is a kind of race; different lexical entries compete in the race, but there can be only one winner of this **lexical competition**—we recognize a word when it has been identified as the winner. But for there to be a winner, there have to be losers. So what determines whether, and when, a competitor falls by the wayside?

The most obvious factor is compatibility with the acoustic input. There is extensive evidence showing that *acoustic mismatch* leads to a rapid decline in the activation of a lexical entry. Whereas a word like 'book' might prime 'page', the nonword 'boog' (pronounced to rhyme with 'book') would not — changing the voice onset time (see Chapter 3) of the final phoneme from a /k/ to a /g/ would be enough to cause rapid deactivation of the lexical entry for 'book'. But if the smallest deviation of this kind can lead to a rapid decline in the activation of a lexical hypothesis (and see Chapter 5 for further examples), what is going to happen each time we hear a word pronounced slightly differently, or each time a bit (or worse still, a lot) of background noise changes the acoustic signal? There has to be some *tolerance* in the system.

In fact, it turns out that there is; a slight deviation does not cause a lexical entry to self-destruct, it merely causes a decline in the activation, which means that the activation can pick up again if subsequent input is still compatible with that entry. Of course, if that deviation occurs at the start of a word, and is sufficiently large, it may prevent the intended word from being activated in the first place. Cynthia Connine, at the State University of New York, showed in the early 1990s that non-words which were very similar to real words could prime words related to those real words—importantly, the more similar the non-word to the real one, the greater the priming. So there *is* tolerance built into the system, with the activation of each lexical candidate reflecting its goodness of fit to the input. This tolerance is not unlike the tolerance seen in the categorical perception of

DO RHYMES PRIME EACH OTHER?

Marslen-Wilson and Zwitserlood, 1989 in *Journal of Experimental Psychology* found that rhymes (as in 'candle' – 'handle') do not prime one another. This is just as well, or we would confuse 'tongue' with 'dung'. Later research established, however, that there *is* some activation of the rhyme:

Allopenna and colleagues, 1998 in *Journal of Memory and Language* monitored people's eyes as they looked at a board on which were placed a beaker, a plastic beetle, and a small loudspeaker, amongst other objects. When told to 'pick up the beaker', people's eyes looked with roughly equal frequency at the beaker and the beetle during the earliest part of 'beaker'. Once the /k/ of beaker was heard, there were far fewer looks towards the beetle, because of the acoustic mismatch. However, despite the word-initial mismatch between 'beaker' and 'speaker', there *were* looks towards the loudspeaker (and more than towards other objects), due to the overlap.

Lexical competition the process by which the activation values of simultaneously activated lexical candidates (the actual word and its competitors) change to reflect their goodness of fit with the acoustic input and context, leaving just one candidate to emerge as the best-fitting one (hopefully, this is the actual word being heard!)

Connine, 1993 in ***Journal of Memory and Language***

The aim of this study was to explore the effects of acoustic mismatch on lexical activation: non-words were created which deviated from real-words by varying degrees. So from the real word 'captain' (which normally primes 'ship'), a similar non-word would be 'gaptain' (the initial phoneme has been changed from an unvoiced consonant to its voiced equivalent) and a dissimilar one would be 'maptain' (the initial phoneme is voiced instead of unvoiced; its *place of articulation* has changed–from the middle of the roof of the mouth to the lips; and its *manner of articulation* has changed to a nasal). In a cross-modal priming study, people heard either the real word (Experiment 1) or one of the two non-words (Experiment 2). At the offset of this stimulus, a target word was presented on a screen and the people had to decide whether it was a real word or not. Priming was assessed by comparing decision times to 'ship' with decision times to a target that was unrelated in meaning to 'captain'. There was substantial priming from 'captain', somewhat less priming from the similar non-word 'gaptain', and no priming (or at best, very little) from the dissimilar non-word 'maptain'. So the amount of priming depended on the degree of match between the prime word and the lexical representation of 'captain'.

phonemes discussed in Chapter 3—there, variation in the acoustic signal associated with a particular phoneme is tolerated up to a certain degree, beyond which any further variation causes the sound to be perceived quite differently. And just as goodness-of-fit is an important concept in theories of spoken word recognition, so we saw that it is important in categorical perception—there, the system is also sensitive to the slightest variation, even if the final perception (as one phoneme or another) is unchanged.

In general, then, a word can be recognized when there has been sufficient mismatch between the acoustic input and that word's competitors. Often this will be before the word's acoustic offset, but sometimes it may be after. 'Ram' could continue as 'ramp' or 'rampart'. But if the sequence being heard was something like 'The ram roamed around', the lexical entries for 'ramp' and 'rampart' would become deactivated when 'roamed' was encountered, resulting in the eventual recognition of 'ram'.

The Cohort Model A QUICK SUMMARY

The 'Cohort' model is the model of spoken word recognition developed by William Marslen-Wilson and colleagues. It assumes that the acoustic input causes the simultaneous activation of multiple candidates compatible with the input, with each candidate's activation reflecting its goodness of fit to the input. As more input accumulates, a candidate emerges as the best fitting one. A word is recognized when its level of activation is sufficiently higher than that of its competitors. There is some evidence (the Zwitserlood study) that after the initial activation of candidates, activation values can change to reflect not only goodness of fit to the acoustic input, but also goodness of fit to the context. The 'modern' form of the theory was first described by **Marslen-Wilson, 1987** in *Cognition*. In an early version of the cohort model (from **Marslen-Wilson and Welsh, 1978** in *Cognitive Psychology*), acoustic mismatch would return the activation of a candidate to zero (activation was not based on goodness of fit in that early version).

Early models of word recognition I: Morton, 1969 in *Psychological Review*

UK psychologist John Morton devised the *logogen* model of word recognition, in which word detectors (equivalent in some respects to motion detectors in vision) store a word's visual properties (for reading), phonological properties (for speaking), and semantic properties (aspects of its meaning). And just as movement detectors need a particular kind of shape to move in a particular direction before they become activated, so these word detectors (the *logogens*) require a particular acoustic input to become active. The basic idea, then, is that during spoken word recognition, each detector becomes progressively more active as it encounters progressively more acoustic input that matches its 'requirement'. When a threshold has been reached, the detector 'fires', and the word is recognised.

The model could explain a number of phenomena associated with the speed with which words can be recognized: For example, frequency effects (high frequency words are recognized faster) could be explained by supposing that each detector normally has some level of activity even when there is no matching input, but that the detectors of highly frequent words have increased levels of 'resting' activity relative to those of infrequent words. Such detectors would reach threshold sooner than a detector with a lower level of resting activity (which would therefore require more matching acoustic input to bring it to threshold, something that would therefore require more time). A second example concerns the role of context (words which fit in well with the context are recognised faster than words which do not): The idea here was that each logogen could receive not simply acoustic input, but also input from other logogens that were related in meaning, or from some other processing subsystem that could work out the meaning of the sentence (this part was a little vague in the original formulation of the theory!). A logogen that received matching acoustic input *and* additional input from any of these other sources would reach threshold sooner than a logogen which received only matching acoustic input.

The model was compromised by Marslen-Wilson's finding that the point at which a word is recognized is dependent on whether or not other words are still compatible with the input. In the logogen model, each detector's threshold is independent of the activity of other detectors, and so this dependence is not easily explained.

Early models of word recognition II: Forster, 1976 in *New Approaches to Language Mechanisms*

Not all models of word recognition abandoned the dictionary metaphor, and one influential proposal from the 1970s was formulated by Ken Forster, now at Arizona. He proposed that the mental lexicon is indeed much like a conventional dictionary, with word recognition consisting of a search through the dictionary, entry by entry. As each word is heard, its phonological form (its sound structure, in effect) would be determined from the acoustic input and matched against a list of entries that were ordered by frequency. Of course, and as we saw earlier, the access code is crucial when it comes to accessing the entries in a dictionary — it would be no good coding each entry in terms of the acoustic properties of the word it represented if the dictionary was to serve visual word recognition (as during reading). So there were two access lists that could be searched during word recognition: one was coded phonologically, another was coded orthographically (that is, by letter, for use during reading). Once the perceptual form of the word had been matched against an entry in either one of these lists, that entry yielded an 'address' in a master list at which all the relevant information about that word (including its meaning) could be found. This is not unlike how Roget's Thesaurus operates: at the back of the book is an alphabetically ordered list of words – one scans down the list until a match is found, and then one reads off the page number on which the words with similar meanings can be found, and then one turns directly to that page. Because each access list was ordered by frequency, matches against more frequent words would be found faster than matches against less frequent words (hence the finding that more frequent words are recognised faster than less frequent words. And because the access lists were searched entry-by-entry, it would take a relatively long time to recognize a non-word (e.g. 'slomp') because all the real word entries would be searched first (and indeed, it takes longer to recognise non-words than real words).

This model, like the logogen model, was compromised by the observed relationship between recognition and uniqueness point. According to this model, once a match is found in the frequency-ordered list, the search terminates (hence frequency effects), but this means that a lower frequency word that overlaps with the target word should not influence the search time, because the search should terminate before that lower frequency word is reached.

Acoustic mismatch in running speech

Words are rarely spoken in isolation, but are spoken in the (seamless) context of other words coming before and after. And this is important for a number of reasons, not least because people are generally quite lazy in their articulation, and the position and shape of the articulators at any one moment reflects not simply the sound to be produced at that moment, but also the sound that will be produced next. We encountered a version of this phenomenon in Chapter 5 under the guise of co-articulation. Generally, the term is used to describe how a vowel, for instance, can be 'coloured' by the consonants that precede and follow it. The fact that vowels are not perceived categorically allows this colouring to be used in anticipating the identity of the following segment. But something very similar can occur when one consonant is followed by another. And this is where the problems start: if the consonant were to actually change as a result of this process, a mismatch would occur. And this would mean that we would then fail to activate the intended meaning. Just how bad is the problem?

The answer is that it is as bad as having to recognize 'Hameethathimboo' as meaning 'Hand me that thin book'. Word-final consonants such as the /d/ in 'hand', the /t/ in 'that' and the /k/ in 'book' are often dropped completely. And instead of articulating the /n/ in 'thin' by closing off the mouth with the tip of the tongue against the back of the upper teeth (and allowing air through the nasal passage), the speaker might anticipate the following /b/ and instead close off the mouth at the lips (still allowing air through the nasal passage). This would result in 'thin book' being articulated as 'thim book'—an example of **phonological assimilation**. And because the /d/ had been dropped from 'hand me', the preceding /n/ may combine with the /m/ to produce 'hamee'. These kinds of changes, generally at the ends of words, are surprisingly common, although the extent to which they occur, and how they occur, can depend on the person doing the speaking and the language being spoken — again, we encountered such differences in the preceding chapter, when we considered some of the differences between English and French. But if acoustic mismatch leads to the deactivation of lexical candidates, what hope is there of recognizing the intended words after these changes have occurred? If these kinds of effects are more common than not, how could we ever recognize a sentence in its entirety?

Place assimilation this is a form of *phonological variation* that generally occurs when the final consonant of one word is changed to reflect properties of the first consonant of the next word. Broadly speaking, the word-final consonant takes on the *place of articulation* of the following consonant (where in the mouth the airflow is restricted during the articulation of the consonant — chapter 3).

Coping with mismatch

The answer, once again, is tolerance. In this case, the tolerance is context-*sensitive*. The nonword 'thim' will activate the meaning associated with 'thin', but only in the context of a following word. But it cannot be just any old word, it has to be a word in the context of which it would have made sense for what was originally an /n/ to become an /m/. Whereas the 'thim' in 'thim book' would activate the lexical entry for 'thin', the 'thim' in 'thim slice' would not. This was

Gaskell & Marslen-Wilson, 1996 in *Journal of Experimental Psychology*

This study made use of a technique called 'repetition priming', where the lexical decision time to a visually presented word is speeded up by the prior presentation of that same word presented acoustically. They found that 'thin' was primed by 'he liked the thin book' (compared against 'he liked the brown book') *and* by 'he liked the thim book' (compared against 'he liked the browm book'). There was no difference, in fact, in reaction time to 'thin' in either of these two conditions. On the other hand, 'thin' was not as effectively primed, if at all, by 'he liked the thim slice', where the word-final change from 'thin' to 'thim' is not permitted given the following /s/ (but 'thin' *was* primed by 'he liked the thin slice').

demonstrated by another student of William Marslen-Wilson's, **Gareth Gaskell**, in a series of experiments using variations on the priming theme. This naturally begs the question of how the system 'knows' to do this.

Linguists have produced a whole range of rules which describe the range of circumstances in which these different kinds of word-final changes can occur. The *Cambridge encyclopedia of language* writes one such rule as: 'an **alveolar nasal** becomes **bilabial** before a following bilabial consonant'. Despite the complexity of the terminology used in describing such rules, it is certainly tempting to believe that the human mind runs these rules in reverse in order to recover what was originally meant. This would mean that, somewhere, the mind has a store of these rules, and that each time it encounters a pair of consonants (two bilabials, to use that last example), it looks at the rules in order to decide whether the first of those consonants should perhaps be interpreted as something else (an alveolar consonant). But is this what we really do? And if so, where do those rules come from? And if we do not do it this way, what other way is there?

One way to imagine what we might do instead is to recall that the task of the infant is to associate sounds with meaning. The infant must therefore associate not just /thin/ with the meaning of 'thin', but also /thim/ with 'thin', and even /thing/ with 'thin' (as in 'The thin carpet was worn through', where 'thin' would be pronounced as /thing/). But what is actually being associated with the meaning of 'thin' is not just the sound that has been heard, but rather the sound that has been heard within a particular context. This context necessarily includes the surrounding sounds. The infant might therefore associate with the meaning of 'thin' all the following: /thin/ in combination with a following /t/ (e.g. 'thin tree'), /thim/ in combination with a following /b/ ('thin book'), or /thing/ in combination with a following /k/ ('thin carpet', where /k/ is the first phoneme of 'carpet'). As an adult, it is then just a matter of recovering whatever meaning was associated with a particular combination of sounds.

There is, of course, one slight problem, although it is hardly slight—it would require the infant/adult to somehow store all the different pronunciations of each word together with knowledge of which context leads to which pronunciation. This would be extremely wasteful because the system would need to know that /thim/ in combination with a following /b/ is 'thin', and that /skim/ in com-

CONSONANTS AND THEIR PHONETIC LABELS

Different consonants are given different names by *phoneticians* (linguists concerned with the formation of sounds) depending on (1) their *manner of articulation* (nasal – air flows through the nose: /n/, stop – air stops: /t/, fricative – air flow constricted: /s/) and (2) their *place of articulation*:

Bilabial – at the lips: e.g. /b/
Labio-dental bottom lip against top teeth: e.g. /f/
Dental – tongue against top teeth: e.g. /th/
Alveolar – tip of tongue against base of top teeth: e.g. /t/
Palato-alveolar – tongue (back from the tip) towards the front of the roof of the mouth: e.g. /sh/
Palatal – it is easier simply to give the one example in English! – /y/
Velar – the body of the tongue (towards the rear) against the roof of the mouth: e.g. /k/

Place assimilation is most common between alveolar consonants at the end of one word, and bilabial or velar consonants at the start of the next word.

bination with a following /b/ might be 'skin'. It would surely be much easier to simply acquire a very much smaller number of rules to do the same job, each rule applying across a whole range of words (for instance, one rule could apply to all words normally ending in /n/ when followed by a /b/, /p/, or /m/). Of course, in order to learn the rule, the infant would have to be exposed to enough different pronunciations in different contexts to formulate an accurate rule (and how does it do *that*?), but it would not need to remember each such combination. Moreover, it might never encounter /skim/ meaning 'skin' as it formulated the rule that would convert /m/ to /n/, but it would still be able to use this rule to figure out what was intended if it *did* subsequently encounter a 'skim book' (something of considerable use if you are a dermatologist).

So rules have a lot going for them. In which case, on what basis can anyone reasonably claim that rules are not run in reverse? To answer this question, we need to revisit, only very briefly, what we know so far about how we recognize spoken words.

Activation, prediction, and context

We know from the work of Marslen-Wilson and Zwitserlood that at each point during a spoken word (that is, at each moment as the word 'unfolds' in time), all the lexical entries that are compatible with the input up until that point are activated. So during the /sp/ of 'speech', the entries corresponding to 'spike', 'spoke', 'speak', 'spook', and countless others are activated; after /spi/ (corresponding to 'spea..'), the entries for 'spike', 'spoke', and 'spook' become relatively less active compared to those for 'speak' and 'speech' (because of the acoustic mismatch given the actual vowel encountered in the input); and eventually, only 'speech' remains significantly active. But the activation of each entry corresponds, in a sense, to a prediction of what phonemes might be encountered next in the input, and of what the corresponding meaning would be. For instance, the activation of the entry for 'speak', after the /sp/ of 'speech', constitutes a prediction that the next phoneme will be /i/ (corresponding to the 'ea' in 'speak') — which is a correct prediction — and that the phoneme after this one will be a /k/ — an incorrect prediction given that the unfolding word is 'speech'. So one way of thinking about this is that at each moment in time, the mind attempts to anticipate, or *predict*, what may come next — it activates all the entries that are compatible with each possible prediction that it can make (and in fact activates each according to how likely it is — hence those earlier frequency and context effects, where more frequent or better fitting lexical candidates are activated to a greater extent than less frequent or poorer fitting candidates). Why is this important? Because after the /thi/ of 'thin', the entry for 'thin' is activated on the basis of the prediction that the phoneme /n/ will follow. But the entry for 'thin' is also activated on the basis of another prediction: that the phoneme /m/ *and then* the phoneme /b/ will come next. This is hardly different from the case where, after the /sp/ of 'speech', the entries for 'speak' and 'speed' are activated — the activation of 'speak' predicts that the phoneme /i/ and then the phoneme /k/ will come next, whilst the activation of 'speed' predicts the phoneme /i/ and then the phoneme /d/. And that is why this is all so important, because it means that no additional mechanism is required, other than the basic one for recognizing words, to allow the correct interpretation of /thim/ as 'thin' in the context of a following /b/. No explicit rules of the 'alveolar nasal becomes bilabial' type are required, let alone a mechanism for evaluating which of several different rules might apply at any one time.

But *deploying* knowledge (that is, applying it in the appropriate circumstances), and *acquiring* it, are two very different things. A mechanism is required that would allow those different predictions to develop as the child learns new words—as it learns the ways in which words systematically change as a function of the contexts in which they occur. And what really matters is whether, never having encountered the sequence corresponding to 'skim book', we would still interpret it as 'skin book'. Rules are great for making generalisations; but what about our (currently) hypothetical system that learns to predict what will come next? Can it make equivalent generalisations? Well… it *could* make equivalent generalisations if it encountered not just 'thim book' but also 'opem book' (for 'open book'), because it could then learn that what predicts the interpretation of /m/ as /n/ is not '/thim/ in the context of a following /b/', but '/n/ in the context of a following /b/'. In effect, it could learn *the rule*. And in fact, Gareth Gaskell demonstrated just this in a computer simulation that we shall find out more about in Chapter 13: if it learned that 'thim book' and 'opem book' meant 'thin book' and 'open book' respectively, it *did* generalise to cases such as 'skim book' (as 'skin book') which it had never encountered before. In that chapter we shall explore the manner in which computer simulations of neural functioning can give rise to a particular kind of learning that results in exactly the properties of lexical access (including its predictive nature) that we have encountered so far. For the moment, though, it does not really matter how the system solves the 'thim book' problem—what matters is that we know that it *can* solve the problem, and that we know something about what the solutions must look like.

Before leaving the problem altogether, it is worth considering that phonological variation is very widespread indeed: each time you hear someone speaking, they are most likely speaking with an accent that it is distinct from someone else's. And each distinct accent is another example of individual vowels, and occasionally consonants, being uttered in some slightly different, but systematic, way. Yet somehow we adapt, which means that we manage to map those vowels and consonants onto our own internal representations of the words which they form. Most likely, the mental machinery that is able to work out this mapping is very similar, if not identical, to the machinery we have encountered here for dealing with phonological assimilation, and to which we shall return in Chapter 13, on computer simulations of mind and language.

In the meantime, there are still other mysteries waiting to be unlocked. We have come some considerable way in understanding how words are recognized as the speech we hear unfolds before our ears, but we have so far omitted mention of one very important and defining characteristic of many of the words we hear — that they are *morphologically complex*, consisting of stems and affixes.

Recognizing the internal complexity of words

Morphologically complex words present a challenge to models of word recognition: if each word has its own lexical representation, how are we to explain the systematic relationship in meaning between 'he is walking' and 'he is talking'? In each of these, the '-ing' indicates that the actions are ongoing, in much the same way that, in 'he walked' and 'he talked', the '-ed' indicates that the actions happened in the past. Similarly, how are we to explain the systematic relationship between 'unreal' and 'real', or between 'undo' and 'do'? Or between the pairs 'enhance'–'enhancement' and 'entrap'–'entrapment'? Is there some rule, somehow encoded within the neural structures of the brain, which reflects, for example, that '-ing' on the end of a verb means that the action described by the verb is ongoing? And if there is, why would we not mistake 'bring' for an ongoing version of 'br'?

One useful starting point when addressing such issues is to suppose that it is *not* the case that each word has its own lexical representation. Instead, we might suppose that each *morpheme* has its own representation. That way, the relationship between 'walking' and 'talking', and between 'entrapment' and 'enhancement' can be explained in terms of the words sharing morphemes that each contribute to the meaning in a way that is specific to that morpheme. And that way, there could be no mistaking 'bring' as an ongoing instance of the verb 'br' because there simply is no morpheme in the lexicon corresponding to the stem 'br'. As usual, though, things are never so simple. But sometimes the complexities reveal the workings of the mental machinery. Words like 'apartment' and 'department' are not morphologically complex in the same way that words like 'entrapment' and 'enhancement' are — 'entrapment' is related in meaning to 'entrap', but 'apartment' is not related to 'apart'. Neither is 'department' related to 'depart'. So how does the mind deal with such cases?

Recognising affixes

Not surprisingly, the mind deals with these cases in much the same way as it deals with other aspects of word recognition. On hearing the initial sounds of 'enhance', it activates the entry for 'enhance' (that is, it activates the entry corresponding to the stem 'enhance', and this includes the activation of some representation of the meaning associated with 'enhance'). Towards the end of the sequence of sounds corresponding to the stem, it will anticipate a number of possibilities (recall that such anticipations, or predictions, are what it is really doing all along when it activates a particular lexical entry on the basis of only part of the corresponding acoustic input). These possibilities will include, amongst other things, the subsequent morpheme '-ment' (and the activation of some representation of the meaning associated with this morpheme; in this case, 'the act or consequence of whatever is the action associated with the preceding stem'). If the subsequent acoustic input matches this morpheme, then its activation increases. If not (as in the case of 'enhanc*ing*' or 'enhance *his income*'), then the activation decreases due to the mismatch. For a word like 'apartment', things are a little different: After hearing the sequence corresponding to 'apart', two distinct lexical entries are activated — one corresponding to the stem 'apart' and both its associated meaning (to do with being separate) and the anticipation that a new word will come next (which will subsequently be found not to be the case), and another corresponding to the still unfolding, but unrelated word 'apartment' — this is not so different from the case of /capt/, where two distinct entries are activated, one relating to 'captain' and the other to 'captive'. What is important about this last case is that the 'ment' in 'apartment' is not interpreted as a distinct morpheme. And of course, this explains also why the 'ing' in 'bring' is not confused with the morpheme '-ing'.

So much for suffixes (morphemes at the ends of their stems). What about prefixes (morphemes at the start of the word)? Words like 'fasten' (derived historically from the stem 'fast' as in 'to make fast') can be prefixed with 'un-' or 're-' to make 'unfasten' or 'refasten'. As the sequence /unfa../ is heard, the entry for the morpheme 'un-' will become activated, followed by the entry corresponding to the stem 'fast'. Similarly for 'refasten', where the morpheme 're-' is activated followed by the stem 'fast'.

The proof of the pudding

This all sounds quite simple, really. The proof is more complex, and relies on a series of priming studies carried out by William Marslen-Wilson and colleagues during the 1990s, who have explored morphological processing during spoken word recognition in a variety of different languages, including English, Hebrew, and Chinese. They used the priming technique again — people heard a word like 'happiness' but had to decide whether a word presented visually at the offset of 'happiness' was a real word or not (the lexical decision task again). They found that 'happiness' primes 'happy', and 'rebuild' primes 'build' (so faster decision times to 'happy' and 'build' respectively), suggesting that the activation of the stem in the prime ('happiness' or 'rebuild') facilitates, or primes, the subsequent activation of that same stem in the subsequently presented visual target ('happy' or 'build'). The same logic applied to other word pairs used in these experiments shows that words like 'apartment' and 'apart' do not share the same stem (we knew that already, of course! But it is important that the empirical data reflect that). The data are a little more complex than this, because one might argue that priming between 'happiness' and 'happy' does not reflect a shared stem but rather it reflects the common meaning, or even the overlap in the initial phonemes. Using this same priming method, it was found, however, that a word like 'tinsel' would not prime 'tin' (showing that overlap in phonemes is not enough to guarantee priming), and that a word like 'happiness' does not prime 'happily' (showing that overlap in meaning is not enough to guarantee priming between morphologically complex words). But this last finding is very surprising, because 'happiness' and 'happily' share the same stem, and so they *should* prime one another. So what is going on here?

This last question can be addressed by one particularly interesting result: it turns out that 'happiness' can prime 'darkness'! Why? Because they share the same morpheme '-ness'. Its activation in 'happiness' facilitates its subsequent activation in 'darkness'. So whereas a stem with one suffix ('happi<u>ness</u>') will prime a different stem with the same suffix ('dark<u>ness</u>'), a stem with one suffix ('happi<u>ness</u>') will *not* prime the same stem with a different suffix ('happi<u>ly</u>'). One way of thinking about this is that the recognition of the stem in 'happily' *is* facilitated by its prior activation in 'happiness', but that the recognition of the suffix '-ly' is made harder (it is *inhibited*) by the prior activation of the different suffix '-ness' when in combination with the same stem. And why should this happen? Although the answer to this question is somewhat unclear (meaning that there is no hard-and-fast answer as yet), one possibility is that if you hear 'happiness' one time, and immediately hear the stem /hapi/ again, you assume that you're going to hear the same thing as before rather than something different. So you predict with even greater certainty than usual that '-ness' will follow, making it harder than usual to then recognise the '-ly' when it occurs. This is like a local frequency effect — at the offset of /hapi/, the mind has to choose which of many predictions are possible, and which of these are most likely. Generally, higher frequency continuations are more likely than lower frequency continuations, and the morpheme '-ness' is a very frequent morpheme when you have just heard it! Hence the tendency to expect it again, at the expense of expecting anything else.

So much for morphological complexity in the mental lexicon. There is an ever-growing literature on morphological processing during both spoken and written word recognition, and on how different kinds of morphological complexity change both the nature of the representations contained within the mental lexicon and the nature of the recognition processes that use those representations (for example, words in Semitic languages are processed quite differently from words in English, or other languages that lack **Semitic morphology** and its infix system). An entire book, or two, could be devoted to this topic alone!

MORPHOLOGY IN SEMITIC LANGUAGES

In the Semitic languages, which include Hebrew and Arabic, all verbs, and many nouns, are composed of two morphemes: a root, which conveys the core meaning of the word, and what is called a 'word pattern', which conveys the equivalent of the affixes in English. In the following Arabic word, the word pattern is underlined, and the root is not — k<u>a</u>t<u>a</u>b ('wrote'), k<u>u</u>t<u>i</u>b ('was written'), <u>a</u>kt<u>u</u>b ('was writing'), <u>u</u>kt<u>a</u>b ('was being written'). Here, the root is composed of the consonant sequence 'k–t–b', and the different vowel patterns constitute different infixes.

Experiments on written Hebrew by **Frost and colleagues, 1997** in *Journal of Experimental Psychology* have shown that these two types of morpheme are represented independently within the mental lexicon — words with the same root but different word pattern prime each other, and words with the same word pattern but different root also prime each other (although interestingly, this last effect occurs only with verbs, not nouns). To rule out the possibility that priming was due simply to letter overlap, there were conditions where the prime was either 'ktb' (the probe was 'katab'), or 'kat' — in the first case there is both letter overlap and morphological (root) overlap, and in the second, there is just letter overlap. The data demonstrated that the priming effects caused by roots and by word patterns were independent of letter overlap.

Marslen-Wilson and colleagues, 1994 in *Psychological Review*

A large number of findings, across six different experiments, were reported in this study, using cross-modal repetition priming (described a little earlier). The findings can be separated into effects of suffixes and effects of prefixes. Here are the highlights: (1) Suffixes: both 'governor' and 'government' prime 'govern'; this is not due to phonological (sound) overlap, because 'tinsel' does not prime 'tin', nor does 'apartment' prime 'apart' (so priming only occurs between morphological variants of the same stem — 'apartment' is not morphologically related to 'apart'). Finally, 'governor' and 'government' do not prime each other. (2) Prefixes: 'unfasten', 'refasten', and 'fasten' all prime each other; this is not due to phonological overlap, because 'trombone' did not prime 'bone' (we return to the relationship between 'trombone' and 'bone' a little later). Finally, 'depress' and 'express' do not prime each other, because they are not morphologically related.

Early studies of morphological processing

Some of the earliest research into how the mind processes morphologically complex words took as its starting point the idea that words which look (or sound) morphologically complex are broken down into their component morphemes, which are then checked against the mental lexicon. One of the earliest proposals, by **Taft and Forster, 1975** in *Journal of Verbal Learning and Verbal Behaviour*, led to the idea of 'affix-stripping'. For a word like 'revive', the prefix 're-' would be stripped off leaving the morpheme 'vive' which would then be looked up in the lexicon (it is a 'bound' stem morpheme — it cannot occur by itself, but can occur in other words like 'survive'). In the case of a 'pseudoprefixed' word (the 're' in 'relish' is not a prefix, and nor is 'lish' a stem), the pseudoprefix would be stripped off (in error) but the resulting pseudostem would not match against the lexicon. So one consequence of an automatic affix-stripper would be that 'relish' should be harder to recognize than 'revive', because the 're' would be stripped off when it should not be. In support of this idea, they found in a visual lexical decision task that (1) 'vive' takes longer to reject than 'lish'—because 'vive' exists as a stem whereas 'lish' does not; and (2) 'invive' (an inappropriately prefixed stem) takes longer to reject than 'inlish' (a prefixed non-word)—again, because once the prefixes are stripped off, 'vive' is a real stem and 'lish' is not. The first morphological priming study was reported by **Stanners and colleagues, 1979** in *Journal of Verbal Learning and Verbal Behaviour*, who used repetition priming to show that 'happy' can be primed by both 'unhappy' and 'happiness' (all the words were presented visually).

Although there has been a large literature on morphological processing and word recognition, the vast majority of this literature has been based on visual word recognition, with only a handful of studies, such as the Marslen-Wilson studies, exploring spoken word recognition.

Getting at the meaning

We know something about how, and when, lexical entries are activated, and how, and when, they may become deactivated. But what information is contained within a lexical entry? How do we square a question like this with the idea that a lexical entry is simply a kind of neural circuit? Returning to the analogy of a combination lock, we can ask the same kind of question: given the arrangement of its tumblers, what information does a combination lock contain? On the one hand, there is a sense in which a combination lock itself contains no information at all. It is simply a physical arrangement of potentially moveable objects. On the other hand, the precise arrangement of the tumblers determines which exact sequence will open the lock — the appropriate sequence has meaning by virtue of causing an effect to occur that is specific to that sequence, and to no other. In this sense, the combination lock does contain information, and a skilled locksmith would be able to examine the arrangement of the tumblers, and figure out, on the basis of this information, the sequence required to open the lock. Similarly, even if a lexical entry is nothing more than the neural equivalent of a combination lock, it contains information by virtue of the effect that an input sequence can have (and in Chapter 9 we shall discuss further the nature of meaning, and the nature of the effects that a word may cause). And just as we can still refer to lexical entries when what we are really talking about is some complex neural circuitry, so we can refer to meaning when what we are really talking about is the result of this circuitry becoming activated.

So lexical entries are where the meaning of a word resides. But one of the first things one notices when opening up a large written dictionary is that most words have more than one meaning (they are **homonyms**). The word ‘pitch’, for example, was introduced in Chapter 1 without any explicit definition. And yet it would be almost impossible to look up the word and not discover that it has several distinct senses or meanings: to pitch a ball; to pitch a tent; the pitch of a roof; the pitch of a musical sound; the pitch you get from distilling tar; the sales pitch; the football pitch, and so on. Presumably the mental lexicon must also reflect this multiplicity of meaning. But what are the implications for how we retrieve a single meaning? Do we activate all possible meanings of a word that is ambiguous and has more than one meaning? Do we somehow scan all the meanings (to return, momentarily, to the dictionary metaphor) until we get to the first one that is appropriate given the context in which the word occurs, ignoring any others that we have yet to get reach? Or do we somehow activate only the contextually appropriate meaning, so avoiding a cluttering of our minds with all those other, inappropriate, meanings?

Homonyms Words that have the same form (spelling or sound) but which have more than one meaning. There are several distinct subtypes:

Homophones Words that are pronounced the same but which may have different spellings, such as ‘write’ and ‘right’.

Homographs Words that have the same spelling but which may be pronounced differently, such as ‘lead’ (the metal) and ‘lead’ (the verb, meaning to direct). Sometimes called heterophones (from the Greek, meaning ‘different sound’).

The word ‘bank’ is both a homophone and a homograph, although there is no special term to describe such cases.

Synonyms Different words that mean (roughly) the same: ‘speak’ and ‘utter’ (the verb), ‘complete’ and ‘utter’ (as in ‘utter nonsense’).

Polysemy The case of words with multiple senses (meanings that are related): ‘talk’ meaning ‘to talk’ or ‘a talk’; ‘speech’ meaning ‘spoken language’ or ‘a talk’.

Recognizing words with more than one meaning

In the late 1970s, **David Swinney**, then at Tufts University, published a paper that was to prove extremely influential. Not only did it demonstrate (after many years of bitter argument and counter-argument) that the alternative meanings of ambiguous words are activated, but it was also one of the first demonstrations of cross-modal priming, which we encountered earlier. The specific question that Swinney considered was whether or not we activate the alternative meanings of words even when those words are heard in the context of sentences which are compatible with only one of the meanings of the word. For instance, in 'He swam across to the far side of the river and scrambled up the bank before running off', it would hardly be appropriate to interpret 'bank' as a financial institution. Similarly in 'He walked across to the far side of the street and held up the bank before running off', it would hardly be appropriate to interpret 'bank' as a river bank (or to interpret 'hold up' as 'support'). In order to explore what actually happens when we hear sentences such as these, Swinney played people sentences similar in principle to these ones and immediately after they heard the word 'bank', he flashed up on a screen either 'money' or 'river'. The people knew to make a lexical decision as soon as they saw a word appear on the screen. Swinney found that, irrespective of which of the two sentences had been used, both 'money' and 'river' were primed. This showed that both meanings of 'bank' must have been activated.

Swinney, 1979 in ***Journal of Verbal Learning and Verbal Behaviour***

Two passages were created for a cross-modal lexical decision study: 'The government building had been plagued with problems. The man was not surprised when he found several bugs in the corner of his room' and 'The government building had been plagued with problems. The man was not surprised when he found several spiders, roaches, and bugs in the corner of his room'. The passages were played over headphones, and immediately at the offset of 'bugs', one of the following words appeared on a screen: 'ant' (related to the preferred meaning of 'bugs'), 'spy' (related to the less preferred meaning), and 'sew' (unrelated to either meaning, but with the same length and frequency as the other two words). All three words yielded identical lexical decision times when presented in isolation.

As expected, 'ant' and 'spy' were both primed (relative to 'sew') in the 'several bugs' case. They were also both primed after 'several spiders, roaches, and bugs'—so even when the context made it clear that the insect-related meaning was intended, the less-preferred spy-related meaning was nonetheless activated. When the word 'bugs' was replaced, in two other passages, by the word 'insects', there was only priming of 'ant', not 'spy' (so 'spy' was not activated just because of the previously mentioned government buildings). When the visual probes ('ant', 'spy', 'sew') appeared three syllables later (part way through 'corner'), only 'ant' was primed—the spy-related meaning of 'bugs' was no longer sufficiently active to cause priming of 'spy'. In the 'several bugs' case, the (preferred) insect-related meaning remained active (and the less preferred spy-related meaning did not) even when the context did not make it clear which was the intended meaning of 'bugs'.

Of course, at some stage, the inappropriate meaning of 'bank' must make way for the appropriate meaning, and, sure enough, Swinney found that if he presented the target words two or three syllables later (that is, *downstream* from the ambiguous word), only the target related to the contextually appropriate sense of the word was primed.

These findings aroused an enormous amount of interest, not least because some subsequent studies failed to show quite the same results. These later studies manipulated two factors that had not been systematically manipulated in the original Swinney studies (though this should not detract from the importance of those initial studies): the degree of contextual support for one meaning rather than another, and the preference for one meaning over another. The basic finding was that when the context supported the *less* preferred (or less frequent) meaning, then both the more frequent and the less frequent meanings would show signs of activation; but when the context supported the more frequent meaning, the less frequent meaning would show little or no sign of activation. One of the earliest demonstrations of this was by **Susan Duffy** and colleagues at Amherst, and although their experiment required people to *read* words like 'bank' rather than *listen* to them, the basic principle holds for spoken language also. So context is important. But as pointed out in a series of elegant studies by **Patrizia Tabossi**, now at Trieste University in Italy, the context has to be of the 'right' kind to prevent the activation of the less frequent meaning: it has to support a particularly salient feature of the more frequent meaning. For example, if the most frequent meaning of 'bank' is the financial institution, a context such as 'he walked across to the far side of the street and held up the …' would not be sufficient to cause activation of just the financial interpretation of 'bank'. This is because there is nothing in the context that supports that specific interpretation—there is nothing in the context that is specifically to do with what

Duffy and colleagues, 1988 in ***Journal of Memory and Language***

This study uses the fact, previously reported by Rayner and Duffy in 1986 that it takes longer to read an ambiguous word than an unambiguous word (the different meanings 'compete' with one another and slow down processing). In this study, people had to read sentences containing ambiguous words which were 'balanced' (both meanings equally frequent) or 'unbalanced' (one meaning very much more frequent). 'Pupil' is balanced (a student or a part of the eye), but 'port' is unbalanced (the ship version is much more frequent than the drink). In contexts which did not bias towards one meaning or another, 'pupil' took longer to read than the unambiguous word 'nurse'—both meanings were active. But 'port' took no longer to read than 'soup'—in the absence of any context, it is effectively unambiguous. In contexts which *did* bias one meaning of the balanced words ('After waiting in the rain for a bus, the pupil..') 'pupil' took no longer to read than 'nurse'—in effect, the word was now unambiguous. In contexts that biased towards the less frequent meaning of unbalanced words ('Even though it had a strange flavour, the port…'), 'port' took longer to read than 'soup'—the context activated the less frequent meaning to the extent that it could now compete with the other meaning.

Tabossi, 1988

Tabossi, 1988 in ***Journal of Memory and Language*** There were two sentence types: 'I arrived for an appointment at the bank' (Experiment 1) and 'I opened a checking account at the bank' (Experiment 3). At the offset of each sentence, a word related to the dominant meaning of 'bank' ('money') or to the other (*subordinate*) meaning ('river') or to neither ('honey') appeared on a screen for a lexical decision. In isolation, decision times to all three words were the same. After the 'appointment' sentence, reaction times to 'money' and 'river' were the same, and both faster than to 'honey'—suggesting activation of both meanings ('bank' is not so unbalanced that only one meaning is apparent). But after the 'checking account' sentence, 'money' was reacted to faster than 'river', which had the same reaction time as 'honey'—suggesting activation of just the dominant meaning.

Tabossi, 1988 in ***Journal of Experimental Psychology*** This study explored the priming of features related to certain aspects of the meanings of *unambiguous* words. In Italian, in which language the experiment was conducted, the word for 'rose' is unambiguously a flower. She found priming of 'thorns' at the offset of 'The girl was pricked by a rose' but not at the offset of 'The girl smelled a rose'—something about pricking primed that part of the meaning of 'rose' to do with thorns. Importantly, she did not find priming of 'thorns' after 'The girl was pricked by a wasp'—so it was an interaction between the pricking and 'rose' that caused this effect.

banks 'mean' (their function, their properties, and so on). On the other hand, in the context 'he was worried about the contents of his safety-deposit box so he visited his bank', there *is* something now which supports the specific interpretation of bank as a depository for valuables and money, and in this case, one would see activation of the meaning of bank-(finance) with no activation of bank-(river). What is less clear with results such as these is whether there really is no activation at all of the less frequent meaning, or whether activation is so low that the combination of the context and the more frequent meaning make it exceedingly hard to observe that activation.

At around the same time that Swinney performed his cross-modal priming experiments, **Michael Tanenhaus** and colleagues, then at Wayne State University in Detroit, performed a similar experiment, using words that were ambiguous between a noun (e.g. 'watch' as a time-piece) and a verb (e.g. 'watch' as a kind of looking). In a sentence like 'John began to watch the game', only a verb can follow the fragment 'John began to . . .'. Armed with this information, we could scan a written dictionary and look only at the entry for 'watch'-as-verb, ignoring the entry for 'watch'-as-noun, and hence ignoring the time-piece meaning of 'watch'. But does the same thing happen when we search the mental lexicon? Can we eliminate from the lexical search all the words whose syntactic categories are inappropriate given the preceding words in the sentence? Apparently not. Tanenhaus found that *both* meanings of 'watch', related to time-piece and looking, were activated when people listened to sequences such as 'John began to'. So knowledge of the type of word that must follow (that is, knowledge of its syntactic category) is not used to help constrain the possibilities. But why

Tanenhaus & colleagues, 1979 in ***Journal of Verbal Learning and Verbal Behaviour***

Like the Swinney study, this was one of the first cross-modal priming studies to be reported. In this case, people heard either ‘She held the rose’ or ‘They all rose’ and at the offset of these fragments one of the words ‘flower’ (related to one meaning), ‘stand’ (related to another), or ‘drink’ (unrelated to either meaning) appeared on a screen. People had to say aloud this visual probe word as quickly as possible (a ‘naming latency’ task). Naming times were faster to both ‘flower’ and ‘stand’ relative to ‘drink’, and it did not matter which sentence had preceded the probe—so even the grammatically inappropriate meaning of ‘rose’ was activated. When the visual probe words appeared 200 milliseconds after the offset of ‘rose’, priming was found only to the contextually appropriate meaning.

not? Is this some arbitrary property of the workings of the mental lexicon? Or is there some reason behind this?

Earlier, the processes involved in recognising spoken words were likened to a series of predictions—at each moment in time, as a word unfolds, the mind anticipates the range of possibilities that could unfold at the next moment in time, and the next after that. The Tanenhaus result is therefore all the more surprising; surely, we should predict, after ‘John began to’ that a verb will come next, and after ‘John checked his’ that a noun should come next. But whilst these predictions can indeed be made (and we return to such predictions in Chapter 8, on how we process grammatical information), they are not, in fact, much use. You might guess that a noun is coming up next. But *which* noun? Or, if it is likely to be a verb, *which* verb? The particular word that is most likely after ‘John began to’ is no more likely to be ‘watch’ than ‘snore’. So we have no firm basis, before encountering ‘watch’, on which to selectively activate (or ‘anticipate’, or ‘prime’) a meaning corresponding to ‘look’. This is why, when we do encounter it, there is no selective activation of one meaning (‘look’) versus another (‘clock’). Generally, the most reliable cue to the intended meaning of a word at any moment in time is in fact the acoustic signal itself, not the context. Which is why, when we hear ‘watch’, the alternative meanings are activated on the basis of that acoustic signal, although that initial activation rapidly changes to reflect the goodness of the fit between each alternative and the context.

Once again, we see the perils of adopting the convenient abbreviations that our written dictionaries provide us with; just because written dictionaries include the syntactic categories corresponding to the alternative meanings of each word does not mean that the mental lexicon does likewise, or that the mental lexicon uses that information if it has it available. After all, the *OED* also includes the approximate date at which each word entered the language — but just because the *OED* includes this information, and can search for words on the basis of this information, does not mean that the mental lexicon does likewise. Syntactic categories need not be listed separately in the mental lexicon as in a dictionary. Why should they be? If the syntactic category of a word is nothing more than a reflection of its meaning (and there will be more about this in Chapters 9 and 13), it will not—indeed, could not— be listed separately.

The view that has developed of how we access the mental lexicon is substantially different from the view we might originally have had on the basis of how we access a dictionary like the *OED*. Accessing the mental lexicon is far less restrained. We activate all the lexical entries compatible with the developing sequence of sounds entering the ear. If the same sequence of sounds has more than one meaning, we generally activate all the meanings compatible with that sequence and only subsequently does the activation of the contextually inappropriate meanings fall away. The 'generally' back there refers to the Tabossi and Duffy findings that in certain very constraining contexts, when the context does in fact prime some specific aspect of a meaning, it looks as if just the contextually appropriate meaning is activated (although we cannot be absolutely certain that there was *no* activation of the inappropriate meaning). An inevitable by-product of all this is that we must activate all manner of spurious, unintended, meanings. But given that the mental lexicon must reside within the neural circuitry of the brain, this is in fact a natural (if initially counter-intuitive) way for the system to work — the neural circuits are like so many combination locks, and as the speech input unfolds through time, so do the tumblers of the different combination locks move around, until eventually, just those combination locks whose sequences are completed spring open.

Unlocking the combination

How far should the combination lock analogy be pushed? To return to the opening theme of this chapter, might this not be just another example of an analogy that is simply inappropriate, simply wrong, or simply confusing? Maybe. But the continued use of the analogy helps explain one further fact concerning the manner in which we access the mental lexicon. The fact itself concerns a prediction that can be made on the basis of the way a combination lock operates, although when the prediction was originally tested, combination locks were probably the furthest things from the mind of the tester. The prediction, quite simply, is that wherever in the speech input a sequence is found that could correspond to a word, the lexical entry for that word should be activated. After all, rotate the dial of an old-fashioned mechanical combination lock and so long as the sequence of rotations contains, somewhere within it, the correct sequence for that lock, the lock will open.

Recognizing words within words

In the late 1980s, **Richard Shillcock** carried out an experiment to determine whether, for example, listeners activate the lexical entry corresponding to 'bone' if they hear the sequence 'He carefully placed the trombone on the table'. He used the cross-modal priming paradigm described earlier in which a word related to 'bone' (e.g. 'rib') would be flashed up on a screen at the offset of 'trombone'. Crucially, he also included sentences like 'He carefully placed the bone on the table'. Shillcock found that the amount of priming he got from 'trombone' to 'rib' was the same as that from 'bone' to 'rib'. In other words, the lexical entry corresponding to the word 'bone' is activated even when 'bone' is heard simply as part of the word 'trombone'. Subsequent studies (by **Patrizia Tabossi** and colleagues, in Italian) have shown similar effects with sentences equivalent to 'He broke all records for the new distance'. Here, the lexical entry corresponding to the word 'nudist' is activated (embedded in the sequence 'new distance').

There are, it turns out, limits on just when these kinds of effect occur. Jean Vroomen and colleagues at Tilburg University showed that it is not the case, for example, that 'speech' causes the activation of 'peach', or that 'teeth' causes the activation of 'tea'. This latter case is relatively easy to explain, and draws on the effects of *coarticulation* which we encountered in the previous chapter—the vowel in 'teeth' would not be the same as in 'tea' because aspects of the /th/ sound would be reflected in the vowel (because during the articulation of this vowel, the articulators would be moving towards the position required for the articulation of /th/). The vowel would also be shorter than that expected for 'tea'. So coarticulatory information, as well as duration, would lead to an acoustic mismatch against the input that would be expected for the word 'tea'. The first result ('speech' not activating 'peach') reflects the fact that the beginnings of words tend to coincide with the beginnings of syllables, and anything coming after the /s/ of 'speech' is more likely to be the continuation of a word starting with /s/ than the start of a new word.

Shillcock, 1990
Tabossi & colleagues, 1995

Shillcock, 1990 in ***Cognitive Models of Speech Processing*** After 'He carefully placed the trombone on the table' or 'He carefully placed the bone on the table' (played over headphones), Shillcock found that lexical decision times to the related probe word 'rib' were faster than to the unrelated probe word 'bun'. The amount of priming to 'rib' was the same for both sentences.

Tabossi, 1995 in **Journal of Memory and Language** This study explored the activation of the embedded word 'nudist' (the final /t/ is often dropped), underlined in 'the new discovery' and 'the nudist camp'. At the offset of the embedded word, lexical decision times to a word associated with 'nudist' ('beach') were recorded. In both these cases, the times were faster than after 'the summer camp'—suggesting that 'nudist' was activated after both 'new dis' and after 'nudist'. To check that this was not due to activation of 'nudist' after just 'new' (as predicted by the Cohort model), a second experiment included 'the new recovery', but there was no priming of 'beach' after 'new re'—suggesting that the priming seen before really was due to the 'nudist' in the 'new discovery'.

Too good to be true…

There is one important caveat to all these results. Although Shillcock, for example, found that words like 'trombone' speed lexical decision times to words like 'rib', Marslen-Wilson and colleagues, in their experiments on morphologically complex words, found that words like 'trombone' *did not* speed up lexical decision times to words like 'bone' (this result was described earlier on in the context of morphological processing). This is clearly at odds with the Shillcock result. So what is going here? Which result are we to believe? One difference between the studies is that Shillcock measured decision times to words related in meaning to 'bone', whereas Marslen-Wilson measured decision times to 'bone' itself. But it is unlikely, if the meaning of 'bone' is activated at the 'bone' in 'trombone', and if this is sufficient to speed the processing of 'rib', that it does not also speed the processing of 'bone'. So something else must be going on. A second difference is that Shillcock's 'trombone' was presented in a context ('He carefully placed the trombone on the table') whereas Marslen-Wilson's 'trombone' was simply a word heard in isolation. One possibility is that Shillcock's and Tabossi's contexts caused certain aspects of the meaning to be activated which would not have become activated if the word had been heard in isolation—indeed, Tabossi found in some of her other experiments that different contexts changed which aspects of the meaning of a word were activated (this was described earlier). Perhaps the shape aspect of the meaning of 'bone' (to take just one aspect of its meaning) was activated in both 'He carefully placed the bone' *and* in 'He carefully placed the trombone', and this contributed to the priming of 'rib'. In the absence of any context, the 'bone' in 'trombone' may not activate sufficiently those aspects of the meaning of 'bone' (perhaps including its shape) which would support the priming of 'rib' (or, come to that, 'bone').

This is, of course, pure conjecture. But it does exemplify the kinds of problems that we encounter when seemingly similar experiments can lead to quite different empirical results. Sometimes no clear resolution is possible, and one must simply wait for new data to definitively resolve the issue. In this case, no resolution is possible because the de Vroomen study mentioned earlier also found that 'trombone' primed 'rib', and yet in that study (performed in Dutch, in fact), the words were also heard only in isolation. So there is still something to be explained. Perhaps the explanation will turn out to hinge on subtle differences in the experimental procedure. In which case, what matters is that each researcher was consistent in respect of his or her experimental procedures. That way, we need simply suppose that subtle differences in procedure can change the sensitivities of that procedure to different kinds of mental process. So in effect, Shillcock and de Vroomen used a procedure sensitive to one lot of processes, and Marslen-Wilson a procedure sensitive to another (perhaps overlapping) lot. But the end result is that the conclusions they each reach remain entirely valid. So for the moment, we have to accept that there can be slight inconsistencies between different studies, and that the conclusions we reach as we survey the field need not change as a result. Inconsistencies of this kind are a hallmark, in fact, of science.

Too many words, too little time

If, as seems likely now, we do consider all possible hypotheses about what could be present at any point in the incoming speech signal, there surely would be an explosion of possibilities. How do we determine which are the right ones? How do we know that we should be reading about 'a new discovery' and that nudists have nothing to do with it? Why are we not even conscious of all these spurious words? The answer can be found in the manner by which we string the meanings of words together to give a coherent meaning to the entire sentence. Although the 'bone' in 'he placed the trombone' may cause the activation of aspects of the meaning of 'bone', the hypothesis that 'bone' is the intended interpretation of this sequence would have a poor fit with the context—the sequence corresponding to 'bone' in 'trombone' is more likely to be a part of 'trombone' than a part of (or all of) 'bone'; the hypothesis that it corresponds to the word 'bone' does not fit with any lexical interpretation of the sounds coming before it ('trom' is not a real word, and so whatever will come after 'trom' is more likely to be the continuation of a word starting 'trom' than a new word starting at the offset of 'trom'). Similarly, 'a new discovery' may well start off being compatible with 'a nudist' as well as 'a new dis(covery)' but when the 'covery' is encountered, the activation of the hypothesis corresponding to 'a nudist' will drop off relative to that for 'a new discovery', because only the latter hypothesis predicted 'covery' at that point in the signal. So although there may be some initial activation of (some aspects of) the representation of these spurious words, their poor goodness-of-fit will ensure that their activation will die down soon after—and on the basis of the Swinney and Tanenhaus results, we can anticipate that this dying down will occur within around 200 milliseconds.

Why it is that we are not conscious of the multiplicity of meanings that become activated at each moment in time is far less clear. Perhaps, what enters consciousness is the more salient (or active) hypothesis at any one time. Of course, how it is that anything enters consciousness at all is a mystery—perhaps the biggest mystery awaiting a solution. And until we solve that one (which is beyond the current remit) we must make do with the limitations that accompany the advances that have been made. One thing is for certain: whatever the limits of our understanding, we now know not to trust whatever intuitions we may have had on the basis of the dictionaries on our bookshelves.

ADDITIONAL READING

Aitchison, J. *Words in the mind: an introduction to the mental lexicon* (1994).
A very readable introduction to the topic. Surprisingly, spoken word recognition attracts fewer researchers than does visual word recognition, and there are thus few accessible textbook introductions to the subject (in fact, just this one).

Miller, G. *The science of words* (1991).
Although an older book, this is as close to a 'coffee table' book as you can get on this topic. It is richly illustrated, and covers many different aspects of words—their writing, their origins, their meanings, their mental representation, and so on.

EXPERIMENTAL METHODS

Word monitoring
People are told to listen out for a 'target' word, and to press a button as soon as they hear it in a sentence that is then played to them. Often used to assess the effects of context on word recognition.

Lexical decision
People have to decide whether a word they see or hear is a real word or not. The nature of the task requires trials in which this target word is in fact a non-word. Often used as part of a **priming** task

Naming latency
People have to say aloud as quickly as possible a word that is presented to them either on a screen or through headphones (in which case they repeat the word as quickly as possible). Like lexical decision, it is often used as part of a **priming** task.

Cross-modal priming
This task is based on the finding that lexical decision times or naming latencies are speeded up if the target word is preceded by a word or sentence which is related in meaning to that target. Typically, a person listens to a sentence or a word, and partway through this stimulus (the 'prime'), or at its end, a target word appears on a screen and the person has to perform a lexical decision on that target, or name it as quickly as possible. The task has been used to explore the factors that influence the activation of lexical candidates.

Form priming
This is similar to cross-modal priming, except that the prime and target are not related through meaning but through physical 'form'—'shark' and sharp' share physical (or acoustic) form, but not meaning. The prime tends to be spoken, and the target can be either spoken (and the person has to repeat it) or visual (in which case they might perform a lexical decision or naming task). This task has been used extensively to explore mismatch effects in word recognition.

Morphological priming
This is similar to cross-modal and form priming, except that the prime and target are related morphologically. This task has been used extensively in research on morphological processing.

Eye-tracking
The most recent method, it involves measuring where the eyes look, and when, as someone hears a word that refers to an object in front of them (or on a screen). It enables the researcher to plot how the probability of looking towards something varies as a spoken word 'unfolds'. Usually it involves the person responding to an instruction to pick something up or move it. It is a surprisingly sensitive measure, and has been used to explore the factors that influence the activation of lexical candidates, and importantly, the timing of that activation.

Gating
This is a task that we have not encountered in this chapter, but which was very influential in the 1980s. It involves playing to people increasingly larger fragments of a spoken sentence; for example, the first 50ms., then the first 100ms, then the first 150ms. and so on. After each fragment, the person has to write down the word they think they are hearing and, in some variants, how confident they are of their response. It provides an estimate of the amount of phonetic information (i.e. sound!) needed to identify a word.

KEY TERMS

Word formation: Morphology: morphemes: stems and affixes: prefixes, infixes, suffixes. (1–2)
Derivations and Inflections. (1–2)

Word recognition: Uniqueness point. (5–6)
Frequency and neighbourhood effects. (7–8)
Priming (methods): cross-modal, repetition. (8, 15, 19)
Priming (types): semantic, associative. (9)
Lexical competition (9, 11)
Acoustic mismatch effects: tolerance (11–12), place assimilation (14–17)
Marslen-Wilson's 'cohort' model. (12)
Morton's 'logogen' model. (13)
Forster's 'search' model. (13)
Morphological processing. (18–21)

Multiple meanings: Homonyms: homophones: balanced and unbalanced homophones: dominant and subordinate meanings (22–26)

Words within words: Embedded words (28–30)

KEY FINDINGS

Word recognition times
Words are often recognised before their acoustic offsets (4–5), and recognition times are faster for words with earlier uniqueness points (5–6); words that fit better with the context (5–6); words that are more frequent in the language (7); words that have fewer neighbours (7–8); and words that are preceded by words related in meaning or with which they are strongly associated (7, 9).

Activation of lexical candidates
All candidates compatible with the acoustic input are activated (8–10). Subsequent mismatch with the acoustic input reduces the activation level (11–12); similarly for mismatch with the context (5–6, 10). Acoustic mismatch that is 'legal', and due to word-final place assimilation, does not reduce the activation level of the lexical hypothesis corresponding to the 'unassimilated' form (15).

Morphology and the mental lexicon
Words are not represented as indivisible wholes, but are represented as stems and affixes (19–21).

Activation of multiple meanings
Alternative meanings of the same word form (homophones) are activated even in contexts that rule out one meaning (23–26). When one meaning is more common than the other, the alternatives are most likely activated in proportion to their relative frequencies of usage (24–25). If the context favours the dominant meaning, it can appear as if this is the only meaning that is activated (24); when the context favours the subordinate meaning, the activation of this subordinate meaning can be brought up to the same level as that of the dominant meaning, and it then appears as if both meanings are activated (24–25).

Activation of embedded words
Words embedded within other words cause their corresponding lexical representations to be activated, even if that activation is spurious (28). The onset of such spurious words must occur at a permissible syllabic boundary for such activation to occur (28).