Lexical Access

Intermediate article

Anne Cutler, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

CONTENTS

Lexical access occurs in speaking when the pronounceable word forms are found which encode the concepts the speaker wishes to communicate; and it occurs in listening when the meaning is found which is expressed in the word forms the listener has heard.

THE LEXICON

Communication of a message from a speaker to a hearer can take the form of word sequences which the speaker has never before produced and the listener never heard before. *This article tries to give a plain introduction to lexical access* - most people have probably not heard this sentence, but it is no problem for listeners presented with such a new utterance to understand it. This is because utterances, though they may be uniquely constructed, are built up of discrete units (such as *article, give, plain, access)* which speakers assume their listeners will already know. These discrete units we call 'words', and the stock of words which speakers of a given language use in speaking and listening constitutes the vocabulary. The mental lexicon is the mental representation of the vocabulary.

Entries in the mental lexicon may correspond to words such as *give* and so on, but they may also be other forms which speakers store as discrete units: fixed phrases such as *bon appetit,* manipulable idiomatic phrases such as *let the cat out of the bag,* productive derivational affixes such as *re-* or *un-* or *-ish,* inflections for pluralization, tense and so on, stems which occur in multiple words. That is, the forms in the mental lexicon are those which language users store as discrete entities, and they may or may not coincide with forms which are written as discrete words. (Nevertheless, in this text 'words' will serve as shorthand for lexical representations.)

Speakers begin with the intention to communicate a message, and to achieve this they must encode the message in words and articulate the resulting string of words. Listeners hear the string of words, which they must decompose into its word parts, and their first step in decoding the message is identification of the meaning associated with each word. Lexical access in speaking is the process of finding the lexical representations to express the desired meaning. Lexical access in understanding is the process of finding the lexical representations which correspond to the heard sounds. Both speaking and understanding thus require lexical access, but the two processes are the reverse of one another. It has been a matter of dispute whether there is a unitary mental lexicon that is drawn upon both in speaking and understanding.

LEXICAL ACCESS IN SPEAKING

The speaker begins with an intended message and must convert this into spoken sound patterns. Models of word production (Dell, 1986; Levelt *et al.*, 1999) agree that the conversion process consists of multiple stages, and in particular that retrieval of meaning and retrieval of sound are separate processes.

Retrieval of Meaning

In the model proposed by Levelt *et al.* (1999) it is assumed that the speaker effectively translates the intended message into individual lexically represented concepts. Each such concept has a unique lexical representation (the lemma) associated with it, and activation of the concepts expressed in the message causes activation of the lemmas to which they are connected. The lemmas contain the syntactic constraints associated with the concepts.

For example, Figure 1 depicts a single part of the lexical network which could be involved in expressing the message that this article is trying to give - *a plain introduction.* The lemma for *plain* is assumed to be activated by a corresponding conceptual node selected to express an attribute; the lemma *plain* accordingly includes the syntactic attribute of being an adjective. Other syntactic information which can form part of a word's lemma representation includes the expression of, for instance, tense and number (in verb lemmas), case (in noun and adjective lemmas for languages which mark case), or grammatical gender (in noun lemmas for languages with gender).

Related conceptual nodes are connected to one another, but each conceptual node is connected to only one lemma. Lemmas are not connected to one

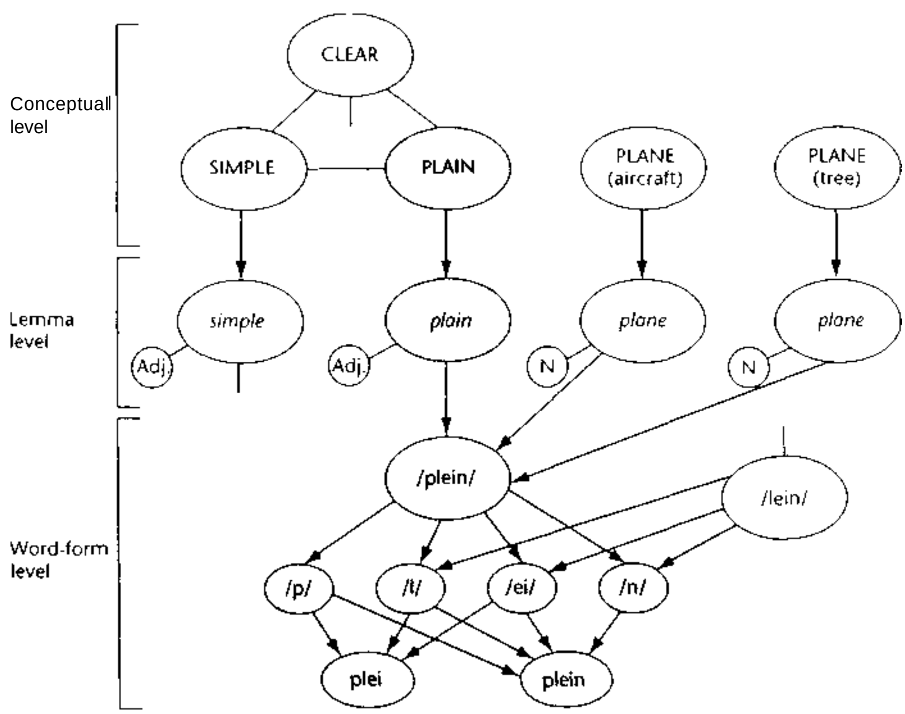

Figure 1. Fragment of a lexical network for word production on.

another at all; they receive activation from concepts, and pass activation to the phonological forms (the word-forms) which express them as sound patterns, but they do not pass activation to other lemmas.

Activation from the conceptual node may spread to other conceptual nodes to which it is connected, and these nodes may in turn send activation to their associated lemmas. At the lemma level, however, a selection takes place; the most-activated lemma is chosen to pass activation on to the word-form level.

Retrieval of Sound

The lexicon includes separate representations of phonological form, the word-forms. Each lemma is connected to just one word- form. (An exception may perhaps occur in the special case of a single word for which a speaker knows two pronunciations; for example, some British speakers can pronounce the word *garage* to rhyme with either *marriage* or *mirage.)*

Word-forms, of course, may be connected to more than one lemma, because word-forms are very often homophonous - two words can sound the same. Thus the word-form to which the adjective lemma *plain* sends activation has many other connections - from the noun *plain* (a level tract of land), and from the various lemmas *plane* (an aircraft; a kind of tree; a surface; to make level; and so on). Figure 1 shows just two of these. None of these lemmas are connected to one another, nor to the adjective *plain.*

Evidence for the connection of word-forms to more than one lemma, and in general for separation of the lemma and word-form levels, appeared in some studies of frequency effects in production. Speakers can produce common words more rapidly than uncommon words; but if an uncommon word happens to be homophonous with a common one (e.g. *plane* the name of the tree and *plain* the adjective), then the uncommon word is produced as rapidly as the common one. Apparently it is the frequency of the word-form which matters.

The word-form representation contains information about how the word is pronounced, in the form of connections to the appropriate phonemic representations, together with instructions for the compilation of these phonemes into possible syllables. In actual utterances, syllable boundaries are determined by the string of words as a whole rather than only by the lexical representations of the individual words. A syllable-final phoneme may become syllable-initial if the word following it begins with a vowel (and, in English and similar languages, especially if that second word is a function word). In *plane that wood!,* the [n] is syllable-final, but in *plane it!,* the [n] may be syllable-initial, and the precise phonetic realization of the [p] can differ in the two positions. Therefore the phonemic representations activated at the word-form level are not fully compiled, because their final compilation depends on the rest of the utterance. In Figure *1,* the word-form of *plain* is connected to four phoneme nodes (three of which, for example, it shares with the word-form of *lane),* and can activate two potential syllable nodes.

Slips of the Tongue

Errors can arise at all points in the process of producing an utterance. Slips of the tongue (see Fromkin, 1973) fall into two major classes: mis-selections and mis-orderings. Among the former are errors of lexical access, in which the wrong word is selected. These in turn are of three main kinds. Most common are mis-selections in which the meaning is similar but not identical - *next* instead of *last.* Mis-selections also occur in which the erroneously selected word is similar to the target word in sound though not in meaning - *single* instead of *signal.* And finally, blends of two words can occur - *bookstop* as a blend of *bookstore/bookshop.*

How do such errors arise? Semantic errors, understandably, arise in the process of accessing word meaning. They can be explained in terms of activation spreading among related conceptual nodes, leading to a lemma related in some way to the target being activated to the same or a greater extent than the target lemma. The nonintended lemma *(next)* is selected and from that point on word production proceeds just as if *next* had been the intended word.

Blends could also arise in the same way during the access of meaning representations. The difference between semantic errors and blends is that in the latter case the problem is carried beyond the lemma level: two equally activated lemmas could simultaneously pass activation to the word-form level, the two word-forms would then activate phonetic components at the articulatory level, and at this point the potential problem of competing candidates would perforce be resolved since the output would allow utterance of only one of the available syllable onsets, one of the available syllable nuclei, and so on.

Finally, form-based selection errors (malapropisms) could arise at the word-form level; activation spreading to the phonemic segments could be misassigned, leading to the syllabic packets appropriate for a similar form *(single)* being articulated, instead of those required by the target form *(signal).* The existence of both purely form-based and purely meaning-based slips lends further support to the two-stage model of lexical access in speaking. The 'tip-of-the-tongue' phenomenon also supports this division; a speaker who has a word on the tip of the tongue has accessed the word's meaning (and, in languages with grammatical gender, the speaker has also accessed the gender), but the pronunciation, i.e. the word-form representation, is for some reason temporarily inaccessible.

LEXICAL ACCESS IN UNDERSTANDING

The listener's task is to identify the words making up the speaker's message. Languages do not make life easy for listeners. The vocabulary of any language consists of tens or hundreds of thousands of words on average made up of only about 30 speech sounds, or phonemes (English has - depending on dialect - 40 or more phonemes, and is thus a relatively phoneme-rich language). Thus words inevitably resemble one another, and short words may fortuitously occur within longer ones; in consequence, in any spoken utterance the words uttered by the speaker are not the only words present in the speech signal. This would pose the listener little problem if words in speech were reliably demarcated with signals indicating where each word ends and the next begins. In many written texts (such as this one in English) such helpful signals are indeed available - white spaces between the words. That is not true of speech; spoken utterances reach the listener's ear as a continuous stream.

Segmentation

Figure 2 is a spectrogram of the phrase *A plain introduction to lexical access,* uttered by an American speaker. The gaps in the spectrogram correspond to speech sounds, not to pauses between words. Certain speech sounds cause a momentary obstruction of the vocal tract, resulting in a brief period of silence - the consonants [p], [t], [k] and [d] are among such stop consonants and they occur in this utterance - [p] in *plain,* [k] in *introduction, lexical* and *access,* and so on. In contrast, there is no pause at word boundaries - *plain* runs continuously into *introduction, lexical* into *access.* The listener must find the words despite the absence of clear word boundary signals. This utterance in Figure 2 also shows how easy it is for shorter words to appear by accident within longer ones - thus *plain* contains *play, lay,* and *lane; introduction* contains *duck; access* contains *axe.* Words may even occur across word boundaries, because of the continuous flow of one word into another: thus across the boundary of *lexical* and *access* there is a string of sounds consistent with *lack* and *lax.*

Figure 2. Spectrogram of the utterance 'A plain introduction to lexical access'. The display represents frequency on the vertical axis against time on the horizontal axis, with greater energy represented by darker shading. The transcription is aligned as closely as possible with the corresponding sounds on the spectrogram. There are clear breaks in the speech signal, but these do not correspond to breaks between words. The breaks occur whenever a speech gesture actually stops the flow of air through the speaker's vocal tract for a brief period - for instance the sounds 't' and 'd' in 'introduction'. In contrast, the individual words adjoin to one another continuously, without a break.

Listeners can use knowledge about their language to help them segment continuous speech into its component words. For instance, certain phoneme sequences cannot occur within syllables (e.g. [nl] in English), or even within words (e.g. [mg] in English). Listeners can use this information to find words; the beginning of *lexical* would thus be clearer in the phrase *in lexical access* than it is in *to lexical access.* In English, words more often than not begin with a stressed syllable, and this information is also useful in segmentation; *plain, lexical* and *access* all begin with a primary stressed syllable, and *introduction* has secondary stress on its initial syllable. Only the function words *a* and *to* are unstressed; this also is typical of English, and also used by listeners to assign grammatical category. Of course, all these types of information differ across languages, so that the way they are exploited in listening will also be language-specific.

Activation and Competition

Models of spoken-word recognition (e.g. Marslen-Wilson and Welsh, 1978; McClelland and Elman, 1986; Luce *et al.*, 1990; Norris, 1994) agree that words which are present in the speech signal are automatically activated in the listener's mind. Even words which are accidentally present (e.g. *lane* in *plain)* can be activated.

Simultaneous activation of all candidate words which are supported by information in the input means that a unique selection within the lexicon is not immediately possible; selection is achieved by allowing the activated words to compete with one another until one or more winners emerge. This process of competition provides a solution to the problems arising from the similarities between words and the embedding of words within other words; *plain* and *lane* compete for one portion of the input, *axe* and *access* for another, and so on.

Concurrent activation has been a feature of all models of spoken-word recognition since Marslen-Wilson and Welsh (1978). Competition was first proposed in the TRACE model (McClelland and Elman, 1986), and in the same form - competition via lateral inhibition between competitors - it is the central mechanism of the Shortlist model (Norris, 1994). In other forms it is also found in the other main models currently available, such as the Neighborhood Activation Model (Luce *et al,* 1990) and the latest Cohort model (Gaskell and Marslen-Wilson, 1997).

There is substantial evidence of activation of words embedded within other words, and of simultaneous activation of partially overlapping words. This sort of evidence is gathered in laboratory studies in which listeners hear words or parts of words, and also perform another task such as deciding whether a visually presented letter string (e.g. GIVE or FLERK) is a real word or not. Hearing a word facilitates acceptance of a related word - it is easier to decide that GIVE is indeed a word after just hearing *take,* for instance. Activation of embedded words has been proposed because it has been observed that they too provide such facilitation - e.g. the recognition of GIVE could be facilitated by hearing *mistake.* And simultaneous activation of partially overlapping words has been supported by experiments which showed that hearing a fragment such as *lec-*, which could be the beginning of several words (e.g. *lexicon, lecture)* facilitated words related to all of them (such as DEFINITION, COLLEGE).

The activation process is continuous, and can effectively use early co-articulatory information, as is shown by experiments in which words are cross-spliced with other words and with nonwords. It is hard to decide that *shrud* is indeed a nonword if its spoken form includes a 'shru-' taken from an utterance of *shrub* or *shrug.* Thus even nonwords which resemble words will activate lexical information, and the more similar the nonword is to a real word, the more effective it will be in activating word candidates. Activation of a lexical representation thus does not obligatorily require full presentation of the corresponding word form; partial information (in partial words, for instance, or in nonwords which in part overlap with real words) suffices to produce partial activation.

Simultaneous multiple activation of words does not, of course, necessarily entail competition between those words. Models such as TRACE and Shortlist predict that simultaneously activated words will compete by passing inhibition to one another. This too has been demonstrated in the laboratory. One way to show such effects is in experiments in which listeners hear nonsense strings, some of which have a real word in them; the task is to find any such real word. In the strings *obzel crivlish lakfid,* for example, only the last string contains a real word - *lack.* When such nonsense strings activate competing words, listeners find it harder to find the embedded words. Thus, *mess* is harder to find in *domess* (which could partially activate *domestic)* than in *nemess* (which activates no competitor). The more competing words may be activated, the more the recognition of embedded words will be inhibited. This kind of finding is direct evidence for competition between words.

Figure 3 shows simulations of the input *lexical access* in the Shortlist model (Norris, 1994). The input activates temporarily words such as *lecture* and *excema* which are eventually defeated by competition from *lexical; collapse* and *lax,* which straddle the word boundary, are also briefly active, but joint competition from *lexical* and *axe* defeats them in turn; initially, *axe* enjoys stronger activation than *access,* but by the end of the input the two most activated words are, as intended, *lexical* and *access.* Competition between candidate words which are not aligned in the signal helps to achieve segmentation of the speech stream into individual words. Thus although the recognition of *lexical access* involves competition from other words, this is eventually overcome by joint inhibition from the actually spoken words. The competition process, and its concomitant constraints, can so efficiently result in victory for words which are fully present in the signal that concurrent activation of partially present words, or of words embedded within other words, is simply a low-cost by-product of the efficiency with which the earliest hints of a word's presence can be translated into activation.

An activated word form passes activation on to the meaning associated with it. In word recognition models activation is assumed to cascade continuously through the levels, rather than to wait for a clear winner at each level. Thus activation of the word form *plain* can pass activation on - in this case, of course, to several meanings rather than just one. The string *a plain* is ambiguous as to whether the word-form *plain* is noun or adjective, and which of several interpretations of either syntactic form might be intended. Empirical studies (e.g. with the priming tasks described above) have indeed shown that multiple meanings of a homophone can be simultaneously accessed. However, the integration of the word's meaning with the rest of the syntactic and semantic context is also assumed to be a continuous process, as a result of which the intended meaning can very rapidly be selected. In the context of a story about flight, the message interpretation component might very quickly select the noun *plane* in the sense of aircraft, on semantic grounds. Syntactic considerations would suggest that *A plain* followed by *introduction* is best interpreted with the adjectival meaning of *plain.*

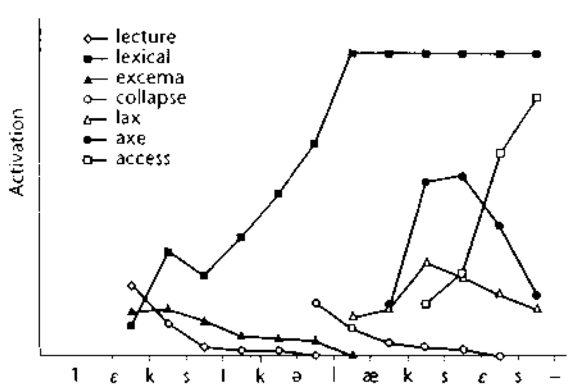

Figure 3. Shortlist simulations of word activations given the input 'lexical access'.

Slips of the Ear

Errors can, again, arise at any point in the process of comprehending an utterance. Errors specifically in the process of lexical access imply that the outcome of the competition process is a word or sequence of words which does not correspond to what the speaker actually said. Such slips of the ear are particularly likely when the input is unclear - due to background noise, for example, so that weak acoustic support is provided for a number of possible candidate words. Thus an old popsong contained the words *she's a must to avoid;* very many listeners independently interpreted this as *she's a muscular boy.*

Slips of the ear often (as in this example) produce implausible results, and may produce reports of low frequency words or even nonwords (see Bond, 1999). This suggests that frequency and plausibility do not play a strong role in the activation and competition process.

The operation of the strategies constraining the competition process can be seen in the patterns of slips of the ear. In English, for instance, slips are much more likely to produce interpretations in which strong syllables are word-initial and weak syllables are not word-initial - as in the example, in which *must to avoid* (strong, weak, weak-strong) is thought to be *muscular boy* (strong-weak-weak, strong).

THE INDEPENDENCE OF LEXICAL ACCESS

A recurring question in research on language processing is whether the different components of the process are independent, or whether they can be influenced by processing decisions made by other components which are logically later in the processing chain. In models which preserve independence ('autonomous models'; e.g. Levelt *et al.,* 1999 for word production, or Norris, 1994 for word recognition), information feeds forward only - in speaking, from conceptual formulation to articulation, and in listening, from acoustic-phonetic processing to message interpretation. In models which allow non-independence ('interactive models'; e.g. Dell, 1986 for word production, or McClelland and Elman, 1986 for word recognition), there is feedback, e.g. from phonological encoding to lemma selection in speaking, or from word activation to phonetic processing in listening.

This is one of the most controversial issues in cognitive science, and it is a subject of continuing debate. For lexical processing in both speaking and listening, however, strictly feedforward models have been proposed which adequately account for the known empirical evidence (Levelt *et al.,* 1999; Norris *et al.,* 2000). Considerations of parsimony suggest that the simplest model consistent with the evidence should be preferred; the proponents of these models therefore argue that lexical processes should not be held to involve feedback connections.

In speaking, considerations of added efficiency have not been invoked in favor of feedback; the strongest evidence which has been used to argue for feedback in fact comes from performance failure, i.e. from slips of the tongue. First, semantic errors are more likely if the error is similar in sound to the intended word; that is, *plane* is more likely as a semantic error for *train* than for *boat.* And second, single-phoneme errors tend to produce another real word rather than a nonword: *plane* is more likely to be mispronounced *plate* than *plake.* However, both of these tendencies also arise in Levelt *et al.'s* (1999) feedforward model. Errors which result in real words are more likely to be missed by the monitoring component in that model. The split-stage architecture in the model further provides a natural explanation for phonological similarity in semantic errors. If two equally active lemmas are passed on to the word-form level instead of one, the error word-form is more likely to be selected at that level if it is similar to (and its phoneme connections hence receive activation from) the intended word-form.

In listening, similarly, the performance of the lexical processor cannot actually be improved by feedback to prelexical processing. Suppose the input *plai-,* with co-articulatory information in the vowel, is enough to activate *plain* to such an extent that it is the most well-supported word in the ongoing lexical competition process. Feedback from the lexicon can then determine that phonetic processing identifies an [n]. But this offers no assistance to the recognition of *plain,* since *plain* is already the most activated word. It would also not assist word recognition to pass down information from a number of equally supported competing words, if *plain, place, plate* and *plague* all passed down activation to their separate terminal phonemes, the phonetic processing level would simply receive more incorrect than correct information. The information in the signal would ultimately have to determine the decision at the phonetic level, so that no useful role would be played by the feedback.

References

Bond ZS (1999) *Slips of the Ear: Errors in the Perception of Casual Conversation.* New York: Academic Press.

Dell GS (1986) A spreading-activation theory of retrieval in sentence production. *Psychological Review* 93: 283-321.

Fromkin VA (1973) *Speech Errors as Linguistic Evidence.* The Hague: Mouton.

Gaskell MG and Marslen-Wilson WD (1997) Integrating form and meaning: a distributed model of speech perception. *Language and Cognitive Processes* 12: 613—656.

Levelt WJM, Roelofs A and Meyer AS (1999) A theory of lexical access in speech production. *Behavioral and Brain Sciences* 22: 1-38.

Luce PA, Pisoni DB and Goldinger SD (1990) Similarity neighborhoods of spoken words. In: Alrmann GTM (ed.) *Cognitive Models of Speech Processing,* vol. 7, pp. 122-14. Cambridge, MA: MIT Press.

Marslen-Wilson WD and Welsh A (1978) Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology* 10: 29-63.

McClelland JL and Elman JL (1986) The TRACE model of speech perception. *Cognitive Psychology* 18: 1-86.

Norris D (1994) Shortlist: a connectionist model of continuous speech recognition. *Cognition* 52: 189-234.

Norris DG, McQueen JM and Cutler A (2000) Merging information in speech recognition: feedback is never necessary. *Behavioral and Brain Sciences* 23: 299-370.

Further Reading

Brown C and Hagoort P (eds) (1999) *Neurocognition of Language,* chaps 4 (Levelt), 5 (Cutler and Clifton), and 7 (Price, Indefrey and van Turennout). Oxford: Oxford University Press.

Friederici A (ed.) (1998) *Language Comprehension: A Biological Perspective,* chaps 1 (Frauenfelder and Floccia), 2 (Cutler), and 3 (Zwitserlood). Heidelberg: Springer.

Grosjean F and Frauenfelder UH (eds) (1996) Spoken word recognition paradigms. Special issue of *Language and Cognitive Processes* 11: 553-699.