Word Recognition and Lexical Access

Gregory W. Yelland
Department of Psychology
Monash University

The recovery of meaning from any sentence (spoken or written) requires more than a consideration of the meanings of the individual words it contains. Words, though, are the building blocks of comprehension: it is the properties they represent that provide the primary data the language processing system is to use in computing sentence meanings. The processes responsible for the recovery of the properties of words are those of lexical access, and more generally, word recognition.

1. Word recognition and lexical access

Word recognition is the product of a set of processes charged with the task of rendering a belief about what word was just heard or seen. This set of processes is required:

(a) relate a word's sensory input to a corresponding mental representation stored in memory;
(b) check the accuracy of any match between memory and stimulus representations;
(c) make the contents of the accessed memory representation, such as the word's form descriptions (phonological and orthographic), it's meaning(s) and it's syntactic properties, available to higher level processors, where single words can be integrated into interpretable phrases and sentences; and where necessary
(d) to resolve any discrepancy that may exist between the input and accessed representations (e.g., in the case of a speech or spelling error), so that the product of recognition is the speaker's (or writer's) intended form.

(For more detailed discussion to this point, see Norris, 1986, and Bradley and Forster, 1987.)

There is no single theory of how word recognition is achieved. The point of dissension between theorists stems primarily from vastly different views about the processes that support the first step, the mapping between a stimulus description and its representation in memory. These are the processes of lexical access, and are the focus of this article.

The term lexical access derives from the concept of a mental lexicon (Oldfield, 1966), a memory system dedicated to the storage of all that we know about the words in our vocabulary. Any single representation, or entry, in the lexicon is said to contain a language user's collective knowledge about the form, syntactic and semantic properties of a given word. The processes of lexical access, then, are those that locate a representation of a word in lexical memory that matches some representation of the input.

This is intentionally a very broad definition of "lexical access", since different models of lexical access propose very different mechanisms for the mapping between an input description and an appropriate entry in the lexicon. The differences between models lie primarily in two areas, the means by which the stimulus description is compared with candidate lexical representations, and the extent to which information from higher levels of the language processing system, and more general cognitive systems, can be recruited to aid the
stimulus-driven processes. The latter is of particular interest, for it is central to the more general issue of the architecture (i.e., organization) of the language processing system.

The following sections outline the basic properties of current models of lexical access and how each accounts for some of the more robust research findings reported in the word recognition literature; that is, facts about word recognition that are in need of an explanation. Finally, the implications for the debate about language processor architecture are examined.

2. Spoken and written inputs

One issue to be faced in the modeling of lexical access is the existence of dual input representations for any given word; its spoken form and its written one. The problem is that the characteristics of the sensory data that carry the linguistic information are very different for the two input modalities. In speech, the input is an acoustic waveform. As such, the information necessary for speech perception is deployed over time in such a way that no single instant of input carries sufficient information to support recognition. In addition, there is the problem of speakers' coarticulation of adjacent phonemes. This has the effect of smearing the acoustic features of adjacent phonemes across the speech waveform in a way that rarely yields a linearly-ordered sequence of discrete phonemes for the listener (Liberman, Cooper, Shankweiler and Studdert-Kennedy, 1967; Pisoni and Luce, 1987). That is, the stimulus features thought critical for lexical access (the phonemes) are not available directly from the speech waveform, rather, from some abstraction of the acoustic properties.

This is not the case for the written word. Here the critical stimulus features are routinely available in the surface form of the input. In print, the boundaries of words are clearly defined, as are the individual letters of a word and their orthographic features. Further, input features are not deployed over time as they are in speech, but are static; discrete letters and words are available to a single glance (or fixation). This provides readers with the added advantage of an opportunity to review the stimuli. Research on eye-movements during reading have shown that, when reading sentences, readers sometimes take a second look at a word and spend longer looking at some words than than they do on others. For details of this work see, Rayner and Pollatsek (1987, 1989 pp.113-187) and O'Regan (1983). This is a luxury rarely afforded the listener.

While acknowledging the differences of input form across modalities, the majority of models of lexical access have been developed from data obtained from experiments with written words alone. Not because of any fundamental belief about the written code being the appropriate medium for examining the processes of lexical access, but because of the relative ease of experimentation with the written word. Experimentation in written word recognition afforded the experimenter greater control over the preparation of materials, less complicated measurement techniques, and a more rapid research progress, than was the case for research with the spoken word. Spoken stimuli introduce sources of variation that are not present for written stimuli, such as; differing voice onset times across words, and changes in intonation across conditions. As such, until recent developments in the digitization of acoustic stimuli, the preparation and presentation of spoken word stimuli, and the accurate measurement of responses to those materials (particularily reaction time responses) was very exacting and time consuming work. The assumption of this work has been that the data obtained within the written domain reflect lexical processing generally, and not written word recognition specifically. As such, most models provide a unified account of lexical access, proposing that access processes operate similarly over spoken and written inputs, beyond a point of convergence in their internal representations. As we shall see, this point of convergence differs across models of lexical access; it is that point at which inputs are represented in a similarly abstract form to the lexical representations.
3. Models of lexical access.

After around 25 years of research in lexical processing there is, as yet, no uniformly held model of lexical access. Rather, what has developed over this time is a large variety of views on what such a model should look like (for detailed descriptions of these see Norris, 1986, and Taft, 1991). However, Forster (1989) points out, that the range of existing models fall into basic "families of overlapping models" (p.75): activation models, computational (or search) models, and a group that Forster has called "hybrid" models, for they draw on properties of both of the other families. It is in this light that we shall examine some of the more influential candidates for a theory of lexical access.

3.1 Activation Models

A basic property of activation models is the concept of content-addressable memory: perceptual information extracted from an input (the spoken or written word) feeds directly into a set of "word detectors" of some form, which are themselves addresses in lexical memory. There, perceptual information is seen as something akin to neural activity distributed across all word detectors sharing any part of the information recovered from the stimulus. Lexical access is a function of the level to which a word detector is activated. There are two main classes of activation models: the logogen model (Morton, 1969, 1977, 1979) and it's variants, and the network models of McClelland and Rumelhart (1981, 1986, Rumelhart and McClelland, 1986) and their colleagues.

3.1.1 The Logogen Model

Morton's original logogen model (1969) is the foundation for all activation models of lexical access. Perceptual information recovered from the input word, such as orthographic or phonological features, is fed directly into a set of feature counters. The feature counters are called logogens; each logogen corresponds to a single entry (word) in lexical memory. An input containing a feature of a particular word will increment the feature count of that word's logogen by one. It also increases the feature counts of the logogens of all other words containing that feature.

An increment in the feature count is represented as an increase in the activation level of the logogen. Lexical access occurs when a logogen has collected sufficient features for its level of activation to have reached some predefined threshold of activation. It is then said to "fire", making its semantic and syntactic properties available to output processes. Thresholds are set independently for each logogen. That is, the criterion for a match between stimulus representation and its lexical representation is not uniformly specified for all logogens, but is differentially tuned for each logogen.

Morton proposed that logogens were not only sensitive to phonological and orthographic (i.e., perceptual) features, but that they also count semantic features from related words and contexts, making lexical access context sensitive (this will be discussed further in section 4.3). These semantic features are fed back from the central "cognitive system" (which is responsible for comprehension) to increase the activation of the logogen of any word to which those semantic properties are relevant. As a result, a word presented within a related context, or following a word to which it is related, will already be partially activated by the semantic features it has collected. Thus, fewer perceptual features would be required to push activation above the threshold.

Recently, in light of evidence of differing effects of written and spoken inputs on performance in tasks requiring lexical access – particularly the absence of cross-modal priming (e.g., Winnick and Daniel, 1970; Morton, 1979) – Morton revised his model,
separating visual and auditory inputs. He proposed two sets of input logogens: visual input logogens, which count only orthographic and semantic features, and auditory input logogens, which count phonological and semantic features. More recently still, Harris and Coltheart (1986) have made the distinction between input logogens and output logogens. They propose that information recovered from the cognitive system (such as the semantic and syntactic properties of the input) is fed into separate sets of visual and auditory output logogens, for writing (i.e. spelling) and speaking respectively. However, it is not clear how these output logogens are intended to operate.

3.1.2 Network (Connectionist) models

Network models consist of multiple layers (or banks) of logogen-like detector units. Each layer in the network contains units reflecting a higher level of representation than that below it. What is represented at each of these layers depends on whether the lexical memory has local representations, as in the interactive activation model (McClelland and Rumelhart, 1981; Rumelhart and McClelland, 1982), or distributed representations, like the more recent parallel distributed processing models (Rumelhart, et al., 1986; McClelland, et al., 1986, Seidenberg and McClelland, 1989). With local representations, there is a single detector for each linguistic unit (e.g., each word, phoneme, letter, etc.). In the case of distributed representations, a word, for example, is not represented by a single detector unit: rather, it is represented by a specific pattern of activation distributed across numerous orthographic, phonological and semantic units within the network.

In the interactive activation model (McClelland and Rumelhart, 1981) the layers of detector units were visual feature units, letter units, word units, and semantic and syntactic units, for written inputs. Auditory feature and phoneme detector units replace visual feature and letter level units, for the case of spoken inputs. Each unit in the network is connected to all units in the adjacent layers, and to each unit within its own layer, giving one large network throughout which activation is able to spread. The connections are either excitatory or inhibitory. Activated units inhibit units within their own layer, but excite units in the layers immediately above and below. In this sense, the network is fully interactive, with activation continually spreading in three directions (up, down and across) throughout the network. The amount of activation passed from one unit to another differs with the strength of the association between the two units.

Processing within the network is parallel, in two senses: it is carried out simultaneously within and between all layers, and all letters of a word are processed simultaneously (it is not clear what happens in the case of the spoken word, where all features are not available at the same time). For example, if the input stimulus is the word ‘late’, the word units for all words containing the letters l, a, t and e receive excitatory input from the letter units (e.g., late, date, lake, bark, and so on). These word units activate appropriate semantic units, while at the same time sending inhibitory activation to all other word units, and sending excitatory activation back to letter units for each letter in those words. Initially, there are a large number of active units in each layer. The spread of activation continues throughout the network until the total network approaches a state of equilibrium; this will occur when one word unit is sufficiently activated to strongly inhibit all other word units. This unit is taken to be the most likely representation of the input word. Since the unit for late shares all letters in common with the input, it ought to eventually accumulate the highest level of activation.

Parallel distributed processing (PDP) models of word recognition retain the general principles of the interactive activation model, with a parallel detector-based system in a network of interactive, spreading activation. The main difference is the way in which the properties of words are represented. For example, Seidenberg and McClelland (1989) suggest that word recognition requires only three sets (layers) of representational units, orthographic,
phonological and semantic. The representation of a word is given by the sum of active representational units in each of those sets.

Representational units do not map directly from one to the other (i.e., an orthographic unit cannot directly activate a phonological unit); rather, each is connected to the other by a separate layer of hidden units. These units “are those whose only inputs and outputs are within the system we are modeling.” McClelland, Rumelhart and Hinton, 1986, p.48). They mediate the interactions between the pools of representational units. So, when a letter string is presented, it is first encoded into a pattern of activation over the orthographic units. Encoding is in a form of Wickelgren (1969) triples, where the letter string LATE, for example, would be represented as the set of letter triples _LA, LAT, ATE and TE_ (where _ represents a terminal position). Here, where representations are distributed, such coding preserves information about letter order. However, Seidenberg and McClelland point out that the triples are not represented by single units, but are encoded as a distributed pattern of activation over the set of orthographic units, although, the means by which this is achieved is not clear. Spoken inputs are thought to be represented by similar triplets of phonemes.

Active orthographic units activate units in the hidden layers between the orthographic and both the phonological and semantic representational units. The hidden units then activate the set of potentially related semantic and phonological units. The strength of the connection (and thus amount of activation) between any two units is weighted in accordance with the level of their association, and is continually modified as a function of experience. Activation continues to spread interactively throughout the network until a distinct pattern of activated units emerges across each of the representational layers. Collectively, the distributed pattern of active representational units provides orthographic, phonological and semantic descriptions of the input.

3.2 Computational (Search) Models

The critical features of the models within this family (and that which distinguishes them from the activation models) is that the perceptual information recovered from the input does not have direct access to the lexicon. Instead, some abstract representation of the input is used to search the lexicon for a potential match with the stimulus. In this sense they are computational models of lexical access, since they "assume that recognition is carried out by a representational system that operates by explicitly manipulating and comparing symbols" (Forster, 1989, p.76)

The most widely considered of the computational models is Forster's serial search model of lexical access (Forster 1976, 1979, 1989, 1990; 1992; Bradley and Forster, 1987). This is a two-stage model in which a peripheral access file is searched until a match is found between an abstract representation of the input (the access code) and an entry in access file. A match provides access to the entry in the lexicon where information about the input word can be recovered.

Forster proposes access files of three representational types: orthographic and phonological access files for written and spoken inputs respectively, and a semantic access file for production purposes (which we will consider no further). Forster argues that the access files are divided into "bins" containing a subset of the lexicon over which a search is made, alleviating the need for a search of the entire lexicon each time a word is seen or heard. All entries within a bin carry the same address, determined by the hash coding of some aspect of their orthographic or phonological form specifications, and are ordered according to their frequency of occurrence in the language.

The term "hash-code" is borrowed from the domain of computer programming; it's function is to translate an input pattern into a unique address in memory. As Forster (1992) explains it:
“Whenever data about a particular pattern is to be stored, the hash-code is used to decide where to store this information. Similarly, whenever data about a particular pattern are to be retrieved, the hash-code can be used to determine where that information has been stored.”

In terms of lexical access, the function of hash-coding is to map the features of an input word on to an address in lexical memory. Forster gives a simple example of a hash-coding mechanism where each letter of the alphabet is assigned a number corresponding to it's position in the alphabet. For example, then, the address generated for the word ‘fish’ is 42 (i.e., 6+9+19+8); but 42 is also the address generated for ‘tape’ and ‘rare’. Such “collisions” are likely occur regardless of the hash-coding system. Forster argues that the best way to cope with them is to put words with like addresses together (in a bin), and search among them for a match with the input form.

A critical step, then, is the selection of the correct bin over which the search is to be made. This is achieved by a hash coding of the input's access code; the search is made over the bin with the corresponding hash code. It is a frequency-ordered serial search that proceeds from the entry of highest frequency to that of lowest frequency, until a match with the access code is found. The criterion for a match is the same for all entries, but it can be uniformly lowered or raised according to circumstances (Bradley and Forster, 1987).

Once a match is obtained a post-search check is carried out comparing the orthographic description contained in the lexical entry, with the full orthographic properties of the input word. The equivalent comparison of phonological properties would be carried out for a spoken input. Should this detailed comparison fail to reveal a close match, then the search would resume immediately. When a ‘detailed’ match is obtained, the contents of the corresponding lexical entry are made available to the appropriate processors. (For a more detailed discussion, see Bradley and Forster, 1987.)

A necessary consequence of organizing ‘bins’ according to form and frequency is that, since access to the lexicon can only be achieved via the bins, lexical access must be form-dependent; i.e., the semantic properties of prior inputs are unable to aid lexical access as they are in the interactive (activation) models described above. Thus, one of the central claims of the serial search model is that lexical processing is autonomous, in that it is driven entirely by the form properties of the input.

Recently, Forster (1989, 1992) has posited the notion of a parallel search model, in which serial searches of all bins are carried out simultaneously. This negates the need for the hash coding of input representations, and allows a complete search of the lexicon in the time it takes to search one bin.

3.3 Hybrid Models

The family of ‘hybrid’ models has more members than either the activation or search model families, and there is a greater disparity amongst its members. They are models of lexical access that borrow features from both the activation and search families (Forster, 1989). Most have, as a first step, some form of detector-based (or activation) system responsible for generating a set of potential candidates. This candidate set is subsequently searched (or checked) to find the candidate that provides the best match with the input. Further, for most, the candidate set is usually determined on a purely form basis, with contextual factors (such as, semantic and syntactic information from prior inputs) only coming to bear in the final stage of selection.

Some of the more prominent models sharing these features are: the verification and activation-verification models of Becker (1980) and Paap, Newsome, McDonald and Schvaneveeldt (1982), respectively; Norris's (1986) checking model and Marslen-Wilson's 1987-version of the cohort model (a model of spoken word recognition). Discussion here, will
be restricted to Becker's (1980) verification model, for it is good example of the hybrid type; it combines aspects of both Morton's (1979) logogen model and Forster's (1976) serial search model. Input features feed directly to the logogen-like network. However, rather than waiting for one logogen to reach threshold, Becker argues that the lexical processor assigns the form representations of all partially activated logogens to a temporary candidate set; the sensory set. Candidates are assigned to the sensory set according to their frequency of occurrence. Verification of the most likely candidate is achieved by a frequency-ordered serial search of the sensory set to locate a candidate in the sensory set whose form representation provides the best match with the representation of the input.

When words are presented in the context of other words, Becker argues that a second candidate set is generated; the semantic set. The semantic set contains the form representations of words that are semantically related to the input. The purpose of the semantic set is to aid the recognition of a subsequently presented word. That new word can be verified against the semantic set, while the perceptual set is being generated. As a consequence, a word that is in the semantic set will be recognized faster than one that is not, for verification of the latter must await the completion of the sensory set.

4. Comparison of the Models

This is not really a question we can ask at this stage, for it is difficult to make any definitive statement about which model of lexical access, or even which family of models, will eventually hold sway. What is important at this stage is how the models fare in accounting for the set of known facts about word recognition contained within the wealth of accumulated research data. What will eventually distinguish between models, is not what they can account for, but what they cannot.

We can already make some distinctions between models on the basis of facts they find difficult to explain. For example, one of the most robust effects to be observed in studies of word recognition is the frequency effect (e.g., Rubenstein, Garfield and Millikan, 1970); the observation that it takes less time to make a recognition response to a word that occurs frequently in the language than it does to respond to a word occurring less frequently. For the search model, the frequency effect is a natural consequence of the frequency-ordered organization of entries within the bins of the access files. In any search, the entry for a high frequency word will be encountered first, and so accessed before that for a low frequency word.

Activation models have more difficulty in providing an explanation. For the logogen models, it is argued that logogens for high frequency words have lower activation thresholds than those for words of lower frequency. Thus, fewer input features need to be counted before threshold is reached. For network models the mechanism is similar, with information about word frequency carried in the strengths of the connections (i.e., activation) between units in the network; units representing high frequency words are more strongly activated than are those representing low frequency words. The problem for these accounts is that a low frequency word (e.g., ‘blight’), that is very similar to a word of higher frequency (e.g., ‘bright’), may never be recognized, since the combination of frequency and form activation should make ‘bright’ the more highly activated unit. The basis of the problem is that information about input form and frequency are carried by the same mechanism. Consequently, in order to ensure that low frequency words will be recognized, the additional activation to come from information concerning frequency has be kept to a minimum; a proposition that is not in line with the usually large effects of word frequency on the speed of word recognition. (See Balota and Chumbley, 1985; cf. Bradley and Forster, 1987; Savage, Bradley and Forster, 1990; for further discussions to this point.)
Another well documented set of findings in the literature are the context effects, amongst which the semantic priming effect stands prominent (Meyer and Schvaneveldt, 1971). Here, words preceded by a related word (e.g., ‘king - queen’) are responded more rapidly than when preceded by and unrelated word (e.g., ‘cloud - queen’). The shoe may be on the other foot, here, for semantic priming effects are potentially problematic for a search model. Activation models, logogen and network alike, are interactive models of lexical access. They allow semantic information from a prior context to be fed back to all related word units, either as semantic features (logogen models) or excitatory activation (associative networks). Units activated in this way require less activation from input features for access to be achieved.

Semantic priming is a problem for the serial search model, only if the priming effects are shown to be effects on lexical access; that is, if they influence the access procedures. The search models has no mechanism whereby semantic information has access to lexical procedures; lexical access can only be achieved via an access code constructed from the physical features of the input word. As such, it does it does not permit interactions between lexical level processes and those at higher levels in the language processor, where semantic information may be available. The only recourse, is to make an explanation in terms of post-access processes; that is, processes occurring after the mental representation of a word has been accessed (Forster, 1990).

There is, some evidence to this point from recent work examining priming effects of a different type, ‘masked priming’ effects (Forster and Davis, 1984). In the masked priming task subjects make a lexical decision response (i.e., word-nonword decision) to a target following a very brief exposure of a prime that has been forward masked. Subjects have no conscious awareness of the prime, and so are unable to draw on more central processes to aid target recognition. Since priming effects are still observed, but for word targets only (e.g., Bradley, 1980; Davis, 1990; Forster, Davis, Schoknecht and Carter, 1987), it would seem that masked priming reflects purely lexical processing. In a case to this point, Bradley (1990) compared the effects of semantic prime-target relatedness with those for prime-target pairs related only by virtue of orthographic form, under both masked and unmasked (visible) priming conditions. She found large form priming effects when primes were masked, which were not evident when primes were not masked. For semantically related pairs the pattern was reversed, with semantic priming effects largely restricted to the case of unmasked priming. This pattern of results is consistent with the notion that lexical processes are form based, while processes which utilize semantic information operate at some point after lexical access procedures have made a semantic representation available. In this light, semantic priming effects do not pose a problem for a search model.

Word frequency effects and the semantic priming effect are only two examples of the range of phenomena for which models of lexical access need to provide an account, and where the distinctions between models are to be found. Before we can decide which model fares best, we will need to decide which of the reported phenomena are those to be given a lexical explanation, and which are not.

5. Issues for future consideration

It is difficult to speculate about what directions research in lexical access might take in the future, and about which issues are likely to be those of most prominent concern. However, foremost amongst them will be the debate over whether lexical access involves computational processes addressing local representations, or parallel interactive processing across a network of distributed representations. At a lexical level, the notion of parallel distributed processing is a challenge to the concept of "lexical access" (Seidenberg, 1989). With distributed processing, word recognition is the simple consequence of a build up, over time, of activation within an
associative network of orthographic, phonological and semantic units. Seidenberg claims that there is no point in “terming a specific moment in this process ‘lexical access’” (p.55).

At a more general level this issue is one that speaks to the overall organization of language processes and the cognitive system more generally. One of features of parallel distributed processing systems is an underlying assumption of a uniformity of psychological mechanisms supporting cognitive function. This is at complete odds with Fodor's (1983) modularity hypothesis, in which he claims a separation of function between specialized input modules and a more central cognitive system which receives their outputs. In that scheme the language processor is an input module; that is, language processes, like lexical access, are specialized, and not part of a central cognitive system. Two of the properties Fodor attributes to modules are information encapsulation and cognitive impenetrability; taken together they mean that the central system only has access to the outputs of the module, and that the module cannot receive information from the central system.

The latter are testable hypotheses, since jointly they prohibit interaction between any process internal to the language module and the processes of the central cognitive system. We are likely to see considerable future research addressing this issue; this is research that specifically investigates whether there is any evidence for lexical-central interaction.

See also: 10-027; 10-030; 10-032; 10-036

6. References


