3

Visual Word Recognition: Theories and Findings

Stephen J. Lupker

The topic of "visual word recognition" may have the largest literature in Cognitive Psychology and, therefore, a chapter on the topic must be selective. This chapter will first place the relevant issues in a historical context and then review the basic visual word recognition phenomena within the context of current models. It will then be argued that any successful model of visual word recognition needs to incorporate the assumption of "interactivity," that is, that the various components of the visual word recognition system (i.e., orthographic, phonological, semantic) mutually activate and inhibit each other while a word is being processed (see also Van Orden & Kloos, this volume). (Hereafter, the term "word recognition" will be used as shorthand for the term "visual word recognition.")

What is "word recognition"? At least until the appearance of Seidenberg and McClelland's (1989) connectionist model of reading, word recognition was typically thought of as the process of going from a printed letter string to the selection of a single item stored in lexical memory. Lexical memory, or the "lexicon," is a mental dictionary containing entries for all the words a reader knows. Thus, word recognition was essentially synonymous with the terms "lexical access" or "lexical selection." Such a definition, of course, assumes that words are represented as lexical entries in memory. Seidenberg and McClelland's model explicitly denied the existence of such representations, arguing instead that representations were distributed across sets of simple subsymbolic processing units. To the extent that models of this sort have been successful, they have forced theorists to contemplate the possibility that some of the standard assumptions about the architecture of the word recognition system should be altered.

What appears to be an equally important aspect of Seidenberg and McClelland's (1989) model was that it contained a straightforward outline for how semantics should be integrated into the word recognition system. That is, semantic information was assumed to be represented no differently than other types of information (i.e., orthographic and phonological) and all of these mental representations were assumed to follow the same rules of activation. As such, this model represented what I would argue was the first complete model of word recognition. This is a crucial point because, as will be argued in this chapter (see also Balota, Ferraro, & Connor, 1991), any successful model of word recognition will need to have a mechanism for explaining the impact of semantics, both the impact of the semantic context within which a word is processed and the impact of the semantic attributes of the word itself (Whaley, 1978).

Historical Context

Most of the early models of word recognition (e.g., Gough, 1972; Massaro, 1975; Morton, 1969; Smith & Spoehr, 1974; Theios & Muise, 1977) relied on two assumptions. First, the human information processing system involves a series of processing stages that work in a serial, nonoverlapping fashion. Information only flows one way, that is, forward, through the system and, further, each stage is essentially completed before the next begins. The term "thresholded" is used to refer to the assumption that each stage must be completed before the next one can begin. The idea is that a stage is ready to pass information on to the next stage only when the activation at the initial stage reaches a threshold. In contrast, models proposing that information passes between stages as soon as information at one stage begins to be activated are referred to as "cascaded" (McClelland, 1979). The second assumption was that the word recognition system is a fairly autonomous system, that is, it works only with the information stored within it, in particular, the information that can be referred to as lexical information (Forster, 1981). (Theios & Muise's, 1977, model, contained in figure 3.1, is a typical example of this type of model.)

At the risk of overgeneralizing, these models proposed that there is initially a perceptually based process that leads to the activation of sublexical units (typically letter units). The activation of these sublexical units allows the formation of some sort of "prelexical" code. This code activates those word (i.e., lexical) units that are more or less consistent with it. Ultimately, one of these units is selected or accessed. Only at that point does meaning start to become activated. The specific assumption that meaning activation strictly follows lexical selection is referred to as the "form-first" assumption (Forster & Hector, 2002).

One major problem that the early models faced was explaining why there often seemed to be observable effects of "higher-level" information on "lower-level" processing. The classic example is the word superiority effect (Reicher, 1969; Wheeler, 1970). The word superiority effect refers to the fact that letters (i.e., lower-level information) are more accurately reported when presented in words than when presented in nonwords. The experimental task involves the rapid presentation of a letter string often followed by a mask in order to make perception difficult. One letter position is cued for report. To prevent guessing from differentially influencing responding, two alternatives are presented for the identity of the cued letter on each trial. If the letter string had been a word, both alternatives would create a word (e.g., if the word had been WORD and the final position had been

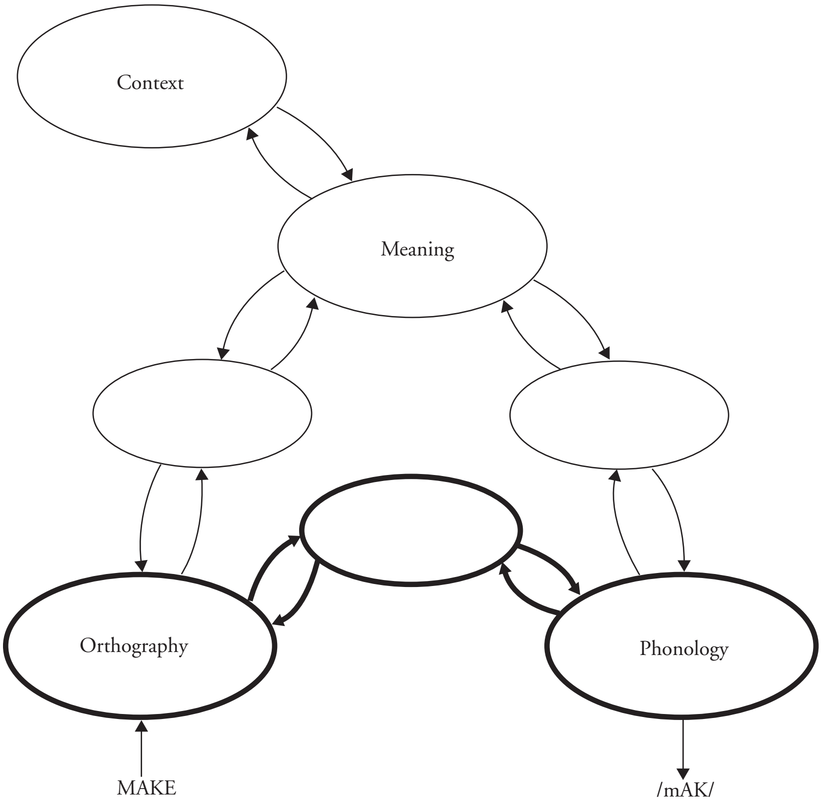

Figure 3.1 Model of word recognition (Theios & Muise, 1977).

cued for report, the alternatives might be D and K). If the letter string had been a nonword, both alternatives would create a nonword (e.g., VCRD with D and K as alternatives for the final position). The standard result is better performance in the word condition than in the nonword condition (e.g., Johnston & McClelland, 1973; Maris, 2002; Paap, Chun, & Vonnahme, 1999; Paap, Newsome, McDonald, & Schvaneveldt, 1982).

The problem for models based on the principles of autonomy and thresholded processing is obvious. How can the existence of a mental representation for a word (e.g., a lexical unit) influence the processing of letter information if that mental representation itself is not accessed until the identity of the letter in question is known? Do changes have to be made to the functional architecture of the models to explain these findings, or is it only necessary to change a single assumption of the model? Alternatively, can these effects be explained in terms of some process (e.g., decision) not actually described by the model itself? It now seems clear that it was the impetus provided by these types of questions that led to the explosion in word recognition research witnessed since the early 1970s. (For a discussion of these issues in auditory word recognition, see Norris, McQueen, & Cutler, 2000, and the invited commentaries.)

The basic phenomena

Although a complete model of word recognition will need to account for an extensive set of phenomena, at present, it is premature to expect any model to do so. Some phenomena will ultimately turn out to be task dependent and, hence, not informative about the nature of the word recognition system per se. Others will only arise in such restricted circumstances that their impact on models of the process will necessarily be limited. What, then, are the basic phenomena that all models should address? Clearly, this list is subjective. The main criteria for inclusion are replicability and the likelihood that the phenomenon reflects the basic architecture of the word recognition system. This second criterion appears to be especially challenging. For three of the four phenomena listed below, there are already arguments in the literature that these phenomena arise outside the word recognition system.

The word superiority effect. Based on its historical import, an obvious phenomenon to include would be the word superiority effect (Reicher, 1969; Wheeler, 1970). It should be noted, however, that some researchers have recently argued that this effect may actually have more to do with phonology than with lexical processing (e.g., Hooper & Paap, 1997; Maris, 2002).

Unlike the word superiority effect, the next three effects all arise in speeded response tasks, that is, tasks in which participants are instructed to respond as rapidly and accurately as possible, and response latency is the main dependent variable. The two standard tasks of this sort are naming, where participants simply have to pronounce a presented word, and lexical decision, where subjects have to decide whether a letter string is a word in the language (e.g., CAT vs. SLINT).

The word frequency effect. The second phenomenon is the word frequency effect (Becker, 1976; Forster & Chambers, 1973; Monsell, 1991; Monsell, Doyle, & Haggard, 1989). Words that are seen more often are responded to more rapidly. Once again, however, this effect is controversial. Balota and Chumbley (1984) have argued that this is a decision phenomenon and, hence, may have little to do with the word recognition system. Further, some researchers (e.g., Morrison & Ellis, 1995) have suggested that observed frequency effects are at least partly due to confounding frequency with age-of-acquisition – that words learned at younger ages are more rapidly processed and, due to the fact that higher-frequency words are typically learned at younger ages, frequency effects may be, to some degree, age-of-acquisition effects.

The semantic priming effect. The third phenomenon is the semantic priming effect (Meyer & Schvaneveldt, 1971; see Neely, 1991, for a review). The experimental task involves the presentation of two words. The first, the "prime," establishes a context. Typically, no response is required to the prime. The second word, the "target," requires either a naming or lexical-decision response. Targets (e.g., DOG) that are related to the semantic context provided by the prime (e.g., CAT) are responded to more rapidly than targets that are not (e.g., NURSE) although there is some controversy as to whether all types of semantic context (e.g., category, antonym) produce priming effects (Lupker, 1984; Shelton & Martin, 1992; Williams, 1996). It should also be noted that there is general agreement that at least some of the observed priming effects are due to processes outside the word recognition system (although see Plaut & Booth, 2000, for an attempt to explain semantic priming solely in terms of lexical processing).

The masked repetition priming effect. The fourth and final phenomenon is the masked repetition priming effect (Evett & Humphreys, 1981; Forster & Davis, 1984). In the masked priming technique, a prime word is briefly presented followed immediately in the same physical position on the computer screen by the target. The presentation of the prime and target in this way means that the target masks the prime such that participants typically report that no stimulus other than the target had been presented. The prime and target are in different cases so that there is very little figural overlap between them (e.g., dog-DOG). Targets are responded to more rapidly if the prime and target are the same word.

There are undoubtedly some phenomena that are noticeable by their absence from the above list. Some are absent because they form the core of the eventual discussion about interactivity (e.g., ambiguity effects, homophone effects). Others are absent because the stability of the effect is still being challenged. Among these are neighborhood effects and form-priming effects. A word's "neighborhood" is defined as all the other words that share letters at all but one letter position (Coltheart, Davelaar, Jonasson, & Besner, 1977). Thus, the word PINE has as neighbors, LINE, PANE, PILE, and PINT, among others. A number of researchers have reported that words with large neighborhoods are processed more rapidly than words with small neighborhoods (Andrews, 1992; 1997; Sears et al., 1995; although see Forster & Shen, 1996, and Grainger, 1990). In addition, a number of researchers have reported that words without higher-frequency neighbors are processed more rapidly than words with higher-frequency neighbors (Grainger, 1990; Grainger & Jacobs, 1996; Grainger, O'Regan, Jacobs, & Segui, 1989, although see Sears, Hino, & Lupker, 1995; Sears, Lupker, & Hino,1999; and Siakaluk, Sears, & Lupker, 2002). Form priming refers to priming that arises when the prime and target have similar forms (e.g., tile–PILE). Although it seems likely that form priming effects do exist, the experimental conditions under which they exist is still unclear in both masked prime (Forster, Davis, Schoknecht, & Carter, 1987; Segui & Grainger, 1990) and unmasked prime (Colombo, 1986; Lupker & Colombo, 1994; Martin & Jensen, 1988) experiments. Other effects are absent because they appear to be restricted to the naming task (e.g., regularity, length). These effects do not appear to represent characteristics of the word recognition system per se and would be better dealt with in a discussion of the processes involved in naming.

Finally, there are a number of effects that are based on responses to nonwords in the lexical decision task, such as the nonword legality effect (Rubenstein, Lewis, & Rubenstein, 1971a; Stanners & Forbach, 1973), the pseudohomophone effect (Coltheart et al., 1977; Dennis, Besner, & Davelaar, 1985), and the nonword neighborhood size effect (Coltheart et al., 1977). While an argument can be made that it is precisely when the word recognition system fails that we can learn most about it, these types of effects seem to have more to say about task specific processes, in this case, in lexical decision, than about the word recognition process per se.

The Models

Search models

The bin model. Search models best represent the way in which one can build a model based on the assumption of thresholded, autonomous processing. According to search models, readers recognize a word by comparing a prelexical code against a set of lexical codes until a match is obtained. The search is not through all of lexical memory but rather, some process designates a section of lexical memory as the optimal search area and the search is confined there. The model that best exemplifies this idea is Forster's bin model (1976; 1989).

According to Forster's (1976) model, the lexical system involves three peripheral access files and a master file, each containing information about all the words in our lexicon. The three peripheral files are orthographically-, phonologically- and semantically-based and each serves as a means of getting to word entries in the master file where all the information about the word is contained. It is relevant to visual word recognition to focus on the orthographic file in which each word in our lexicon contains an entry (this is also true for the other two peripheral files). In each entry in the orthographic file are two things, an "orthographic access code," which is a description of the orthographic properties of the word, and a pointer to the location for that word in the master file.

When a word is viewed, a perceptual process turns that word into a prelexical code that is format compatible with the access codes in the orthographic file. The orthographic file is then searched by comparing the prelexical code with the orthographic access codes. As noted, this search is constrained to a section of the orthographic file. In particular, the orthographic file is organized into bins that contain similar orthographic access codes. So, for example, the words CAT and CAN would probably be in the same bin. In essence, the search is constrained to the bin that is most likely to contain the word being viewed.

The idea of bins may be better understood by drawing a partial parallel to looking up a word in a dictionary. When using a dictionary, one checks the words at the top of each page and only looks at the individual items on the page if it is a page that the word is likely to be on (e.g., the word COMET is virtually certain to be on the page with the heading COMBO–COMFORT). Each bin is like a page in the dictionary and the reader goes directly to the bin most likely to contain the word being viewed. The parallel is not perfect, however, because the words in the bin are not ordered alphabetically, as they are on a dictionary page, but in descending order of frequency. Thus, the entries in the bin are searched in descending order of frequency.

If the search through the designated bin turns up a close match with one of the entries, the location of this entry is flagged while the search continues, looking for other close matches. If a match is close enough, the entry is opened and the pointer to the master file is used to access the word's entry in that file. This process engages a second analysis, referred to as "post-access check," which compares the properties of the stimulus with the properties of the word in the master file. If this comparison is successful, the word has been successfully recognized. Note also that if none of the words in the bin are successfully recognized in the initial search, close matches that had been flagged but not had their entries opened are then evaluated (Forster, 1989; Forster, Mohan, & Hector, 2003).

In terms of the four basic phenomena, the model has no difficulty explaining the frequency effect and the masked repetition priming effect. The more rapid processing of high-frequency words follows directly from the fact that the bins are searched in descending order of frequency. Masked repetition priming arises because the prime begins the word recognition process and, if the target is a repetition of the prime, its processing has a head start. In particular, it is assumed that the prime begins to open the correct entry in the orthographic file. Thus, the entry opening time for the target is shortened, producing more rapid processing. In contrast, the model does not have any obvious way of explaining the word superiority effect.

The other phenomenon, semantic priming, can be explained in terms of cross-referencing in the master file, at least according to the original version of the model (Forster, 1976). Entries for semantically related words are directly linked in the master file. Thus, after the prime DOG has been recognized, the CAT entry in the master file can be easily accessed. As a result, the post-access check of the properties for CAT against the properties of the stimulus can be started without bothering with the search process.

This proposal concerning the (limited) impact of semantics on the word recognition process has a number of implications. One is that, because semantically primed words do not engage the search process, there should be no frequency effect for those words. In fact, Becker (1979) has demonstrated that the frequency effect is smaller when words are semantically primed. A second implication is that semantic priming effects should only exist when the prime's entry is successfully accessed in the master file. Thus, semantic priming effects from primes that are masked in order to avoid recognition (e.g., Carr, McCauley, Sperber, & Parmelee, 1982; Fischler & Goodman, 1978; Hines, Czerwinski, Sawyer, & Dwyer, 1986; Marcel, 1983) are problematic for the model. Finally, because the only impact of semantics on lexical processing is due to the structure of the master file, the model cannot explain any effects of semantics on word recognition with the exception of semantic priming effects. As will be discussed subsequently, there are a number of such effects.

The activation-verification model. Paap et al.'s (1982) activation-verification model (see also, Paap, Johansen, Chun, & Vonnahme, 2000) is also a search model; however, it differs from Forster's (1976) in that, although it is an autonomous model, it invokes cascaded processing. In the model, first letter units and then word units are activated in a serial, but cascaded, fashion (so that information passes through the system before initial processing is complete). Letter activation occurs in position-specific channels and is conceptualized as a feature matching process. Thus, there is some probability that an incorrect but featurally similar letter will be activated at each letter position. Activity at the letter level continuously feeds into the lexicon with the activation of any lexical unit being a function of the activity levels of that word's constituent letters.

It is the activity levels in the lexicon that determine which set of word candidates is selected for further processing. The nature of that further processing is crucially dependent on whether the reader has also been able to establish a “refined perceptual representation of the word” (Paap et al., 1982, p. 574), which is the situation in normal reading. In this case, the set of candidates is serially verified against the perceptual representation (this is the search process). If there is a sufficient match between a candidate and the perceptual representation at any point, the candidate is accepted and the verification process is terminated. As in Forster’s (1976) model, the verification process is frequency-based (higher-frequency words are verified first). Further, if there is a semantic context (i.e., a prime), words semantically related to the prime will enter the candidate set and be verified first. If a refined perceptual representation cannot be established, as in perceptual identification tasks, it is not possible to carry out the verification process. Thus, a probabilistic selection is made from among the candidates based on the activation levels of those candidates.

In terms of the four basic phenomena, this model has had its greatest success explaining the word superiority effect (see Paap et al., 1982). The model can also explain both frequency effects and semantic priming effects. Frequency effects arise due to the serial, frequency-based verification process, whereas semantic priming effects are due to the inclusion of semantically related words in the candidate set. Masked repetition priming effects are more problematic for the model. Presenting a masked prime that is identical to the target will activate target representations at both the letter level and the lexical level. The most important determinant of processing speed, however, is the search process, which only begins once the candidate set has been established. The prime’s activation of the target’s representations may increase the probability that the target will be in the candidate set; however, it should not change its position in the search order. Hence, unless additional assumptions are added, the prediction would be that a masked repetition prime would not have any effect on target processing.

Two additional points should be made about the model. First, in data-limited tasks, tasks in which it is hard to establish a refined perceptual representation, the verification process cannot be carried out. Thus, there is no mechanism for producing a frequency effect. Indeed, there does not appear to be much evidence for frequency effects in word superiority effect experiments (Manelis, 1977; Paap & Johansen, 1994; Paap & Newsome, 1980). Second, as is true of Forster’s model (1976), this model has no means of explaining any semantic effects other than semantic priming effects.

Activation models

The interactive activation model. Activation models represent the other end of the continuum from the search models in terms of cascaded and autonomous processing. The preeminent activation model is McClelland and Rumelhart’s (1981) interactive activation model. This model represents the first real implementation of activation and inhibition processes. It also forms the core of a number of other models in the literature (e.g., Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Grainger and Jacobs, 1996).

The interactive activation model was specifically intended to be a model that would explain the effects of higher-level information on lower-level processing, in particular, the word superiority effect. In the model, there are three levels of representation: feature,

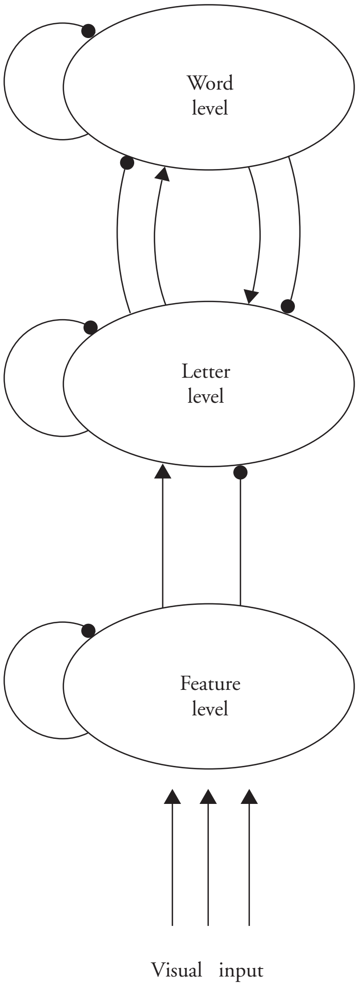

Figure 3.2 Interactive activation model (McClelland & Rumelhart, 1981).

letter, and word. When processing begins, there is a continuous flow of activation upstream from feature-level representations to letter-level representations to word-level representations, as well as downstream from word-level representations back to lower-level representations ("feedback activation"). There is also a flow of inhibition between representations at the same level. Lexical selection is achieved when the activation in a lexical representation exceeds a threshold. (See figure 3.2 for a graphic description of the interactive activation model.).

As McClelland and Rumelhart (1981) argue (see also Rumelhart & McClelland, 1982), this type of system can readily account for the impact of higher-level representations on lower-level representations and, hence, it can explain the word superiority effect. It can also explain frequency effects due to the fact that the resting level activations of word-level representations are frequency dependent. Thus, once activated, representations for high-frequency words will reach their activation threshold more quickly than representations for low-frequency words. Masked repetition priming effects would be explained in terms of the residual activation left in a word's representations as a result of the brief exposure to the masked prime. The model also has the potential to explain semantic priming effects as well as effects due to semantic aspects of the word itself, for example, the fact that imageable words are responded to more rapidly than nonimageable words in lexical decision tasks (e.g., Bleasdale, 1987; de Groot, 1989; James, 1975; Kroll & Merves, 1986). These types of effects would be due to "higher-level input" (i.e., semantic information) impacting word-level representations through feedback activation.

The proposal of a fully interactive system, like that in the interactive activation model, has had its critics (e.g., Massaro, 1988; Paap et al., 2000). However, many modelers have found the idea of interactivity attractive. In the dual-route cascaded model (Coltheart et al., 2001), for example, the basic interactive activation system is used to describe the actions of the first part of the model's "lexical route." The model also has a number of additional structures. In particular, the system containing word-level representations (the "orthographic input lexicon") is directly linked to a phonological output lexicon. The phonological output lexicon contains phonological codes of words known to the reader and it is linked to the phoneme system, which allows translation of the phonological codes into speech. There is also the second route in the model, the "nonlexical route," which connects the letter-level representations directly to the phoneme system. Finally, there is a semantic system, which indirectly connects the orthographic input lexicon and the phonological output lexicon (see Coltheart, this volume, for further discussion). In theory, this system, through feedback operations, would provide a means of explaining both semantic priming effects and effects due to the semantic nature of the word itself.

A second well-known extension of the interactive activation model is Grainger and Jacobs's (1996) multiple read-out model. The main goal of the multiple read-out model was to explain neighborhood effects (and any interactions with word frequency) in both lexical decision and perceptual identification tasks. In order to explain these effects in lexical decision, two new features had to be added: first, a time-based criterion mechanism was added to explain how participants make negative, "nonword" decisions. Second, the assumption was made that positive, "word" decisions could be based on something other than a word-level unit reaching an activation threshold. In particular, positive decisions can be made if there is a high overall level of activation at the word level (i.e., if that activity level exceeded a criterion, it would be taken as sufficient evidence that a word had been presented). The model has had some success at explaining neighborhood effects (although see Sears et al., 1999, and Siakaluk et al., 2002). What the model does not have, however, is any real mechanism for explaining semantic effects. Given the level of precision at which this model is attempting to predict lexical decision latencies, and given the clear impact of semantics on this process (as will be described later), this omission would seem to represent an obvious problem for the model.

Parallel distributed processing models. The models discussed so far contain different assumptions about cascaded processing and the autonomy of lexical processing, but they all agree on one assumption. The core process in word recognition is isolating (i.e., "select-

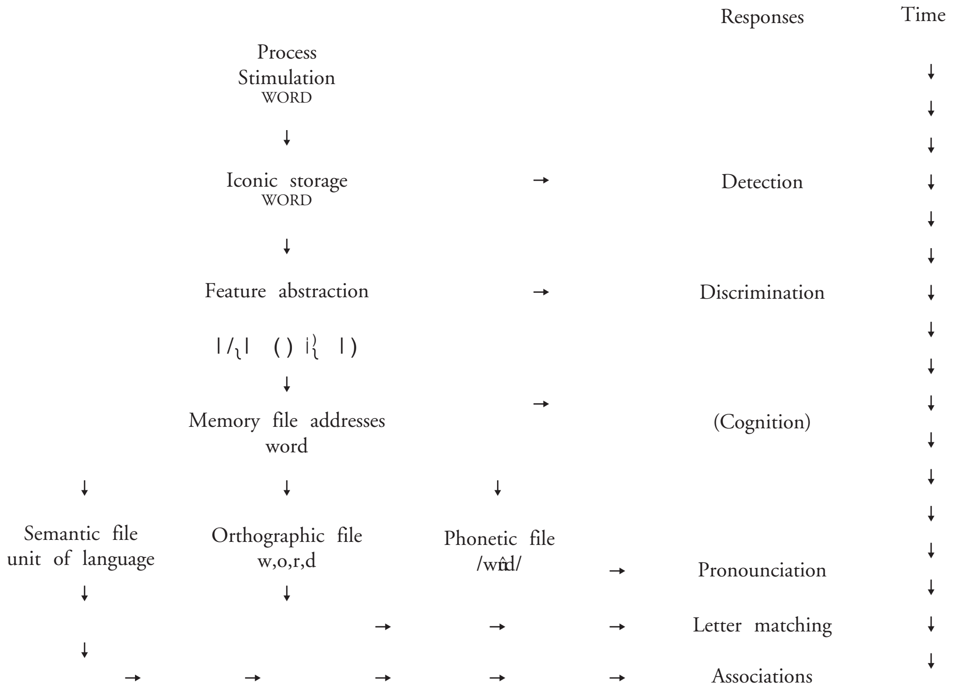

Figure 3.3 Triangle framework (Seidenberg & McClelland, 1989).

ing," "accessing") the relevant lexical unit. However, there are now a number of models in the literature that propose there is no such thing as a lexical unit. Instead, these models are based on the idea that what makes up our lexical system are sets of distributed, subsymbolic codes representing the attributes of the words we know. The word recognition process is the process of activating the appropriate sets of these codes. The models are referred to as parallel distributed processing (PDP) models and they are typically represented with a triangle framework (see figure 3.3 and Plaut, this volume).

The first of these models was proposed by Seidenberg and McClelland (1989). The basic idea is that the word recognition system involves three types of mental representations (orthographic, phonological and semantic representations). Units of each type are assumed to be connected to units of the other types, producing the triangle representation (see figure 3.3). The appropriate connections between sets of units have to be learned, just as a young reader must learn to read. Within the model, learning is essentially an error correction process. When presented with a word, the units at all levels begin to activate (and inhibit) each other, resulting in a pattern of activation across all the units. These activation patterns, which initially will be quite inaccurate, are compared with the correct patterns and then weights between units are adjusted in order to make processing more accurate the next time. This process continues with each new exposure to the word. As a result, over time, activation in one set of units comes to produce the appropriate activation in the units in the other pools (e.g., orthographic processing of the visually presented word CAT allows the activation of the phonological units for the phoneme sequence [kat]). In addition, as shown in figure 3.3, these models also incorporate "hidden units." These units help define the relationships between units in the main pools (for an explanation of why hidden units are necessary see Hinton, McClelland, & Rumelhart, 1986).

The nature and number of representations in each domain (orthographic, phonological, semantic) are often model specific. For example, in the Seidenberg and McClelland (1989) model, the orthographic units are not intended to represent psychologically real concepts; nonetheless, the pattern across units does give rise to sets of letter triples (e.g., MAK). Due to this fact, although the model had considerable success in explaining word recognition data, it also had some serious limitations, particularly in its ability to name nonwords (Besner, Twilley, McCann, & Seergobin, 1990; Fera & Besner, 1992; Coltheart, Curtis, Atkins, & Haller, 1993). In contrast, the orthographic units in Plaut, McClelland, Seidenberg, and Patterson's (1996) model directly represent either letters or letter combinations, allowing some of the problems noted by Besner and colleagues to be fixed. Semantic units, which, until recently (e.g., Harm & Seidenberg, 2004), have played a smaller role in the models' development are typically assumed to represent semantic features (Hinton & Shallice, 1991; Plaut et al., 1996; Plaut & Shallice, 1993), even though the features themselves are often not specified (e.g., Masson, 1991; although see Cree, McRae, & McNorgan, 1999; Harm & Seidenberg, 2004; and McRae, Seidenberg, & de Sa, 1997).

In terms of the four basic phenomena, the model can clearly account for frequency effects. Indeed, the frequency of exposure to a word is the main determinant of the values of the connection weights for the word's units. Further, because the model is an activation model, it should also be able to explain masked repetition priming. That is, the briefly presented masked prime would activate some of the prime's units, allowing for more rapid processing of the target if that target is a repetition of the prime. With respect to semantic priming, modeling work by Masson (1991, 1995), Cree et al. (1999), Plaut (1995b), Plaut and Booth (2000), and McRae et al. (1997) has demonstrated that these types of models can produce semantic priming effects, at least when the concepts share semantic features. In particular, according to these models, the processing of related targets (e.g., DOG following the prime CAT) is easier because the two concepts share semantic units. As a result, some of the semantic units for the target will have already been activated when the target appears, speeding target processing. Notice that the interactivity inherent in these models, specifically, the feedback processes presumed to occur between the semantic units and the "lower-level" orthographic and phonological units, plays essentially no role in this explanation. As will be discussed below, explanations in which these feedback processes do play an important role may provide an even better way of explaining semantic priming effects.

Explaining the word superiority effect is more of a challenge for the model. It would seem like the feedback processes at work in the model should, as with the interactive activation model, produce a word superiority effect. A key distinction here, however, is that it is the feedback from word units in the interactive activation model that produces the word superiority effect. Those units do not exist in PDP models. However, as Plaut et al. (1996) note, units in network models tend to organize themselves into stable patterns called "attractors." These patterns of units that group together to become attractor units may function somewhat similarly to word units, providing the necessary feedback. Thus, it is possible that, with the correct assumptions, the model could also explain the word superiority effect.

The various models discussed above all have their strengths and weaknesses based on their ability to account for the basic phenomena. At present, there is no clear winner. The proposition to be argued in the remainder of this chapter, however, is that other evidence indicates that any successful model of word recognition will need to assume that there is an interactive flow of activation among the processing structures (see also Stone & Van Orden, 1994; Taft & van Graan, 1998; Van Orden & Goldinger, 1994). That is, not only does activation flow forward from activated units to other sets of units, but also, once those units start to become activated, they send activation back to the appropriate units at other levels (see Van Orden & Kloos, this volume). The term "feedback" is used to refer to the flow of activation back to units that were initially activated (i.e., the orthographic units when the word CAT is read). As noted, the interactivity notion is embodied to various degrees by the activation models and is a direct contradiction of the autonomy assumptions that tend to characterize the search models. The framework for this discussion will be the triangle framework, which is most clearly an attribute of the PDP models. In theory, it would be possible to refer instead to a model like Coltheart et al.'s (2001), as it also has the triangle structure embedded within it (i.e., with its orthographic input lexicon, phonological output lexicon, and semantic system). However, the units in the first two of these systems are lexical, while some of the effects to be discussed below (e.g., Stone, Van Hoy, & Van Orden, 1997) are based on sublexical units, making it more difficult to see how those effects fit within this framework.

The Orthographic-Phonological Interaction

Feedback from phonology to orthography

In visual word recognition tasks, the units initially activated are the orthographic units. Thus, evidence for feedback activation would come from experiments demonstrating that phonological activation affects the activation in those orthographic units. For example, Stone et al. (1997) and Ziegler, Montant, and Jacobs (1997) have shown that words that have multiple possible mappings from phonology to orthography (i.e., words like GAIN, which could have been spelled GANE) produce longer lexical decision latencies than words like TENT which have only one possible mapping (see also Perry, 2003). Words like GAIN are referred to as "feedback inconsistent" because the mapping from phonological units to orthographic units are one-to-many. Words having one-to-one mappings, between phonology and orthography, like TENT, are referred to as "feedback consistent." The explanation for these findings is that, with inconsistent words, feedback slows the activation of the correct orthographic code because at least some of that feedback is misdirected to the incorrect orthographic code (e.g., ANE), creating competition.

The effects reported by Stone et al. (1997) and Ziegler et al. (1997) were not large and their reliability has been challenged by Peereman, Content, and Bonin (1998). One could argue, however, that the reason the effects were small was because the manipulations were weak. The feedback directed to the incorrect orthography (i.e., ANE) does not activate a strong competitor for GAIN because neither GANE nor ANE are words and, hence, neither is strongly represented within the orthographic units. The use of homophones allows for a much stronger manipulation. Homophones are words that have different spellings but the same pronunciation (e.g., PAIN and PANE). According to a feedback account, if either is presented visually, the activation of /pAn/ would lead to activation being fed back to the orthographic codes for both PAIN and PANE. The result should be strong competition, leading to a delay in responding. That is, there should be a homophone disadvantage in tasks based on orthographic processing (e.g., lexical decision).

The available data are firmly supportive of this prediction. Rubenstein et al. (1971a) were the first to report a homophone disadvantage in lexical decision in comparison to a control condition. While there was considerable controversy about this finding (e.g., Clark, 1973) and some failures to replicate (e.g., Coltheart et al., 1977), more recently the pattern has emerged very clearly (Davelaar, Coltheart, Besner, & Jonnasson, 1978; Pexman & Lupker, 1999; Pexman, Lupker, & Jared, 2001; Pexman, Lupker, & Reggin, 2002).

Early accounts of homophone effects were based on the idea that visual word recognition was phonologically mediated. Part of the reason was that, originally, these effects were only found when processing the lower-frequency member of the homophone pair (e.g., PANE). The explanation was that both PAIN and PANE activated the phonological code /pAn/ and it then led to the selection of the lexical unit for the higher-frequency member of the homophone pair (i.e., PAIN). Further processing allowed the discrepancy to be noted, at which point the lexical selection process was restarted. The result, of course, was longer latencies for low-frequency homophones. The Pexman et al. (2001) paper is especially important in this regard. Here the nonwords in the lexical decision task were pseudohomophones (nonwords that sound like words when pronounced; e.g., BRANE). These nonwords produced longer word latencies and not only did the homophone effect increase for the low-frequency words, but there was also a significant homophone effect for the high-frequency words. This should never happen if the homophone effect were due to selecting the higher-frequency member of the pair first in lexical search because that event should not be altered by changing the type of nonword being used. In contrast, this result is quite consistent with the claim that these effects are feedback effects.

Feedback from orthography to phonology

The key issue for word recognition is the impact of phonology on orthographic processing. However, for completeness, it is important to discuss the impact of orthography on phonological processing. Research directly relevant to this issue is fairly extensive (e.g., Borowsky, Owen, & Fonos, 1999; Dijkstra, Roelofs, & Fieuws, 1995; Ziegler, Muneaux, & Grainger, 2003); I will focus on two of the earlier papers. A key finding suggesting that orthographic feedback has an impact on speech perception was reported by Seidenberg and Tanenhaus (1979) (see also Donnenwerth-Nolan, Tanenhaus, & Seidenberg, 1981). Seidenberg and Tanenhaus presented participants with a cue word (typically auditorily) followed by auditorily presented target words. The subject's task was to respond as soon as one of the target words rhymed with the prime. Although accurate performance in this task must be based on an evaluation of the phonological code, there was a clear impact of the orthographic relationship between the cue and target. That is, participants were much faster to respond when the two words were spelled the same (e.g., GREED–DEED) than when they were not (e.g., BEAD–DEED).

This effect indicates that orthographic information is automatically activated when a spoken word is heard. It also indicates that orthographic information plays a role in the processing of subsequently presented spoken words. Although it might be the case that participants evaluate the spelling of the words in these experiments in spite of the fact that they are explicitly told to do something else, a more reasonable explanation is that orthographic information was automatically activated when the cue word was processed, and it fed back to the phonological codes for similarly spelled words. Thus, those words were more activated and, hence, easier to process.

A second finding suggesting the impact of orthographic feedback on speech perception was reported by Ziegler and Ferrand (1998). As those authors note, effects like those reported by Seidenberg and Tanenhaus's (1979) derive from the processing of an initial stimulus. Many strategies are available to participants in such a situation, allowing a number of alternative explanations for the findings. Ziegler and Ferrand investigated an on-line effect using an auditory lexical decision task. The key variable was whether the target word had only one or multiple possible spellings. That is, as before, because the word TENT has only one way that it could possibly be spelled, its phonology-orthography mapping is referred to as "consistent" while the word GAIN, which could have been spelled GANE, has an "inconsistent" phonology-orthography mapping. The results showed more rapid latencies for consistent words than for inconsistent words. The explanation offered is that words like GAIN, when presented auditorily, activate incorrect orthographies (e.g., GANE) reducing the support the phonological code /gAn/ receives through feedback activation. Thus, its activation is slowed.

Interactions with Semantics

Facilitative effects of feedback

To provide a satisfactory explanation of any effect, there needs to be at least an implicit assumption about how an experimental task is performed. More specifically, it is necessary to take a position on what units are important in each task. The following discussion will focus on interactions involving semantic and both orthographic and phonological units. We have argued that the interaction between semantics and orthography manifests itself in effects in lexical decision (Hino & Lupker, 1996; Pexman &

Lupker, 1999), whereas the interaction between semantics and phonology manifests itself in effects in naming (e.g., Hino & Lupker, 1996; Pexman, Lupker, & Reggin, 2002). In short, the process of making a lexical decision is driven mainly by activity within the orthographic units, while the naming task is mainly based on activity within the phonological units.

The first effect to be discussed is the ambiguity effect in lexical decision. The standard finding is that words with more than one meaning (e.g., BANK) have shorter latencies than words with a single meaning (e.g., EVENT – Borowsky & Masson, 1996; Hino & Lupker, 1996; Hino, Lupker, & Pexman, 2002; Hino, Lupker, Sears, & Ogawa, 1998; Jastrzembski, 1981; Jastrzembski & Stanners, 1975; Kellas, Ferraro, & Simpson, 1988; Millis & Button, 1989; Pexman & Lupker, 1999; Rubenstein, Garfield, & Millikan, 1970; Rubenstein, Lewis, & Rubenstein, 1971b). This result has a ready explanation in terms of feedback. An ambiguous word has a single set of orthographic units linked to multiple semantic units. When these semantic units are activated, they provide feedback to the correct set of orthographic units, supporting and increasing their activation. As a result, activation rises faster in the units of ambiguous words, producing an ambiguity advantage. A similar expectation holds for the naming task. That is, higher levels of semantic activation provide increased activation from the semantic units to the phonological units. As a result, naming latencies should be shorter for ambiguous words. Again, this result seems to hold (Gottlob, Goldinger, Stone, & Van Orden, 1999; Hino & Lupker, 1996; Hino et al., 2002; Hino et al., 1998; Lichacz, Herdman, LeFevre, & Baird, 1999; Rodd, 2004; although see Borowsky & Masson, 1996).

These effects are based on a many (sets) to one (set) feedback relationship from semantics to either orthography or phonology. However, any variable that affects the semantic richness of concepts (i.e., the amount of activity created at the semantic level) should produce a processing advantage in both lexical decision and naming tasks. For example, words that are highly imageable are assumed to have a richer semantic representation than low-imageable words, and, thus, there should be a processing advantage for such items. (Although a distinction can be made between the concepts of imageabilty and concreteness, the distinction seems somewhat artificial.) In fact, a processing advantage for more highly imageable low-frequency words has been consistently found in lexical decision tasks (e.g., Bleasdale, 1987; de Groot, 1989; James, 1975; Kroll & Merves, 1986). (See Schwanenflugel, Harnishfeger, & Stowe, 1988, for an argument that the active semantic variable here is better described as "context availability" rather than imageability/concreteness.) More recently, similar effects have been found in naming in both English (Cortese, Simpson, & Woolsey, 1997; Strain, Patterson, & Seidenberg, 1995) and Persian (Baluch & Besner, 2001), but again only for the more slowly processed words (cf. Ellis & Monaghan, 2002).

The reason that these effects tend to be restricted to words of low frequency is that low-frequency words are processed more slowly and feedback operations take time. A set of orthographic units must begin to be activated, the activation has to flow forward to the semantic units, a set of semantic units must be activated and, finally, activation has to feed back to the orthographic units (in lexical decision) or forward to the phonological units (in naming) soon enough to actually have an impact. Thus, only words that are more difficult to process would be expected to show an effect.

Pexman, Lupker, and Hino (2002) provided another examination of these ideas by selecting two sets of words that differed in semantic features. This was done using norms for 190 concepts that McRae et al. (1997) derived by asking participants to list the "physical (perceptual) properties . . . functional properties . . . and encyclopaedic facts" (p. 104) for those concepts. The prediction was that words with more features (a richer semantic representation) should produce more activation flowing from semantics to both the orthographic and phonological units. In line with this prediction, Pexman et al. (2002) found that words with more semantic features produced shorter latencies in both lexical decision and naming tasks.

A final point to make is that this type of framework also provides a rather straightforward explanation for semantic priming effects (Meyer & Schvaneveldt, 1971; Neely, 1977, 1991). Feedback activation from the prime's semantic units goes not only to the prime's orthographic units but also to any orthographic units connected to those same semantic units. Thus, the orthographic units for DOG will be partially activated by the prime CAT. As a result, if DOG is presented as a target, activation of its orthographic units will reach the necessary threshold more rapidly. This type of explanation is not very different from that offered by classic "spreading-activation" models that were based on the assumption that words are represented as lexical units (e.g., Collins & Loftus, 1975). In such models, the presentation of CAT as the prime causes activation to spread from its semantic representation to the semantic representation for DOG and then back to the lexical unit for DOG, making DOG easier to process. As noted earlier, however, this type of idea differs noticeably from the explanations offered by Masson (1991, 1995), Cree et al. (1999), Plaut (1995b), Plaut & Booth (2000), and McRae et al. (1997). In these accounts, semantic priming is due to the fact that the prime establishes a position in semantic space by activating a set of semantic units similar to those of the target. Moving from the activation pattern created by the prime to that created by the target is then easier if the target and prime are related (and, hence, share semantic units) than if they are not.

An explanation of semantic priming based on changing activation patterns in semantic space (like those proposed by Masson, 1991, 1995, and others) is only a viable explanation, however, if it is assumed that responses in both lexical decision and naming tasks are based on the results of semantic processing (i.e., the time it takes for the system to settle at the semantic level). This seems to be an unlikely assumption, particularly when one considers the naming task. Rather, a feedback explanation would seem to be a more parsimonious explanation for the effects of semantic priming in both tasks (also see Becker, Moscovitch, Behrmann, & Joordens, 1997, and Cree et al., 1999, for the argument that explanations based on changing position in semantic space are more applicable to semantically based tasks).

Inhibitory effects of feedback

To this point, discussion of the interaction between semantics and orthography or semantics and phonology has been noticeably different from the discussion of the interaction between phonology and orthography. In the one case, the argument has been that a richer semantic representation creates stronger feedback to a single set of orthographic or phonological units, producing more rapid processing. In the other case, the argument has been that a single phonological representation feeds activation to two sets of orthographic units, producing competition and, hence, a processing cost. Both of these predictions are based on the nature of the links between units and the presumed nature of the processing required for the task (e.g., an evaluation of orthographic codes in lexical decision). However, there is nothing special about these particular links. Any linkages between units allow for an analysis and predictions.

The type of relationship between orthography and phonology that produces a homophone effect (i.e., two sets of orthographic units are linked to one set of phonological units) has a parallel in the relationship between orthography and semantics and in the relationship between phonology and semantics. In particular, when considering words that have synonyms, there are two sets of orthographic (phonological) units being mapped into one set of semantic units. Thus, when that set of semantic units is activated, it will feed activation back not only to the correct set of orthographic (phonological) units but also to the set of orthographic (phonological) units appropriate to the synonym, producing a competition. The prediction is that there should be a processing cost in both lexical decision and naming. The results in both Dutch (Pecher, 2001) and Japanese Katakana (Hino et al., 2002) support this prediction.

Two Other Emerging Issues

Representing ambiguous words

As Joordens and Besner (1994) note, models based on distributed representations make a clear prediction about the semantic processing of ambiguous words. They predict that there will be competition at the semantic level between the sets of units for the different meanings, producing a processing cost. Thus, ambiguous words should be more difficult to process than unambiguous words, completely the opposite of what is typically reported in lexical decision and naming experiments (Borowsky & Masson, 1996; Gottlob et al., 1999; Hino & Lupker, 1996; Hino et al., 1998; Hino et al., 2002; Jastrzembski, 1981; Jastrzembski & Stanners, 1975; Kellas et al., 1988; Lichacz et al., 1999; Millis & Button, 1989; Pexman & Lupker, 1999; Rubenstein et al., 1970; Rubenstein et al., 1971b). When considering just these two tasks, a feedback explanation within a PDP framework gets around the problem if one assumes that lexical decision responses are based on activity at the orthographic level and naming responses are based on activity at the phonological level. Thus, semantic level competition has little impact in either task (see also Borowsky & Masson, 1996, and Kawamoto, Farrar, & Kello, 1994, for other ways of addressing this problem). The question still lingers, however, as to whether there is any behavioural evidence for this rather key prediction of PDP models.

There are now three sets of results in the literature supporting this prediction. First, Rayner and colleagues (e.g., Duffy, Morris, & Rayner, 1988; Rayner & Duffy, 1986) reported that in some circumstances, ambiguous words receive longer fixations. Second, Gottlob et al. (1999) and Piercey and Joordens (2000) have reported an ambiguity disadvantage in a relatedness-judgment task in that subjects found it more difficult to decide that a word (e.g., MONEY) was related to an ambiguous word (e.g., BANK) than to an unambiguous word (e.g., DOLLAR). Finally, Hino et al. (2002) have shown that ambiguous words take longer to classify (on negative trials) in a semantic categorization task (i.e., BANK – is it a living thing?).

On closer inspection, however, this evidence appears to be quite weak. As Duffy et al. (1988) note, their results are perfectly compatible with a decision-based explanation. That is, it is possible that all meanings of the ambiguous word are activated simultaneously (and without competition) and that the delay in gaze duration is due to the time taken to select the intended meaning for the sentence. In a similar vein, the effects reported by Gottlob et al. (1999) and Piercey and Joordens (2000) can be explained by assuming that all meanings of the ambiguous words are activated without competition but there is then a competition between response tendencies. When the stimulus is MONEY–BANK, the financial meaning of BANK produces a drive to respond “yes” (the correct response), while the river meaning of BANK produces a drive to respond “no.” Indeed, recent work by Pexman, Hino, and Lupker (2002) has shown that when there is no response competition, there is no effect. That is, when the correct response is “no” (e.g., TREE–BANK vs. TREE–DOLLAR), there is no ambiguity disadvantage. Finally, Hino et al.’s (2002) effect in the semantic categorization task has been shown to be category dependent. That is, it arises when the category is broad (e.g., living things) but not when the category is narrow (e.g., animals, vegetables) (Forster, 1999; Hino, Lupker, & Pexman, 2001). These results also point toward a decision-based, rather than a semantic-processing, explanation.

One issue that might be relevant here is that there are essentially two types of ambiguous words (see Klein & Murphy, 2001, 2002). One type is words that have completely unrelated meanings; for example, BANK. The fact that this word has at least two meanings – ‘a place where you keep your money and the edge of a river’ is an accident of history. These types of words are called *homonyms* and, presumably, the various meanings are represented separately in semantic memory. There are also words that have multiple senses that have evolved from a single meaning; for example, the word BELT. The fact that this word means ‘the strip of material that you have around your waist’, ‘a thin area of land’, ‘a hard blow’, and ‘a drink of an alcoholic beverage’ is not an accident of history. These senses all evolved from the basic meaning of BELT. These types of words are called *polysemous* and the different senses may not be represented separately in semantic memory. If so, there could be different processing implications for the two word types; in particular, only homonyms may produce any real competition during semantic processing.

Klein and Murphy (2001, 2002) examined the question of how different senses of a word are represented in memory using their “sensicality judgement” task. In this task subjects see adjective-noun pairs, like SHREDDED PAPER, and their task is to decide whether this combination of words makes sense. Klein and Murphy reported that participants were much faster and more accurate in responding to the second of two adjacent word pairs when the two pairs tapped the same sense of the polysemous noun (e.g., WRAPPING PAPER followed by SHREDDED PAPER) than when the two pairs tapped different senses of the polysemous noun (e.g., DAILY PAPER followed by SHREDDED PAPER). A similar pattern emerged when they considered homonyms. That is, participants were faster and more accurate in responding to the second word pair when the pairs tapped the same meaning of the noun (e.g., COMMERCIAL BANK followed by SAVINGS BANK) than when the pairs tapped different meanings of the noun (e.g., CREEK BANK followed by SAVINGS BANK). Based on these results, Klein and Murphy suggested that the separate senses of polysemous words appear to be represented separately in semantic memory and, in fact, they are represented essentially the same way that the separate meanings of homonyms are.

In contrast, Azuma and Van Orden (1997) and Rodd, Gaskell, and Marslen-Wilson (2002) have argued that there is an important distinction between homonyms and polysemous words that has processing implications for lexical decision tasks. Rodd et al. selected a set of words in which the number of senses (few vs. many) varied orthogonally with the number of meanings (one vs. more than one). Although words with multiple senses showed shorter latencies than words with few senses (in line with previous research), words with multiple meanings produced slightly longer latencies than words with one meaning, although this effect was quite small and nonsignificant. Further, these results only emerged when the nonwords were pseudohomophones (as did the relevant results in Azuma and Van Orden's, 1997, experiments). If a number of assumptions are made about how participants perform the lexical decision task, Rodd et al.'s inhibitory effect of multiple meanings supports the PDP model prediction about the semantic processing of ambiguous words. However, their findings have not yet proved replicable (Hino & Lupker, 2003; Hino et al., 2001; Pexman et al., 2002) and there appears to be no other example of an ambiguity disadvantage of this sort ever reported in the literature.

Prelexical coding

So far, little has been said about the nature of the prelexical code that is used to access the core components of the word recognition system. Indeed, research on this issue is sparse. In all of the models that exist as simulations, assumptions about this code have had to be made. Even in those situations, however, the assumptions have been driven mainly by modeling convenience. Nonetheless, as Andrews (1997) has argued, these assumptions are important. For example, most models are based on the idea that this code allows at least partial activation of all word units in which the orthography is similar to what is contained in this code. In order to produce legitimate simulations of the word recognition process, it is necessary to specify which set of words is activated in this fashion and to what degree. A second, more concrete, example of why this is important is the fact that it was a change in the assumptions about the orthographic codes that the prelexical code contacts that allowed Plaut et al.'s (1996) model to account for the nonword naming data that Seidenberg and McClelland's (1989) model could not. The nature of the prelexical code is, of course, constrained by the nature of the orthographic code because the only purpose of the former is to activate the latter.

Most of the standard models of word recognition (e.g., Grainger & Jacobs, 1996; McClelland & Rumelhart, 1981; Paap et al., 1982) assume a "channel specific" coding scheme for the prelexical code. That is, each letter in a word is immediately assigned to a channel and then identified within that channel. So, when SALT is being read, there would be some activation of word units for HALT, MALT, WALT, SILT, and SALE, those words overlapping in three of the four channels, but much less, if any, activation of SENT, SLAT and SAT. However, extant evidence, suggests that this assumption is incorrect. Humphreys, Evett, and Quinlan (1990), for example, have shown that shorter letter strings can prime longer words (e.g., oitk–WHITE) in a masked prime, perceptual identification task (see also Perea & Carreiras, 1998; and de Moor & Brysbaert, 2000). Such effects have forced researchers, more recently, to adopt "relative-position" coding schemes (e.g., Coltheart et al., 2001); however, the problem does not appear to be fully solved even with those schemes. For example, data suggest that letter strings containing transposed letters (i.e., SALT–SLAT) are actually more similar to one another than letter strings matching in N-1 positions (e.g., SALT–HALT). For example, transposed letter nonwords (e.g., JUGDE) are harder to reject than one-letter different nonwords (e.g., JUDPE) or control nonwords (e.g., SLINT) in lexical decision tasks (Andrews, 1996; Chambers, 1979; Holmes & Ng, 1993). In addition, Forster et al. (1987) showed that masked priming effects for transposed letter primes (e.g., anwser–ANSWER) were as large as those for repetition primes (e.g., answer–ANSWER) and larger than those for one-letter different primes (e.g., antwer–ANSWER). Finally, Perea and Lupker (2003) have reported significant facilitation in a masked semantic priming experiment with transposed letter primes (e.g., jugde–COURT) and little, if any, priming with one letter different primes (e.g., judpe–COURT) (although see Bourassa & Besner, 1998, for a demonstration that these latter effects can become significant if the experiment has enough power). Taken together, these results suggest that, for example, the letter string JUGDE has more potential to activate the lexical structures for the word JUDGE than the nonword JUDPE does, a conclusion that is quite inconsistent with the assumptions of virtually all of the models of word recognition discussed in this chapter. Future empirical work should be directed at a better understanding of the nature of these prelexical codes (see e.g. Davis, 1999, Ratcliff, 1981, and Whitney, 2001, for some possibilities) and that knowledge should then be used to modify current models.

Parting Thoughts

Since the early 1970s, tremendous strides have been made in terms of understanding the visual word recognition process. A major trend that has emerged during this time period has been a movement toward word recognition models that assume considerable interactivity among the various types of lexical and semantic structures. This is not to suggest that the more autonomous conceptualizations of word recognition, such as those described in the search models, can never make a comeback. Nor is it to deny that certain local components of the word recognition system may work on more autonomous principles (e.g., the establishment of prelexical codes; see Norris et al., 2000). The picture is far from complete and future work is likely to capitalize on insights not only from experimental cognitive psychology but also from the neuroscientific study of reading development and reading disorders.[1]

Note

1. Much of the author's research discussed in this chapter was supported by Natural Sciences and Engineering Research Council of Canada Grant A6333. I would like to thank Penny Pexman, Yasushi Hino, Manuel Perea, and Sachiko Kinoshita for their splendid contributions to that research and for their comments on earlier drafts of the chapter.