CHAPTER

3

Visual Word Recognition

Melvin J. Yap *and* David A. Balota

Abstract

Visual word recognition is an integral aspect of reading. Although readers are able to recognize visually presented words with apparent ease, the processes that map orthography onto phonology and semantics are far from straightforward. The present chapter discusses the cognitive processes that skilled readers use in order to recognize and pronounce individual words. After a historical overview of the broad theoretical developments in this rich field, the chapter provides a description of methods and a selective review of the empirical literature, with an emphasis on how the recognition of an isolated word is modulated by its lexical- and semantic-level properties and by its context. The final section of the chapter briefly considers some recent approaches and analytic tools in visual word recognition research, including megastudy, analysis of response time distributions, and the important role of individual differences.

Key Words: visual word recognition, lexical decision, speeded pronunciation, masked priming, semantic priming, orthographic priming, phonological priming, megastudies, individual differences, response time distributional analysis

Skilled reading is a remarkably complex and multifaceted behavior, which relies on the recognition of individual words. The squiggly marks on the page need to somehow map onto a word representation so that the meaning of the word can be accessed. At first blush, this appears to be a relatively straightforward process of pattern recognition. However, words code and convey multiple domains of information, including orthography, phonology, morphology, and ultimately meaning. Indeed, because of the multidimensional nature of word recognition, this literature has made seminal contributions to (1) the distinctions between automatic and attentional mechanisms (e.g., Neely, 1977), (2) the development of computational models (e.g., McClelland & Rumelhart, 1981), and (3) cognitive neuroscience (e.g., Petersen, Fox, Posner, Mintun, & Raichle, 1989). Given the extensive influence of word recognition research on cognitive science, attempting to provide a concise overview of this area is a daunting task. We have chosen to first provide a brief historical overview of the area, with an emphasis on the wide-ranging theoretical contributions. We then turn to some basic findings in the literature and conclude with more recent developments in studying word recognition. Our goal is to expose the reader to the major issues, as opposed to providing detailed expositions of each of the research topics.

Historical and Theoretical Overview

Although a number of writing systems exist, reading research has been dominated by the study of alphabetic writing systems, where the unit of language symbolized by writing is the phoneme (Treiman & Kessler, 2007). In alphabetic writing systems, the building blocks of words are letters, and so the recognition of letters was central to early models of visual word processing. If printed words are recognized via their constituent letters, then it is natural to wonder whether letters are also recognized via their constituent features (see Grainger, Rey, & Dufau, 2008, for a review). An important approach in this area is the feature analytic approach. According to this view, there is a set of visual features (e.g., vertical lines, horizontal lines, diagonal lines, curved closed forms, closed open forms, intersections) that are critical for discriminating among the letters. So, the letter ‹H› would be defined by the convergence of two vertical lines and one horizontal line. Indeed, component features such as these laid the foundation for the first computational model of letter perception (pandemonium model; Selfridge & Neisser, 1960). About the same time, Hubel and Wiesel (1962) were able to identify receptive fields of cortical neurons in alert cats; these receptive fields appeared to be sensitive to vertical lines, horizontal lines, oblique lines, and intersections. Although it is likely that such features play an important initial role in letter perception, many questions remain. These include (1) how the features are bound together to form a letter (see Treisman, 1999, for a review of the binding problem); (2) how the system flexibly codes different sets of features that are necessary for recognizing letters across fonts, visual angles, and levels of degradation; and (3) how the system adjusts to handwritten text wherein the features appear to be very different from standard text (see Plamondon & Srihari, 2000, for a detailed review).

Moving on to the letter level, letters vary in the extent of feature overlap, and, as expected, this influences the ease of searching for a letter in a background of letters (e.g., it is more difficult to locate ‹Z› when it is embedded within the letters ‹F›, ‹N›, ‹K›, and ‹X›, than when it is embedded within ‹O›, ‹J›, ‹U›, ‹D›; see Neisser, 1967). Appelman and Mayzner (1981), in a comprehensive review of isolated letter recognition, considered studies that measured (1) participants' accuracy for identifying single letters under varying levels of degradation or (2) their response times for letter naming, letter matching, and letter classification (i.e., letter vs. nonletter forms). The results, based on over 800,000 observations from 58 studies, revealed that the frequency of a letter in the language (e.g., ‹T› is approximately three times more frequent than ‹C›) had no effect on accuracy-based studies where participants simply report letters. Interestingly, however, there was a clear effect of frequency on response latencies. Appelman and Mayzner (1981) suggested that the consistent absence of letter frequency effects in accuracy was incompatible with the idea that early letter encoding is modulated by letter frequency. We primarily note this pattern because it is surprising that the simple effect of frequency of exposure would produce varying influences across tasks, and hence it is important to remind the reader that there are always important between-task differences when considering the influence of a variable on performance.

Recognizing Letters Within Words

Letters are rarely presented in isolation, but are typically embedded in words. Interestingly, Cattell (1886) argued that letters (e.g., ‹n›) were more easily reported when presented in the context of letters that form words (*born*) than in the context of letters that form nonwords (*gorn*). There are many interpretations of this simple effect. For example, partial information from words (*bor_*) might be more useful for helping participants guess the identity of the critical letter ‹n›. This led to the development of an experimental paradigm that involved a forced-choice test for letters embedded in words, nonwords, and in isolation (Reicher, 1969; Wheeler, 1970). By providing the participant with two plausible response alternatives (e.g., *bore* vs. *born*), guessing is ruled out as an explanation, along with other interpretations of Cattell's original observation. Remarkably, the superior reporting of letters embedded in words, compared with when they were embedded in nonwords or presented in isolation, was upheld. This became known as the *word superiority effect* or the *Reicher-Wheeler effect*.

The theoretical significance of the word superiority effect is profound because one is confronted with the following conundrum: If letters are a necessary first step for recognizing a word, how can word-level information influence the perception of the letters making up the word? This effect stimulated the highly influential interactive activation model of letter perception developed by McClelland and Rumelhart (1981) and Rumelhart and McClelland (1982) (see Figure 3.1). This powerful computational model involves three levels (features, letters, and words) and two types of connections across representations—facilitatory (represented by arrows) and inhibitory (represented by filled circles). Presenting a word activates the feature-, letter-, and word-level representations consistent with that word. Importantly, as word-level nodes receive activation, they begin to provide feedback to position-specific letters. This additional top-down influence of word-level on letter-level representations drives the word superiority effect.

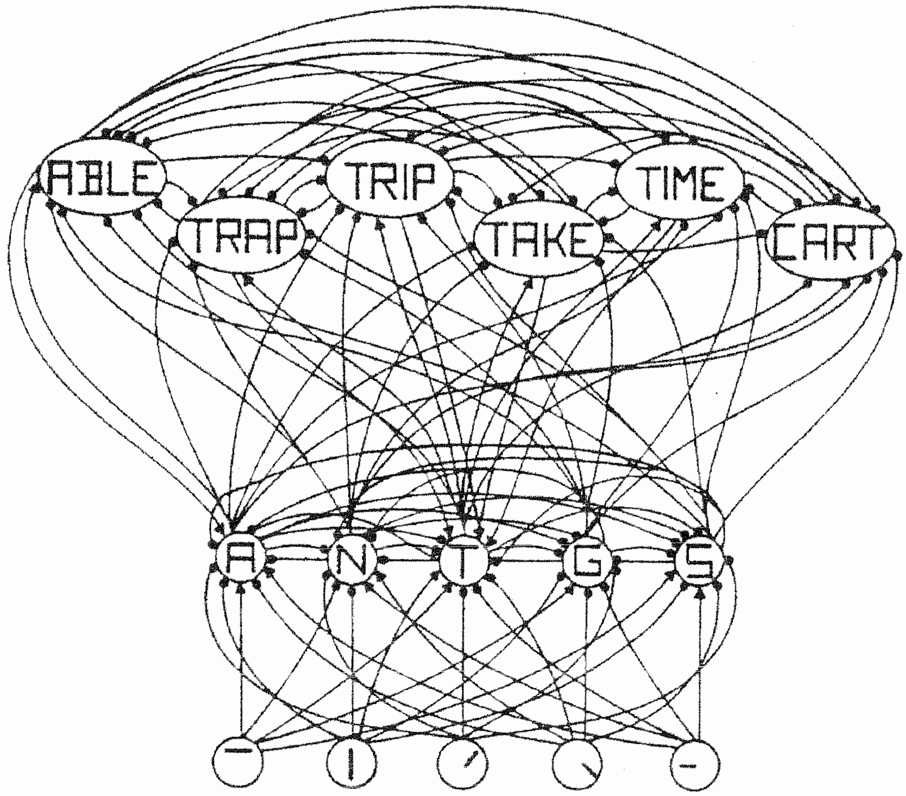

Fig. 3.1 McClelland and Rumelhart's (1981) interactive activation model of letter recognition.

The interactive activation model is historically important for many reasons. First, the model emphasized cascaded, rather than staged, processing (see McClelland, 1979), wherein all nodes accumulate activation across time via the spread of activation and inhibition across the connection paths. Second, the activation dynamics of all units are constrained by the activation and inhibition of other similarly spelled words (i.e., neighbors). This is an important difference from the classic Logogen model developed by Morton (1970), wherein lexical representations (logogens) accumulate activation across time independently of each other. Third, the interactive activation framework is a critical component of a number of computational models of visual word recognition, and predates the principles of the parallel distributed processing (PDP) approaches described in the next section.

Models and Tasks of Lexical Processing

Although the interactive activation model (McClelland & Rumelhart, 1981) contains word-level representations, it was primarily developed to explain letter-rather than word-recognition performance. However, forced-choice letter recognition is rather removed from word-level processing, and one should consider tasks that reflect processes at the word level. Many tasks have been developed to investigate lexical-level processing, including *category verification* and *semantic classification* (e.g., classifying a word as living or nonliving), *perceptual identification* (identifying a perceptually degraded stimulus), and *reading* (with eye-fixation durations on a target word measured). Although all of these tasks have important advantages and some disadvantages, here we focus on two tasks that have been dominant in work on isolated word recognition, *speeded pronunciation* (reading a word or nonword, e.g., *flirp*, aloud) and *lexical decision* (classifying letter strings as words and nonwords via a button press). In these two tasks, researchers respectively measure the amount of time needed by participants to initiate the pronunciation of a word or to press a button. Both tasks a priori appear to map onto processes involved in a word-level representation, reaching threshold to produce the appropriate response, either the correct pronunciation or the correct word/nonword response.

MODELS OF SPEEDED PRONUNCIATION

We will first consider computational models of word-pronunciation performance, since this task has been particularly influential in model development. Our focus is on models of English pronunciation, although it should be noted that models have been implemented in other languages (e.g., French; Ans, Carbonnel, & Valdois, 1998). Historically, there have been two major classes of models of speeded pronunciation: dual-route models and single-route models. The dual-route cascaded (DRC) model (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001) has two distinct pathways for pronouncing a word aloud: a direct *lexical* route that maps the full visual letter string onto a lexical representation and an assembled *sublexical* route that maps the letter string onto its pronunciation based on abstract grapheme-phoneme correspondence rules (see Figure 3.2). These rules (e.g., ‹k› → /k/) were selected on purely statistical grounds; that is, /k/ is the phoneme most commonly associated with ‹k› in English monosyllables. The DRC model accounts for many findings in the visual word recognition literature. One particularly important finding is the frequency by regularity interaction. That is, *regular* words that adhere to abstract grapheme-phoneme correspondence rules (e.g., ‹k› → /k/) are pronounced faster than *irregular* words (those that violate the rules, e.g., *pint*), and this effect is exaggerated for words that are rarely encountered in printed language. This result follows the assumption that the lexical route (based on whole-word representations) is frequency modulated, but the assembled route (based on smaller sublexical units) is insensitive to whole-word frequency. Hence, irregular low-frequency words (e.g., *pint*) are recognized more slowly than regular low-frequency words (e.g., *hint*), because the two routes produce conflicting pronunciations for *pint*, and extra time is needed to resolve the competition before the correct pronunciation can be produced. In contrast, for high-frequency words, the difference in recognition times for regular (e.g., *save*) and irregular (e.g., *have*) words is attenuated or absent, because the lexical route produces an output before there is competition from the slower sublexical route.

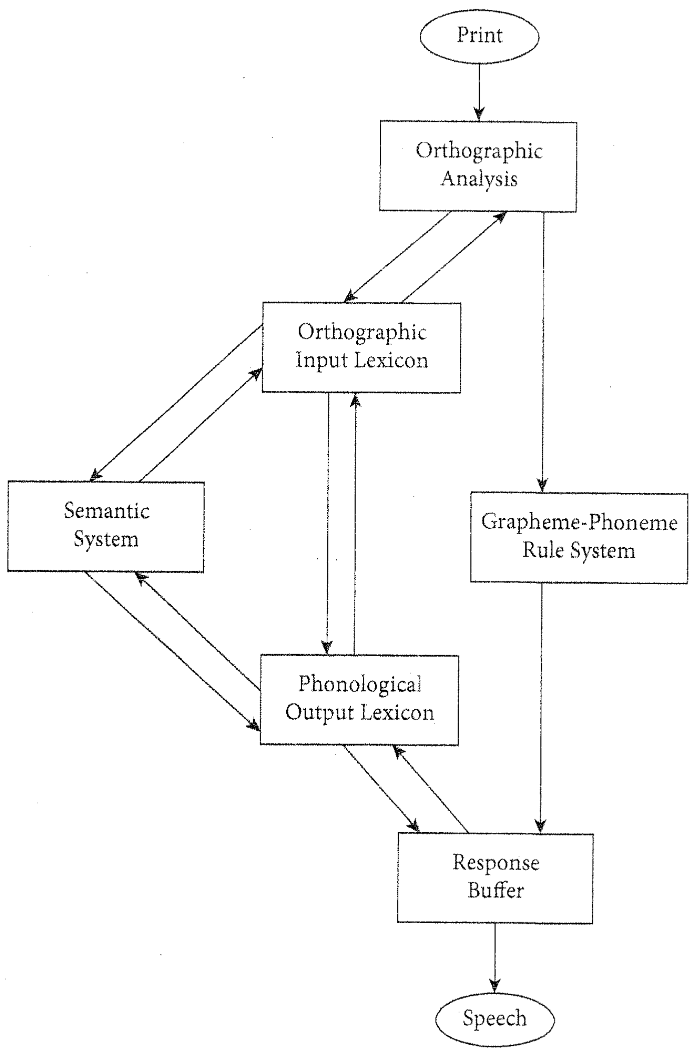

Fig. 3.2 Coltheart et al.'s (2001) DRC model of visual word recognition and reading aloud.

Coltheart et al. (2001) noted that a dual-route model also easily accommodates an important neuropsychological dissociation between acquired surface and phonological dyslexia. Individuals with surface dyslexia appear to have a breakdown in the lexical route, since they are relatively good at pronouncing nonwords and regularize words that do not conform to English spelling-to-sound rules (i.e., they pronounce *pint* such that it rhymes with *hint*). In contrast, individuals with phonological dyslexia appear to have a breakdown in the sublexical route such that they have particular difficulty with nonwords but are relatively good at pronouncing both regular and irregular words, which have lexical representations.

The second major class of models of speeded pronunciation is nicely reflected in the parallel distributed connectionist model developed by Seidenberg and McClelland (1989). The general structure of this model is displayed in Figure 3.3, in which a set of input units codes the orthography of the stimulus and these units map onto a set of hidden units, which in turn map onto a set of phonological units that code the pronunciation of the stimulus. Initially, the pathway weights are set to random levels. Gradually, through the learning mechanism of *backpropagation* (a common method for training computational neural networks), the connections across levels are adjusted to capture the correct pronunciation when a given orthographic string is presented. This model was trained on over 2,400 single-syllable words; the number of times a word

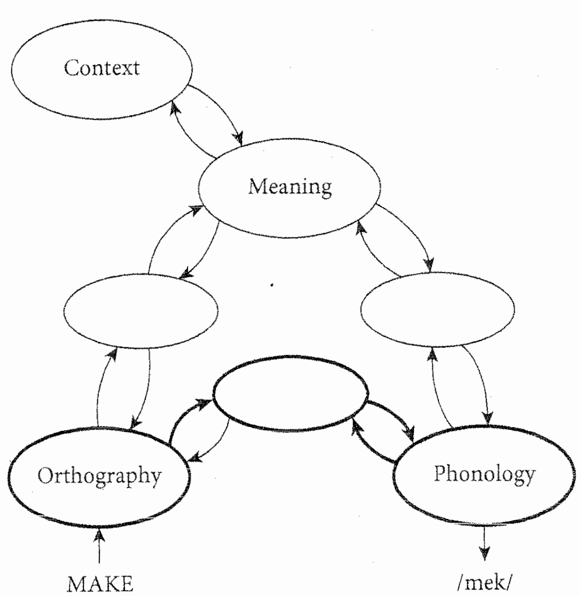

Fig. 3.3 Seidenberg and McClelland's (1989) parallel distributed processing model.

is presented to the model is related to its frequency of occurrence in the language. Remarkably, after training, Seidenberg and McClelland found that the network produced many of the effects observed in speeded pronunciation performance. A particular noteworthy finding is that this connectionist network was able to account for the frequency by regularity interaction noted above. Importantly, the connectionist perspective is appealing because (1) it includes a learning mechanism; (2) it does not contain any formal spelling-to-sound "rules," but instead mimics rule-like behavior based on the statistical properties of spelling-to-sound mappings (see discussion of *consistency effects* later); and (3) it involves one, as opposed to two, pathways for pronunciation.

A hybrid model of speeded pronunciation called developed by Perry, Ziegler, and Zorzi (2007) was the CDP+ (connectionist dual process) model. The CDP+ model is very much like Coltheart et al.'s (2001) model, except that the DRC model's rule-based sublexical route is replaced by a two-layer connectionist network that learns the most reliable spelling-sound relationships in the language. This model is important because it not only accommodates the major empirical benchmarks in the literature but also accounts for considerably more item-level word recognition variance in large-scale databases (see discussion of *megastudies* later). A disyllabic version of this model, the CDP++ model, is also available (Perry, Ziegler, & Zorzi, 2010). The extension to disyllabic words is important because most major word recognition models have focused on single-syllable words (for an exception, see Ans et al., 1998). However, the majority of English words are multisyllabic, which involve additional processing demands such as syllabification and stress assignment. In this light, the CDP++ model is an important advance that extrapolates dual-route and connectionist principles to a much larger set of words.

MODELS OF LEXICAL DECISION PERFORMANCE

The modeling of lexical decision performance has taken a somewhat different path than the modeling of speeded word pronunciation. This is not surprising, since the demands of producing the correct pronunciation for a visual letter string are quite different from the demands of discriminating familiar words from unfamiliar nonwords. For example, within the DRC model, a deadline mechanism has been implemented to simulate lexical decision (Coltheart et al., 2001). That is, a word response is produced when lexical activity in the orthographic lexicon exceeds some threshold, while a nonword response is made if lexical activity does not exceed that threshold after some deadline has elapsed (see also Grainger & Jacobs, 1996). The connectionist network can also be embellished to distinguish between words and nonwords by monitoring a measure of familiarity based on semantic activity (Plaut, 1997). Both approaches are useful for making contact with the lexical processing literature.

In contrast to these models, there are more general approaches that focus on the binary decision processes involved in the lexical decision task. One early model in this area was proposed by Balota and Chumbley (1984; also see Balota & Spieler, 1999). According to this model, lexical decisions can be based on two processes: a relatively fast-acting familiarity-based process and a slower, more attention-demanding process that checks the specific spelling or meaning of a given stimulus. This model was useful for emphasizing the decision-related processes in this task, further underscoring the distinction between task-general and task-specific processes in lexical decision. More recently, computational models of lexical decision have been developed that also emphasize the decision process. For example, Ratcliff, Gomez, and McKoon's (2004) diffusion model assumes that decisions are produced by a process that accumulates noisy information over time from a starting point toward a word or nonword boundary. This model is noteworthy because it captures not only mean response time and accuracy but also response time distributions for both correct and incorrect responses. Hence, this model captures the full range of behavior within the lexical decision task, a problem for previous models. An alternative approach is the Bayesian Reader model developed by Norris (2006). This model assumes that readers in the lexical decision task behave like optimal decision-makers who compute the probability that the presented letter string is a word rather than a nonword, given the input (see Kinoshita, this volume, for further discussion).

It should be evident from the foregoing discussion that models of lexical decision performance are quite different from their speeded-pronunciation counterparts. The latter emphasize processes mediating spelling-to-sound translation, whereas the former emphasize processes mediating word/nonword discrimination. Indeed, the effect sizes of major variables differ remarkably across lexical decision and speeded pronunciation (e.g., Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004). Hence, a flexible and adaptive lexical-processing system is more consistent with the extant literature than one that is relatively static and modular. One such framework is presented in Figure 3.4, wherein one can see how task demands may emphasize different pathways within a more general lexical architecture (Balota & Yap, 2006). Of course, this is simply a general perspective, but the potentially crucial point

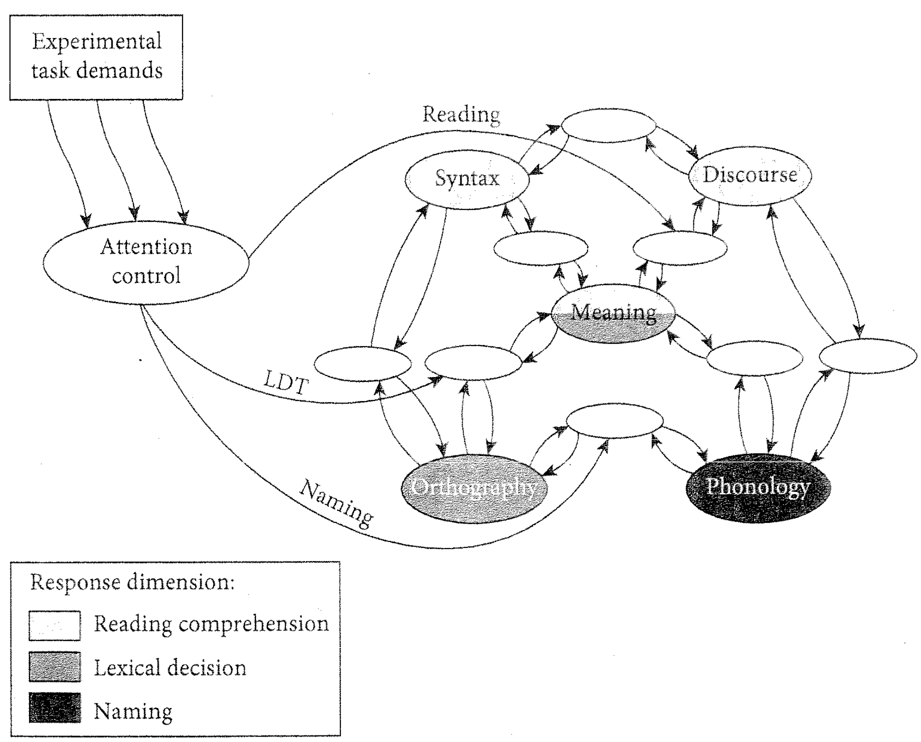

Fig. 3.4 The flexible lexical processor.

is that the lexical processing system adaptively considers different sources of information to maximize performance in response to the demands of a task.

In sum, the visual word recognition domain has provided a powerful test bed for the development of both metaphorical and computational models of mapping visual patterns onto phonology and meaning. This section provides only a snippet of some of the historical developments. Armed with these theoretical perspectives, we now turn to an analysis of how aspects of the empirical literature are interpreted within these models.

Lexical- and Semantic-Level Influences on Word Recognition

In order to better understand the processes underlying visual word recognition, researchers have identified how the many statistical properties associated with words (e.g., frequency of occurrence, number of letters, imageability) influence performance on different word recognition tasks. In this next section, we selectively review the impact of the most important *lexical* variables, which are quantified at the level of the whole word. There is also a rich literature examining the functional *sublexical* units (i.e., representations smaller than a word, such as letters, morphemes, and syllables) mediating word recognition (Carreiras & Grainger, 2004), but this is beyond the scope of the present chapter and is covered in other chapters (see Taft, this volume, and Perea, this volume).

Word Frequency

The frequency with which a word appears in print is the most robust predictor of word recognition performance (Whaley, 1978). Across virtually all lexical processing tasks, participants respond more quickly and accurately to high-frequency than low-frequency words. The word-frequency effect yields important insights into the nature of the human information-retrieval mechanism (Murray & Forster, 2004) and represents a fundamental constraint for all word recognition models. Despite its apparent simplicity, the theoretical interpretation of the word-frequency effect is far from straightforward (see also Kinoshita, this volume).

For example, one general class of lexical access models involves a type of serial *search* or *verification* process (Becker, 1980; Forster, 1976; Paap, McDonald, Schvaneveldt, & Noel, 1987), in which candidates compatible with the initial analysis of the stimulus are compared (or verified) against the visually presented letter string in descending order of frequency. The influential interactive activation model (e.g., Coltheart et al., 2001; McClelland & Rumelhart, 1981; Perry et al., 2007) described earlier assumes that the resting-level activations or activation thresholds of words (logogens in Morton's, 1970, nomenclature) vary with frequency of exposure. High-frequency words are responded to faster because they have higher resting-activation levels (or lower thresholds), thereby requiring less stimulus information to be recognized. Of course, within the connectionist frameworks (e.g., Plaut, McClelland, Seidenberg, & Patterson, 1996; Seidenberg & McClelland, 1989) that rely on distributed, rather than local, representations, frequency is coded by the strength of the weights between input and output representations. The Bayesian Reader model (Norris, 2006), which is predicated on the assumption that people recognize words in an optimal manner, takes a more functional approach to word-frequency effects. Specifically, word-frequency effects are a consequence of ideal observers taking the prior probabilities of words (indexed by their word frequencies) into account when resolving an ambiguous input as the stimulus unfolds during perception.

Researchers have also recently examined how different theoretical frameworks are able to account for the form of the relationship between word-frequency and word recognition measures. For example, a frequency-ordered serial search model predicts a linear relationship between the rank position of a word in a frequency-ordered list and access times, whereas the Bayesian Reader model predicts a logarithmic relationship between frequency and response times (Adelman & Brown, 2008). The work by Murray and Forster (2004) indicated that rank frequency was a better predictor of response times than log-transformed frequency, although this is qualified by more recent analyses by Adelman and Brown (2008) which suggest that word-frequency effects are most consistent with instance models (e.g., Logan, 1988) where each encounter with a word leaves an instance or trace in memory. The functional form of the word-frequency effect has been particularly well studied because researchers have developed large databases of lexical-decision and speeded-pronunciation performance while concurrently generating much better estimates of word frequency within the language (e.g., Brysbaert & New, 2009).

Although printed word frequency plays a central role in lexical access, there is also ample evidence that word-frequency effects partly implicate task-specific processes occurring after lexical access.

For example, in lexical decision, participants may particularly attend to the familiarity and meaningfulness of the letter string to help them discriminate between words and nonwords. This emphasis on familiarity-based information consequently exaggerates the frequency effect in lexical decision, compared with pronunciation (Balota & Chumbley, 1984). Specifically, low-frequency words are more similar to nonwords on the dimension of familiarity/meaningfulness than are high-frequency words. It is therefore more difficult to discriminate low-frequency words from nonwords, thereby slowing response times to low-frequency words and making the frequency effect larger. Indeed, researchers who have manipulated the overlap between words and nonwords by varying nonword wordlikeness (e.g., *brnta, brant, brane*; see Stone & Van Orden, 1993) report that such manipulations modulate the size of the word-frequency effect. The important point here is that frequency effects (and probably most other psycholinguistic effects) do not unequivocally reflect word recognition processes.

Length

Length here refers to the number of letters in a word. In perceptual identification, lexical decision, pronunciation, and reading, one generally observes longer latencies for longer words (see New, Ferrand, Pallier, & Brysbaert, 2006, for a review). Although the length effect is partly attributable to processes (e.g., early visual or late articulatory) that are beyond the scope of word recognition models, simulations indicate that the inhibitory influence of length on pronunciation onset latencies is especially difficult to reconcile with models that fully rely on parallel processing (e.g., Plaut et al., 1996). Instead, length effects are more compatible with models that incorporate serial processing, such as the DRC model (Coltheart et al., 2001), which contains a sublexical pathway that assembles phonology in a serial, letter-by-letter manner (Rastle & Coltheart, 2006). In fact, Weekes (1997) found that length effects are particularly large for nonwords compared with words, consistent with the DRC model perspective that length effects primarily reflect the influence of the sublexical pathway.

Orthographic and Phonological Similarity

In their classic study, Coltheart, Davelaar, Jonasson, and Besner (1977) explored the effects of an orthographic similarity metric they termed *orthographic neighborhood size* on lexical decision. Orthographic neighborhood size is defined by the number of orthographic neighbors associated with a letter string, where an orthographic neighbor is any word that can be obtained by substituting a single letter of a target word (e.g., *sand*'s neighbors include *band, send, said,* and *sank*). Assuming that lexical retrieval involves a competitive process, one might expect words with many neighbors to elicit more competition and hence produce slower response latencies. However, a review by Andrews (1997) suggested that across a number of languages, both lexical decision and pronunciation latencies are generally faster for words with many neighbors, and this effect is larger for low-frequency than for high-frequency words. The facilitatory effects of neighborhood size appear to be difficult to accommodate within any model (e.g., DRC model) that includes an interactive activation mechanism (McClelland & Rumelhart, 1981), because there should be more within-level inhibition to words with more orthographic neighbors. In addition to number of neighbors, researchers (e.g., Sears, Hino, & Lupker, 1995) have also considered the influence of *neighborhood frequency* (i.e., whether the target word possesses a higher-frequency neighbor, see Perea, this volume, for a discussion of such effects).

Like orthographic similarity, phonological similarity is defined by counting the number of *phonological neighbors*, that is, words created by changing a single phoneme of a target word (e.g., *gate*'s neighbors include *hate, get,* and *bait*). Yates (2005) and Yates, Friend, and Ploetz (2008a) have shown that in lexical decision, speeded pronunciation, semantic classification, and reading, words with many phonological neighbors are responded to faster than words with few phonological neighbors. There is also evidence that as the number of phonological neighbors overlapping with the *least supported phoneme* (i.e., the phoneme position within a word with which the fewest phonological neighbors coincide) increases, pronunciation latencies become faster (Yates, Friend, & Ploetz, 2008b). Generally, these results are consistent with the idea that words with many phonological neighbors receive additional activation within the phonological system, and help provide useful constraints for how phonology plays a role in word recognition.

The original definition of neighborhood size is somewhat restrictive. For example, a neighbor had to be matched in length to the target and differing only by the substitution of a single letter or phoneme. More expansive and flexible metrics of neighborhood size have been proposed (see Perea, this volume), including one based on the mean

Levenshtein distance (i.e., the number of single letter insertions, deletions, and substitutions needed to convert one string of elements to another) between a target word and its closest 20 neighbors in the lexicon. This measure (OLD20) has been shown to be a particularly powerful predictor for longer words (Yarkoni, Balota, & Yap, 2008).

Regularity and Consistency

As described earlier, the *regularity* of a word is defined by whether it conforms to the most statistically reliable spelling-to-sound correspondence rules in the language. *Hint* is regular because it follows these rules, whereas *pint* is irregular because it does not. Another theoretically important variable that quantifies the relationship between spelling and sound is *consistency*, which reflects the extent to which a word is pronounced like similarly spelled words. For example, *kind* is considered consistent because most similarly spelled words (e.g., *bind, find, hind, mind*) are pronounced the same way. In contrast, *have* is inconsistent because its pronunciation is different from most similarly spelled words (e.g., *cave, gave, save*). Generally, consistent words are recognized faster than inconsistent words, and the consistency effect is stronger in speeded pronunciation than in lexical decision, because the pronunciation task emphasizes the generation of the correct phonology (Jared, 2002). Such graded consistency effects fall naturally out of the connectionist perspective, where there is no sharp dichotomy between items that obey the "rules" and items that do not. Instead, lexical processing reflects the statistical properties of spelling-sound mappings at multiple grain sizes (Plaut et al., 1996). Consistency effects appear to pose a special challenge for the DRC model (Coltheart et al., 2001), which has some difficulty simulating them (Zevin & Seidenberg, 2006).

Although regularity and consistency correlate highly, these dimensions are separable. Distinguishing between these two variables is particularly valuable for adjudicating between the rule-based DRC approach (which predicts regularity effects) and the connectionist approach (which predicts consistency effects). Indeed, Cortese and Simpson (2000) crossed these two variables factorially in a speeded pronunciation experiment, and compared their results with simulated data from three computational models of word recognition. They observed stronger effects of consistency than regularity, a pattern that was captured best by Plaut et al.'s (1996) PDP model.

The above-mentioned studies have all emphasized the consistency of the *rime* unit (i.e., the vowel and consonant cluster after the onset of a syllable); *bind, find, hind,* and *mind* are all rime neighbors of *kind.* However, Treiman, Kessler, and Bick (2003) showed that the pronunciation of a vowel can also be influenced both by the consistency of its onset and coda. Thus, consistency in pronunciation appears to be sensitive to multiple grain sizes.

Semantic Richness

A growing number of reports in the literature indicate that word recognition is facilitated for semantically *richer* words (i.e., words that are associated with relatively more semantic information; for reviews, see Balota, Ferraro, & Connor, 1991; Pexman, 2012). This is theoretically intriguing because in virtually all models of word recognition, it would appear that a word has to be recognized before its meaning is obtained (Balota, 1990). This is at odds with available empirical evidence which suggests that the system has access to meaning before a word is fully identified, possibly via feedback activation from semantic to orthographic and phonological units (Balota et al., 1991; Pexman, 2012). Although the ultimate goal of reading is to extract meaning from visually printed words, the influence of meaning-level influences on word recognition remains poorly understood.

A number of dimensions have been identified that appear to tap the richness of a word's semantic representation, including the *number of semantic features* associated with its referent (McRae, Cree, Seidenberg, & McNorgan, 2005); its *number of semantic neighbors* (Shaoul & Westbury, 2010); the *number of distinct first associates* elicited by the word in a free-association task (Nelson, McEvoy, & Schreiber, 1998); *imageability*, the extent to which a word evokes mental imagery (Cortese & Fugett, 2004); *number of senses*, the number of meanings associated with a word (Miller, 1990); *body-object interaction*, the extent to which a human body can interact with a word's referent (Siakaluk, Pexman, Aguilera, Owen, & Sears, 2008); and *sensory experience ratings*, the extent to which a word evokes a sensory or perceptual experience (Juhasz & Yap, 2013). Across tasks, words from denser semantic neighborhoods, which possess more meanings and evoke more imagery, and whose referents are associated with more features or are easier for the human body to interact with are recognized faster (e.g., Yap, Pexman, Wellsby, Hargreaves, & Huff, 2012).

Importantly, the different richness variables account for unique (i.e., nonoverlapping) variance in word recognition performance (Yap, Pexman, et al., 2012), implying that no single richness dimension (and its associated theoretical framework) can adequately explain how meaning is derived from print. Instead, semantic memory is best conceptualized as multidimensional (Pexman, Siakaluk, & Yap, 2013).

In addition to the richness dimensions described above, the emotional *valence* (positive, neutral, negative) and *arousal* of a word influence lexical decision and speeded pronunciation performance. For example, *snake* is a negative, high-arousal word, while *sleep* is a positive, low-arousal word. A number of early studies suggested that negative, compared with neutral and positive, stimuli are responded to more slowly. This slowing is consistent with the idea that negative stimuli attract attention in early processing, and more time is needed to disengage attention from these stimuli before a lexical decision or pronunciation response can be made (see Kuperman, Estes, Brysbaert, & Warriner, 2014, for a review). However, this conclusion is qualified by a meta-analysis revealing that the negative and neutral words used in the studies were not always well matched on lexical characteristics (Larsen, Mercer, & Balota, 2006). Although the results of better-controlled studies are somewhat mixed, a recent large-scale analysis of valence and arousal effects for over 12,000 words, which controlled for many lexical and semantic factors, suggests that valence and arousal exert independent and monotonic effects, such that negative (compared with positive) and arousing (compared with calming) words are recognized more slowly (Kuperman et al., 2014).

Finally, an intriguing aspect of the semantic richness literature involves the extent to which is that the strength of these effects is modulated by the specific demands of a lexical processing task (Balota & Yap, 2006). For example, semantic richness accounts for much more item-level variance in the category verification task than in tasks where semantic processing is not the primary basis for responding. Yap, Tan, Pexman, and Hargreaves (2011) also found that words with more senses were associated with faster lexical decision times but less accurate category verification performance. This result is consistent with the notion that multiple meanings can hurt performance in a task that requires participants to resolve the specific meaning of a word.

Context/Priming Effects

Thus far we have described variables that influence isolated word recognition. There is also a rich literature directed at how different contexts or primes influence word recognition processes. In a typical priming paradigm, two letter strings are presented successively that have some dimension of similarity. Specifically, the two strings might be morphologically (*touching-TOUCH*), orthographically (*couch-TOUCH*), phonologically (*much-TOUCH*), or semantically/associatively related (*feel-TOUCH*). Primes can either be *unmasked* (i.e., consciously available) or *masked* (i.e., presented briefly to minimize conscious processing). The key advantage of the masked priming paradigm is that participants are usually unaware of the relationship between the prime and the target, thereby minimizing strategic effects (Forster, 1998; see also Kinoshita & Lupker, 2003). In this section, we limit our coverage to phonological, morphological, and semantic priming effects. Kinoshita (this volume) and Perea (this volume) provide excellent reviews of orthographic priming effects and discuss how this important work constrains models that address how readers code letter position in words (see also Frost, this volume).

Phonological Priming Effects

What is the role of phonological codes in visual word recognition (Frost, 1998)? Do these codes automatically precede and constrain the identification of words, or is phonology generated after lexical access? These controversial questions have been extensively investigated with the masked priming paradigm and other paradigms (see Halderman, Ashby, & Perfetti, 2012, for a review). For example, Lukatela and Turvey (2000) reported that compared with a control prime (e.g., *clep*), phonologically related primes (e.g., *klip*) facilitated lexical decision responses to targets (i.e., *CLIP*), even when primes were presented for only 14 ms. Indeed, in an important meta-analysis of masked phonological priming studies in English, Rastle and Brysbaert (2006) concluded that there were small but reliable effects of masked phonological priming in perceptual identification, pronunciation, and lexical decision. To confirm this, Rastle and Brysbaert (2006) conducted two masked priming experiments that demonstrated that words (e.g., *GROW*) were recognized 13 ms faster on average when they were preceded by phonologically similar primes (*groe*) than by orthographic controls (*groy*). Collectively, these results provide compelling evidence for an early and pervasive influence of phonological processes in word recognition. These phonological processes potentially help in stabilizing the identity of words so that they can be perceived accurately (Halderman et al., 2012; see also Pollatsek, this volume).

Morphological Priming Effects

Morphemes are the smallest units of meaning in words, and many English words are multimorphemic. An important debate in the literature concerns the extent to which the morphemic constituents in a word serve as access units during word recognition (see Taft, this volume). For example, are morphologically complex words such as *painter* automatically decomposed into their morphemic subunits (i.e., *paint* + *er*) prior to lexical access (Taft & Forster, 1975) or does each complex word have its own representation? Relatedly, does the morphological decomposition procedure distinguish between inflected words that are more semantically *transparent* (i.e., the meaning of the word can be predicted from its constituents, e.g., *sadness*) and words that are more semantically *opaque* (e.g., *department*)? The answers to such questions help shed light on the representations and processes underlying morphological processing.

To better delineate the time course of morphological processes, researchers rely heavily on the masked morphological priming paradigm. Using this tool, they have established that recognition of a target word (e.g., *SAD*) is facilitated by the masked presentation of morphologically related words (i.e., *sadness*) (Rastle, Davis, Marslen-Wilson, & Tyler, 2000). By using appropriate controls, Rastle et al. (2000) have shown that such morphological priming effects cannot be simply attributed to semantic or orthographic overlap between primes and targets, and hence provide compelling evidence for early and obligatory decomposition of morphologically complex words into morphemes prior to lexical access.

Interestingly, Rastle, Davis, and New (2004) have also reported that masked morphological priming effects are equivalent in magnitude for transparent (e.g., *cleaner—CLEAN*) and opaque (e.g., *corner—CORN*) prime-target pairs,[1] suggesting that the initial morphological decomposition process is blind to semantics and based entirely on the analysis of orthography. That being said, the role of semantics in morphological processing is still not entirely clear. A meta-analysis of the literature revealed a small but reliable effect of *semantic transparency*. That is, transparent primes facilitate target recognition to a greater extent than opaque primes (Feldman, O'Connor, & del Prado Martin, 2009), consistent with an early semantic influence on morphological processing (but see Davis & Rastle, 2010).

These patterns are theoretically important because they challenge the connectionist frameworks which posit that morphemic effects emerge via interactions among orthography, phonology, and semantics (e.g., Gonnerman, Seidenberg, & Andersen, 2007); such frameworks predict less priming for opaque than for transparent prime-target pairs (Plaut & Gonnerman, 2000). For a more extensive discussion of the morphological processing literature, readers are encouraged to consult Diependaele, Grainger, and Sandra (2012).

"Semantic" Priming Effects

The *semantic priming effect* refers to the robust finding that words are recognized faster when preceded by a semantically related prime (e.g., *cat-DOG*) than when preceded by a semantically unrelated prime (e.g., *mat-DOG*) (Meyer & Schvaneveldt, 1971).[2] The semantic priming literature provides important insights into the architecture of the mental lexicon and the processes used to retrieve information from that network. The "semantic" in semantic priming effect is largely an expository convenience (McNamara, 2005), since the effect may reflect an associative relationship between the two words rather than an overlap in their semantic features. For example, *dog* and *cat* share both a semantic and associative relationship, whereas *mouse* and *cheese* primarily share an associative relationship. While a review by Lucas (2000) suggests there are instances where semantic priming effects truly reflect shared semantic information, a follow-up review by Hutchison (2003) yields the more guarded conclusion that a simple associative account can accommodate most of the priming literature. What else do we know about semantic priming?

Related primes facilitate target recognition even when primes are heavily masked and cannot be consciously identified (Balota, 1983; Fischler & Goodman, 1978), suggesting that the meaning of a prime word can be processed, even if it is not consciously identifiable. This claim is consistent with an intriguing phenomenon known as the *mediated priming effect*. In mediated priming, *lion* is able to prime *STRIPES* (Balota & Lorch, 1986). Although there is no obvious direct relationship between the two words, priming is able to occur through the mediating concept *tiger*. These results are consistent with the classic study by Neely (1977), who demonstrated that semantic priming effects can occur at short stimulus onset asynchronies even when attention is directed to a different area of semantic memory.

A number of theoretical mechanisms have been proposed to explain semantic priming; these mechanisms are not mutually exclusive and may well operate together (see McNamara, 2005, for a review). *Automatic spreading activation* (Posner & Snyder, 1975) is the canonical explanation for semantic priming. That is, a prime (e.g., *cat*) automatically preactivates related nodes (e.g., *DOG*) via associative/semantic pathways, facilitating recognition of these related words when they are subsequently presented (see Collins & Loftus, 1975). Priming may also partly reflect *expectancy*, or the strategic generation of potential candidates for the upcoming target (Becker, 1980); facilitation is observed when the expectancy is correct. Finally, there is evidence that priming effects in the lexical decision task implicate postlexical decision processes. Specifically, participants may engage in backward semantic checking from the target to the prime (Neely, Keefe, & Ross, 1989), since the absence or presence of a prime-target relationship is diagnostic of the target's lexicality (nonwords are never related to the primes). Space constraints preclude a comprehensive survey of this interesting and important area of research, but readers are directed to Neely (1991) and McNamara (2005) for excellent reviews of the semantic/associative priming literature.

Joint Effects of Variables

Heretofore we have emphasized the main effects of variables. However, researchers are typically more interested in the extent to which multiple variables interact to influence word recognition performance. Indeed, such interactions are particularly useful for constraining theoretical models. For example, stimulus length interacts with orthographic neighborhood size, such that there is an increasing facilitatory effect of orthographic neighborhood size for long, compared to short, words (see Balota et al., 2004). In addition, low-frequency words produce larger effects of both orthographic neighborhood size and length than high-frequency words (Balota et al., 2004) in the speeded pronunciation task, but not in the lexical decision task. It is possible that the reduced effects of variables for high-frequency words may reflect better established lexical representations for these items.

There is also considerable evidence for interactions within the priming literature. For example, semantic priming typically interacts with word frequency and stimulus quality, such that priming effects are larger for low-frequency (e.g., Becker, 1979) and degraded (Becker & Killion, 1977) word targets. However, stimulus quality and word frequency produce robust additive effects (Stanners, Jastrzembski, & Westbrook, 1975) in the lexical decision task but not in either the word pronunciation or semantic classification task (Yap & Balota, 2007). There is also recent evidence that priming produces additive effects with the difficulty of the nonword distracters in the lexical decision task (Lupker & Pexman, 2010). Traditional priming accounts (e.g., spreading activation, expectancy) are too simple to capture this complex constellation of additive and interactive effects (McNamara, 2005), and it may be necessary to turn to models that possess multiple stages or levels of lexical-semantic representation (for an example, see Yap, Balota, & Tan, 2013). An important next step within computational modeling will be the development of models that can account for both the additive and interactive effects of targeted variables (see Plaut & Booth, 2000, 2006, for a potential framework, and also Borowsky & Besner, 2006, for a discussion of limitations of this approach).

Newer Approaches and Analytic Tools in Visual Word Recognition Research

Megastudies Versus Factorial Studies of Word Recognition

The most common experimental design in word recognition research is the factorial design, where independent variables of interest are manipulated and extraneous variables are controlled for. Although this approach is useful, like all approaches, it has some limitations (see Balota, Yap, Hutchison, & Cortese, 2012, for a review). The *megastudy* approach allows the language to define the stimuli, rather than the experimenter selecting stimuli based on a limited set of criteria. In megastudies, researchers examine word recognition for very large sets of words, such as virtually all English monosyllabic words (Balota et al., 2004; Treiman, Mullennix, Bijeljac-Babic, & Richmond-Welty, 1995) or multisyllabic monomorphemic words (Yap & Balota, 2009). In addition to identifying the unique predictive power of a large set of targeted variables, along with their interactive effects (Balota et al., 2004), megastudies have proven valuable for adjudicating between computational models of word recognition

(Perry et al., 2007), comparing competing metrics of word frequency (Brysbaert & New, 2009), evaluating the impact of novel psycholinguistic variables (Juhasz & Yap, 2013; Yarkoni et al., 2008), exploring potential nonlinear functional relationships between factors and word recognition performance (New et al., 2006), and investigating the role of individual differences in word recognition (Yap, Balota, Sibley, & Ratcliff, 2012).

The megastudy approach is aided by the availability of freely accessible online databases containing lexical characteristics and behavioral data for large sets of words. For example, the English Lexicon Project (ELP; Balota et al., 2007; http://elexicon.wustl.edu) provides lexical decision and speeded pronunciation measures for over 40,000 English words, along with a search engine that indexes a wide variety of lexical variables (see also the British Lexicon Project; Keuleers, Lacey, Rastle, & Brysbaert, 2011). The ELP has stimulated a flurry of related megastudies in other languages, including the French Lexicon Project (Ferrand et al., 2010), the Dutch Lexicon Project (Keuleers, Diependaele, & Brysbaert, 2010), the Malay Lexicon Project (Yap, Rickard Liow, Jalil, & Faizal, 2010), and the Chinese Lexicon Project (Sze, Rickard Liow, & Yap, 2014). Researchers have been turning to crowd-sourcing tools such as Amazon's Mechanical Turk (Mason & Suri, 2012) or smartphone apps to rapidly collect norms (e.g., concreteness ratings; Brysbaert, Warriner, & Kuperman, 2014) and behavioral data (Dufau et al, 2011). Researchers have also recently started developing databases that explore the influence of context on word recognition. For example, the Semantic Priming Project (Hutchison et al., 2013; http://spp.montana.edu) and the Form Priming Project (Adelman et al., 2014), respectively, serve as behavioral databases of semantic priming and masked form priming performance.

While one might be concerned that large-scale data may not be sensitive to more subtle manipulations (e.g., the interaction between frequency and consistency; Sibley, Kello, & Seidenberg, 2009), recent analyses indicate that databases such as the English Lexicon Project reproduce the standard effects in the literature (Balota et al., 2012). Thus megastudies provide a useful complement to the factorial studies in the literature.

Analyses of Response Time Distributions

In the overwhelming majority of studies in word recognition, researchers compare the *mean* response time across different conditions to determine whether their data are consistent with the predicted hypotheses. To the extent that empirical response time distributions are symmetrical and experimental manipulations primarily shift distributions, this approach works quite well. However, empirical distributions are virtually always positively skewed, and experimental effects can both shift and modulate the shape of a distribution (Heathcote, Popiel, & Mewhort, 1991). Thus, relying solely on analyses comparing means is potentially both inadequate and misleading (Heathcote et al., 1991). Fortunately, a number of approaches have been developed for understanding the influence of variables on the underlying response time distribution. The first and ultimately optimal method is to fit the data to a computational model (e.g., diffusion model; Ratcliff, 1978) that is able to generate specific predictions about experimental effects on the characteristics of the response time distribution. In the absence of such a model, researchers can (1) evaluate the influence of manipulations on the parameters of a mathematical function (e.g., the ex-Gaussian function, the sum of the normal and exponential distribution) fitted to an empirically obtained response time distribution or (2) generate descriptive plots (e.g., quantile plots) of how a manipulation differentially affects different regions of the distribution.

By augmenting conventional means-based analyses with distributional methods, researchers have gained finer-grained insights into the processes underlying isolated word recognition and semantic priming (see Balota & Yap, 2011, for a selective review). Consider the classic semantic priming effect, in which participants recognize *CAT* faster when it is preceded by *dog* than by an unrelated word like *dig*. Across a series of studies, there is evidence that the semantic priming effect in highly skilled readers is purely mediated by distributional shifting (Balota, Yap, Cortese, & Watson, 2008). That is, the benefit afforded by a related prime is constant, regardless of target difficulty (for a replication in masked semantic priming, see Gomez, Perea, & Ratcliff, 2013). Distributional shifting is most consistent with the idea that for such readers priming reflects relatively modular processes, whereby primes preactivate related words through automatic spreading activation and provide readers with a processing head-start when the words are subsequently presented. When word identification is compromised in some way, priming is no longer entirely mediated by a shift; instead, priming effects increase monotonically as target difficulty increases. One sees this pattern when targets are visually degraded (Balota et al., 2008; Yap et al., 2013) or when less skilled readers are processing unfamiliar low-frequency words (Yap, Tse, & Balota, 2009). That is, when target identification is effortful, readers can strategically retrieve prime information to aid in resolving the target (Thomas, Neely, & O'Connor, 2012).

Although it is tempting to map distributional parameters or aspects of the response time distribution onto specific cognitive processes, it is important not to do this in the absence of converging evidence (Matzke & Wagenmakers, 2009). The key point here is that there is a growing literature which suggests that one can gain important insights into lexical processes by moving beyond simple measures of central tendency and considering response time distributional analyses.

Individual Differences

Empirical work and models of word recognition have traditionally focused on group-level performance (but see Zevin & Seidenberg, 2006, for an exception). However, there is compelling evidence that individual differences in reading skill can modulate word recognition performance (see Andrews, this volume; see also Yap, Balota, et al., 2012, for a review). For example, vocabulary knowledge appears to moderate the joint effects of priming and word frequency (Yap et al., 2009). For readers with smaller vocabularies, priming and word frequency interact; priming effects are larger for low-frequency words. In contrast, highly skilled readers with a larger vocabulary produce robust main effects of priming and word frequency but no interaction.

The advent of large datasets containing individual participant data makes it possible to explore individual differences with very large samples. For example, in their analysis of the trial-level lexical decision and speeded pronunciation data contributed by over 1,200 participants in the English Lexicon Project, Yap, Balota, et al. (2012) made a number of noteworthy observations. Importantly, Yap, Balota, et al. reported considerable within- and between-session reliability across distinct sets of items with respect to overall mean response time, response time distributional characteristics, diffusion model parameters, and effects of theoretically important variables such as word frequency and length. Readers with more vocabulary knowledge showed faster, more accurate word recognition performance and attenuated sensitivity to stimulus characteristics. Collectively, results such as these suggest that participants are associated with relatively stable distributional and processing profiles that extend beyond average processing speed. Moving forward, it will be increasingly important to develop models that can capture both group-level performance and the variability across individual readers.

Concluding Remarks

The research in visual word recognition provides exciting insights into the early stages of reading and has been the focus of important principles in cognitive modeling, including interactive activation, rule-based coding, connectionist modeling, and more recently, notions of optimal perceivers from a Bayesian perspective. Although there has been considerable progress, different tasks bring with them task-specific operations that can influence the results. Hence one must be cognizant of the interplay between task-general lexical processes and task-specific processes when considering this literature. Finally, because of space constraints, the reader should be reminded that this is at best a brief snapshot of the visual word recognition literature, and we have focused primarily on behavioral studies in adult readers. For example, research in cognitive neuroscience continues to provide important constraints for word recognition models (Taylor, Rastle, & Davis, 2013; see Woollams, this volume). We anticipate that visual word recognition will continue to be at the heart of fundamental breakthroughs in understanding how people read.

Notes

1 Rastle and colleagues do not distinguish between semantically opaque prime-target pairs that share both an etymological and surface morphological relationship (e.g., *department-DEPART*) and pairs that share only the surface relationship (e.g., *corner-CORN*), because such a distinction is difficult to reconcile with a plausible theory of language acquisition (Rastle & Davis, 2003). However, there are researchers (e.g., Longtin, Segui, & Hallé, 2003) who make this distinction and would consider *corner* a pseudoaffixed word.

2 The extent to which two words (e.g., *cat* and *dog*) are related is typically captured by free association norms (e.g., Nelson et al., 1998), which are derived from participants' responses to cue words. An alternative approach, which assumes that a word's meaning is tied to the contexts it appears in, examines the co-occurrence of words in a large text corpus (Landauer & Dumais, 1997). Word pairs which that co-occur more frequently are considered to be more strongly related (Jones, Kintsch, & Mewhort, 2006).

References

Adelman, J. S., & Brown, G. D. (2008). Modeling lexical decision: The form of frequency and diversity effects. *Psychological Review, 115,* 214–229.

Adelman, J. S., Johnson, R. L., McCormick, S. F., McKague, M., Kinoshita, S., Bowers, J. S., . . . Davis, C. J. (2014). A behavioral database for masked form priming. *Behavior Research Methods, 46,* 1052–1067.

Andrews, S. (1997). The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin & Review, 4,* 439–461.

Ans, B., Carbonnel, S., & Valdois, S. (1998). A connectionist multi-trace memory model of polysyllabic word reading. *Psychological Review, 105,* 678–723.

Appelman, I. B., & Mayzner, M. S. (1981). The letter-frequency effect and the generality of familiarity effects on perception. *Perception & Psychophysics, 30,* 436–446.

Balota, D. A. (1983). Automatic semantic activation and episodic memory encoding. *Journal of Verbal Learning and Verbal Behavior, 22,* 88–104.

Balota, D. A. (1990). The role of meaning in word recognition. In D. A. Balota, G. B. Flores d'Arcais, & K. Rayner (Eds.), *Comprehension processes in reading* (pp. 9–32). Hillsdale, NJ: Erlbaum.

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 340–357.

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General, 133,* 283–316.

Balota, D. A., Ferraro, F. R., & Connor, L. T. (1991). On the early influence of meaning in word recognition: A review of the literature. In P. J. Schwanenflugel (Ed.), *The psychology of word meanings* (pp. 187–218). Hillsdale, NJ: Erlbaum.

Balota, D. A., & Lorch, R. F. (1986). Depth of automatic spreading activation: Mediated priming effects in pronunciation but not in lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 336–345.

Balota, D. A., & Spieler, D. H. (1999). Word frequency, repetition, and lexicality effects in word recognition tasks: Beyond measures of central tendency. *Journal of Experimental Psychology: General, 128,* 32–55.

Balota, D. A., & Yap, M. J. (2006). Attentional control and flexible lexical processing: Explorations of the magic moment of word recognition. In S. Andrews (Ed.), *From inkmarks to ideas: Current issues in lexical processing* (pp. 229–258). New York, NY: Psychology Press.

Balota, D. A., & Yap, M. J. (2011). Moving beyond the mean in studies of mental chronometry: The power of response time distributional analyses. *Current Directions in Psychological Science, 20,* 160–166.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., . . . Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods, 39,* 445–459.

Balota, D. A., Yap, M. J., Cortese, M. J., & Watson, J. M. (2008). Beyond mean response latency: Response time distributional analyses of semantic priming. *Journal of Memory and Language, 59,* 495–523.

Balota, D. A., Yap, M. J., Hutchison, K. A., & Cortese, M. J. (2012). Megastudies: What do millions (or so) of trials tell us about lexical processing? In J. S. Adelman (Ed.), *Visual word recognition: Vol. 1. Models and methods, orthography and phonology* (pp. 90–115). Hove, England: Psychology Press.

Becker, C. A. (1979). Semantic context and word frequency effects in visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 5,* 252–259.

Becker, C. A. (1980). Semantic context effects in visual word recognition: An analysis of semantic strategies. *Memory & Cognition, 8,* 493–512.

Becker, C. A., & Killion, T. H. (1977). Interaction of visual and cognitive effects in word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 389–401.

Borowsky, R., & Besner, D. (2006). Parallel distributed processing and lexical-semantic effects in visual word recognition: Are a few stages necessary? *Psychological Review, 113,* 181–193.

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41,* 977–990.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods, 46,* 904–911.

Carreiras, M., & Grainger, J. (2004). Sublexical units and the 'front end' of visual word recognition. *Language and Cognitive Processes, 19,* 321–331.

Cattell, J. M. (1886). The time it takes to see and name objects. *Mind, 11,* 63–65.

Collins, A., & Loftus, E. (1975). A spreading activation theory of semantic processing. *Psychological Review, 82,* 407–428.

Coltheart, M., Davelaar, E., Jonasson, J., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI* (pp. 535–555). Hillsdale, NJ: Erlbaum.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108,* 204–256.

Cortese, M. J., & Fugett, A. (2004). Imageability ratings for 3,000 monosyllabic words. *Behavior Research Methods, Instruments andComputers, 36,* 384–387.

Cortese, M. J., & Simpson, G. B. (2000). Regularity effects in word naming: What are they? *Memory & Cognition, 28,* 1269–1276.

Davis, M. H., & Rastle, K. (2010). Form and meaning in early morphological processing: Comment on Feldman, O'Connor, and Moscoso del Prado Martin (2009). *Psychonomic Bulletin & Review, 17,* 749–755.

Diependaele, K., Grainger, J., & Sandra, D. (2012). Derivational morphology and skilled reading: An empirical overview. In M. J. Spivey, K. McRae, & M. Joanisse (Eds.), *The Cambridge handbook of psycholinguistics* (pp. 311–333). Cambridge, England: Cambridge University Press.

Dufau, S., Duñabeitia, J. A., Moret-Tatay, C., McGonigal, A., Peeters, D., Alario, F.-X., . . . Grainger, J. (2011). Smart phone, smart science: How the use of smartphones can revolutionize research in cognitive science. *PLoS ONE, 6,* e24974.

Feldman, L. B., O'Connor, P. A., & del Prado Martin, F. M. (2009). Early morphological processing is morphosemantic and not simply morpho-orthographic: A violation of form-then-meaning accounts of word recognition. *Psychonomic Bulletin & Review, 16,* 684–691.

Ferrand, L., New, B., Brysbaert, M., Keuleers, E., Bonin, P., Méot, A., . . . Pallier, C. (2010). The French Lexicon Project: Lexical decision data for 38,840 French words and 38,840 pseudowords. *Behavior Research Methods, 42,* 488–496.

Fischler, I., & Goodman, G. O. (1978). Latency of associative activation in memory. *Journal of Experimental Psychology: Human Perception and Performance, 39,* 1731–1740.

Forster, K. I. (1976). Accessing the mental lexicon. In R. J. Wales & E. C. T. Walker (Eds.), *New approaches to language mechanisms* (pp. 257–287). Amsterdam, the Netherlands: North-Holland.

Forster, K. I. (1998). The pros and cons of masked priming. *Journal of Psycholinguistic Research, 27,* 203–233.

Frost, R. (1998). Toward a strong phonological theory of visual word recognition: True issues and false trails. *Psychological Bulletin, 123,* 71–99.

Gomez, P., Perea, M., & Ratcliff, R. (2013). A diffusion model account of masked vs. unmasked priming: Are they qualitatively different? *Journal of Experimental Psychology: Human Perception and Performance, 39,* 1731–1740.

Gonnerman, L. M., Seidenberg, M. S., & Andersen, E. S. (2007). Graded semantic and phonological similarity effects in priming: Evidence for a distributed connectionist approach to morphology. *Journal of Experimental Psychology: General, 136,* 323–345.

Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review, 103,* 518–565.

Grainger, J., Rey, A., & Dufau, S. (2008). Letter perception: From pixels to pandemonium. *Trends in Cognitive Sciences, 12,* 381–387.

Halderman, L. K., Ashby, J., & Perfetti, C. A. (2012). Phonology: An early and integral role in identifying words. In J. S. Adelman (Ed.), *Visual word recognition: Vol. 1. Models and methods, orthography and phonology* (pp. 207–228). Hove, England: Psychology Press.

Heathcote, A., Popiel, S. J., & Mewhort, D. J. K. (1991). Analysis of response time distributions: An example using the Stroop task. *Psychological Bulletin, 109,* 340–347.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology, 160,* 106–154.

Hutchison, K. A. (2003). Is semantic priming due to association strength or feature overlap? A micro-analytic review. *Psychonomic Bulletin & Review, 10,* 785–813.

Hutchison, K. A., Balota, D. A., Neely, J. H., Cortese, M. J., Cohen-Shikora, E. R., Tse, C-S., . . . Buchanan, E. (2013). The semantic priming project. *Behavior Research Methods, 45,* 1099–1114.

Jared, D. (2002). Spelling-sound consistency and regularity effects in word naming. *Journal of Memory and Language, 46,* 723–750.

Jones, M. N., Kintsch, W., & Mewhort, D. J. K. (2006). High-dimensional semantic space accounts of priming. *Journal of Memory and Language, 55,* 534–552.

Juhasz, B. J., & Yap, M. J. (2013). Sensory experience ratings (SERs) for over 5,000 mono- and disyllabic words. *Behavior Research Methods, 45,* 160–168.

Keuleers, E., Diependaele, K., & Brysbaert, M. (2010). Practice effects in large-scale visual word recognition studies: A lexical decision study on 14,000 Dutch mono- and disyllabic words and nonwords. *Frontiers in Psychology, 1,* 174.

Keuleers, E., Lacey, P., Rastle, K., & Brysbaert, M. (2011). The British Lexicon Project: Lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behavior Research Methods, 44,* 287–304.

Kinoshita, S., & Lupker, S. J. (Eds.). (2003). *Masked priming: State of the art.* New York, NY: Psychology Press.

Kuperman, V., Estes, Z., Brysbaert, M., & Warriner, A. B. (2014). Emotion and language: Valence and arousal affect word recognition. *Journal of Experimental Psychology: General, 143,* 1065–1081.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104,* 211–240..

Larsen, R. J., Mercer, K. A., & Balota, D. A. (2006). Lexical characteristics of words used in emotional Stroop experiments. *Emotion, 6,* 62–72.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95,* 492–527.

Longtin, C.-M., Segui, J., & Hallé, P. A. (2003). Morphological priming without morphological relationship. *Language and Cognitive Processes, 18,* 313–334.

Lucas, M. (2000). Semantic priming without association: A meta-analytic review. *Psychonomic Bulletin & Review, 7,* 618–630.

Lukatela, G., & Turvey, M. T. (2000). An evaluation of the two-cycles model of phonology assembly. *Journal of Memory and Language, 42,* 183–207.

Lupker, S. J., & Pexman, P. M. (2010). Making things difficult in lexical decision: The impact of pseudohomophones and transposed-letter nonwords on frequency and semantic priming effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 1267–1289.

Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods, 44,* 1–23.

Matzke, D., & Wagenmakers, E.-J. (2009). Psychological interpretation of ex-Gaussian and shifted Wald parameters: A diffusion model analysis. *Psychonomic Bulletin & Review, 16,* 798–817.

McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review, 86,* 287–330.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review, 88,* 375–407.

McNamara, T. P. (2005). *Semantic priming: Perspectives from memory and word recognition.* Hove, England: Psychology Press.

McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods, 37,* 547–559.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90,* 227–234.

Miller, G. A. (1990). Word Net: An on-line lexical database. *International Journal of Lexicography, 3,* 235–312.

Morton, J. (1970). A functional model for memory. In D. A. Norman (Ed.), *Models of human memory* (pp. 203–254). New York, NY: Academic Press.

Murray, W. S., & Forster, K. I. (2004). Serial mechanisms in lexical access: The rank hypothesis. *Psychological Review, 111,* 721–756.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General, 106,* 226–254.

Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 236–264). Hillsdale, NJ: Erlbaum.

Neely, J. H., Keefe, D. E., & Ross, K. L. (1989). Semantic priming in the lexical decision task: Roles of prospective prime-generated expectancies and retrospective semantic matching. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 1003–1019.

Neisser, U. (1967). *Cognitive psychology.* New York, NY: Appleton-Century-Crofts.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms. w3.usf.edu/FreeAssociation/.

New, B., Ferrand, L., Pallier, C., & Brysbaert, M. (2006). Reexamining the word length effect in visual word recognition: New evidence from the English Lexicon Project. *Psychonomic Bulletin & Review, 13,* 45–52.

Norris, D. (2006). The Bayesian reader: Explaining word recognition as an optimal Bayesian decision process. *Psychological Review, 113,* 327–357.

Paap, K. R., McDonald, J. E., Schvaneveldt, R. W., & Noel, R. W. (1987). Frequency and pronounceability in visually presented naming and lexical-decision tasks. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 221–243). Hillsdale, NJ: Erlbaum.

Perry, C., Ziegler, J. C., & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: The CDP+ model of reading aloud. *Psychological Review, 114,* 273–315.

Perry, C., Ziegler, J. C., & Zorzi, M. (2010). Beyond single syllables: Large-scale modeling of reading aloud with the connectionist dual process (CDP++) model. *Cognitive Psychology, 61,* 106–151.

Petersen, S. E., Fox, P. T., Posner, M. I., Mintun, M., & Raichle, M. E. (1989). Positron emission tomographic studies of the processing of single words. *Journal of Cognitive Neuroscience, 1,* 153–170.

Pexman, P. M. (2012). Meaning-based influences on visual word recognition. In J. S. Adelman (Ed.), *Visual word recognition: Vol. 2. Meaning and context, individuals and development* (pp. 24–43). Hove, England: Psychology Press.

Pexman, P. M., Siakaluk, P. D., & Yap, M. J. (2013). Introduction to the research topic meaning in mind: Semantic richness effects in language processing. *Frontiers in Human Neuroscience, 7,* 723.

Plamondon, R., & Srihari, S. N. (2000). On-line and off-line handwriting recognition: A comprehensive survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22,* 63–84.

Plaut, D. C. (1997). Structure and function in the lexical system: Insights from distributed models of word reading and lexical decision. *Language and Cognitive Processes, 12,* 765–805.

Plaut, D. C., & Booth, J. R. (2000). Individual and developmental differences in semantic priming: Empirical and computational support for a single-mechanism account of lexical processing. *Psychological Review, 107,* 786–823.

Plaut, D. C., & Booth, J. R. (2006). More modeling but still no stages: Reply to Borowsky and Besner. *Psychological Review, 113,* 196–200.

Plaut, D. C., & Gonnerman, L. M. (2000). Are non-semantic morphological effects incompatible with a distributed connectionist approach to lexical processing? *Language and Cognitive Processes, 15,* 445–485.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103,* 56–115.

Posner, M. I., & Snyder, C. R. R. (1975). Attention and cognitive control. In R. Solso (Ed.), *Information processing and cognition: The Loyola symposium* (pp. 55–85). Hillsdale, NJ: Erlbaum.

Rastle, K., & Brysbaert, M. (2006). Masked phonological priming effects in English: Are they real? Do they matter? *Cognitive Psychology, 53,* 97–145.

Rastle, K., & Coltheart, M. (2006). Is there serial processing in the reading system; and are there lexical representations? In S. Andrews (Ed.), *From inkmarks to ideas: Current issues in lexical processing* (pp. 3–24). New York, NY: Psychology Press.

Rastle, K., & Davis, M. H. (2003). Reading morphologically complex words: Some thoughts from masked priming. In S. Kinoshita & S. J. Lupker (Eds.), *Masked priming: The state of the art* (pp. 279–305). Hove, England: Psychology Press.

Rastle, K., Davis, M. H., Marslen-Wilson, W., & Tyler, L. K. (2000). Morphological and semantic effects in visual word recognition: A time-course study. *Language and Cognitive Processes, 15,* 507–538.

Rastle, K., Davis, M. H., & New, B. (2004). The broth in my brother's brothel: Morpho-orthographic segmentation in visual word recognition. *Psychonomic Bulletin & Review, 11,* 1090–1098.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85,* 59–108.

Ratcliff, R., Gomez, P., & McKoon, G. (2004). A diffusion model account of the lexical decision task. *Psychological Review, 111,* 159–182.

Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology, 81,* 274–280.

Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review, 89,* 60–94.

Sears, C. R., Hino, Y., & Lupker, S. J. (1995). Neighborhood size and neighborhood frequency effects in word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 876–900.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed developmental model of word recognition and naming. *Psychological Review, 96,* 523–568.

Selfridge, O. G., & Neisser, U. (1960). Pattern recognition by machine. *Scientific American, 203,* 60–68.

Shaoul, C., & Westbury, C. (2010). Exploring lexical co-occurrence space using HiDEx. *Behavior Research Methods, 42,* 393–413.

Siakaluk, P. D., Pexman, P. M., Aguilera, L., Owen, W. J., & Sears, C. R. (2008). Evidence for the activation of sensorimotor information during visual word recognition: The body-object interaction effect. *Cognition, 106,* 433–443.

Sibley, D. E., Kello, C. T., & Seidenberg, M. S. (2009). *Error, error everywhere: A look at megastudies of word reading.* Proceedings of the Annual Meeting of the Cognitive Science Society. Amsterdam, The Netherlands.

Stanners, R. F., Jastrzembski, J. E., & Westbrook, A. (1975). Frequency and visual quality in a word-nonword classification task. *Journal of Verbal Learning and Verbal Behavior, 14,* 259–264.

Stone, G. O., & Van Orden, G. C. (1993). Strategic control of processing in word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 744–774.

Sze, W. P., Rickard Liow, S. J., & Yap, M. J. (2014). The Chinese Lexicon Project: A repository of lexical decision behavioral responses for 2,500 Chinese characters. *Behavior Research Methods, 46,* 263–273.

Taft, M., & Forster, K. I. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior, 14,* 638–647.

Taylor, J. S. H., Rastle, K., & Davis, M. H. (2013). Can cognitive models explain brain activation during word and pseudoword reading? A meta-analysis of 36 neuroimaging studies. *Psychological Bulletin, 139,* 766–791.

Thomas, M. A., Neely, J. H., & O'Connor, P. (2012). When word identification gets tough, retrospective semantic processing comes to the rescue. *Journal of Memory and Language, 66,* 623–643.

Treiman, R., & Kessler, B. (2007). Learning to read. In M. G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp. 657–666). Oxford, England: Oxford University Press.

Treiman, R., Kessler, B., & Bick, S. (2003). Influence of consonantal context on the pronunciation of vowels: A comparison of human readers and computational models. *Cognition, 88,* 49–78.

Treiman, R., Mullennix, J., Bijeljac-Babic, R., & Richmond-Welty, E. D. (1995). The special role of rimes in the description, use, and acquisition of English orthography. *Journal of Experimental Psychology: General, 124,* 107–136.

Treisman, A. (1999). Solutions to the binding problem: Progress through controversy and convergence. *Neuron, 24,* 105–110.

Weekes, B. S. (1997). Differential effects of number of letters on word and nonword naming latency. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 50A,* 439–456.

Whaley, C. (1978). Word-nonword classification time. *Journal of Verbal Learning and Verbal Behavior, 17,* 143–154.

Wheeler, D. D. (1970). Processes in word recognition. *Cognitive Psychology, 1,* 59–85.

Yap, M. J., & Balota, D. A. (2007). Additive and interactive effects on response time distributions in visual word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 274–296.

Yap, M. J., & Balota, D. A. (2009). Visual word recognition of multisyllabic words. *Journal of Memory and Language, 60,* 502–529.

Yap, M. J., Balota, D. A., Sibley, D. E., & Ratcliff, R. (2012). Individual differences in visual word recognitions: Insights from the English lexicon project. *Journal of Experimental Psychology: Human Perception and Performance, 38,* 53–79.

Yap, M. J., Balota, D. A., & Tan, S. E. (2013). Additive and interactive effects in semantic priming: Isolating lexical and decision processes in the lexical decision task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39,* 140–158.

Yap, M. J., Pexman, P. M., Wellsby, M., Hargreaves, I. S., & Huff, M. J. (2012). An abundance of riches: Cross-task comparisons of semantic richness effects in visual word recognition. *Frontiers in Human Neuroscience, 6,* 72.

Yap, M. J., Rickard Liow, S. J., Jalil, S. B., & Faizal, S. S. B. (2010). The Malay Lexicon Project: A database of lexical statistics for 9,592 words. *Behavior Research Methods, 42,* 992–1003.

Yap, M. J., Tan, S. E., Pexman, P. M., & Hargreaves, I. S. (2011). Is more always better? Effects of semantic richness on lexical decision, speeded pronunciation, and semantic classification. *Psychonomic Bulletin & Review, 18,* 742–750.

Yap, M. J., Tse, C.-S., & Balota, D. A. (2009). Individual differences in the joint effects of semantic priming and word frequency: The role of lexical integrity. *Journal of Memory and Language, 61,* 303–325.

Yarkoni, T., Balota, D. A., & Yap, M. J. (2008). Beyond Coltheart's N: A new measure of orthographic similarity. *Psychonomic Bulletin & Review, 15,* 971–979.

Yates, M. (2005). Phonological neighbors speed visual word processing: Evidence from multiple tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 1385–1397.

Yates, M., Friend, J., & Ploetz, D. M. (2008a). The effect of phonological neighborhood density on eye movements during reading. *Cognition, 107,* 685–692.

Yates, M., Friend, J., & Ploetz, D. M. (2008b). Phonological neighbors influence word naming through the least supported phoneme. *Journal of Experimental Psychology: Human Perception and Performance, 34,* 1599–1608.

Zevin, J. D., & Seidenberg, M. (2006). Simulating consistency effects and individual differences in nonword naming. *Journal of Memory and Language, 54,* 145–160.