

## Chapter 10

# Visual Word Recognition and Pronunciation: A Computational Model of Acquisition, Skilled Performance, and Dyslexia

*Mark S. Seidenberg and James L. McClelland*

---

Word recognition is an integral part of reading and may be the single most extensively studied topic in cognitive psychology. Aside from its obvious importance to the reading process, the topic has generated interest for several reasons: because learning to recognize words is among the first tasks beginning readers confront, because failure to acquire age-appropriate reading skills is typically associated with deficits in word processing, and because the study of word recognition impairments following brain injury has provided important evidence concerning the neuropsychological bases of complex behavior. The main characteristics of word processing in reading are known as a result of extensive study (for reviews see papers in Coltheart 1987a and Besner, Waller, and MacKinnon 1985). Primary among those characteristics is the fact that for skilled readers the process is very rapid and largely unconscious. The picture that has emerged is that lexical processing yields access to several types of information in a rapid and efficient manner. Readers are typically aware of the results of lexical processing, not the manner in which it occurs. One of the goals of research on visual word recognition has been to use experimental methods to unpack these largely unconscious processes. In this paper we present an overview of a new computational model of visual word recognition, the acquisition of word recognition skills, and the breakdown of these skills in dyslexia.

### *Scope of the Problem*

The problem to be addressed is this: The lexical processor operates so as to rapidly yield several types of information associated with a given word; the skilled reader is able to discriminate an input string from thousands of other vocabulary items within a fraction of a second. What knowledge supports processing of this kind, and how does the child acquire it?

According to the hypothesis-testing view that has dominated reading research for many years, rapid word identification is possible because of information provided by the *literal* context in which a word occurs (see Henderson 1982 for review). Words appear in meaningful contexts; information provided by the context, in conjunction with knowledge of the language and general world knowledge, allows the reader to formulate hypotheses concerning the identity of subsequent words. At any given point in a sentence, the range of likely continuations is thought to be limited, so that less information needs to be extracted from the word itself, and this facilitates recognition. In Kenneth Goodman's phrase, reading is a "psycholinguistic guessing game" (Goodman 1967); the skilled reader is the person able to use contextual information in an efficient manner to facilitate recognition.

Questions concerning the scope of contextual effects on word recognition and the mechanisms responsible for such effects continue to be the focus of considerable attention and debate (see McClelland 1987 and Tanerhaus, Dell, and Carlson 1988). However, studies of children's reading have suggested that differences in the use of contextual information in recognizing words are not the primary source of differences in reading skill (Stanovich 1986). Other knowledge plays a critical role in word processing, specifically, the reader's knowledge of the lexicon itself. Rapid word recognition is possible because readers exploit information concerning the structure and distribution of word-forms in the language. A theory of recognition must characterize the relevant aspects of lexical structure, their representation in memory, and their use in decoding. This information represents the *virtual* context for word recognition, and is at least as important as the information provided by the literal context. The interactive-activation model of McClelland and Rumelhart (1981) was an attempt at characterizing readers' knowledge of the lexicon and its use in decoding, and the present model can be seen as its successor.

### *What Is to Be Learned*

In acquiring word recognition skills, children must come to understand at least two basic characteristics of written English. First there is the alphabetic principle (Rozin and Gleitman 1977), the fact that in an alphabetic orthography there are systematic correspondences between the spoken and written forms of words. Beginning readers already possess large oral vocabularies; their initial problem is to learn how unfamiliar written forms map onto known spoken forms. The scope of this problem is determined by facts about the writing system. The alphabetic system for writing English is a code for representing spoken language; units in the writing system—letters and letter pat-

terns—largely correspond to speech units such as phonemes. However, the correspondence between the written and spoken codes is notoriously complex; many correspondences are inconsistent (e.g., *-ave* is usually pronounced as in *gave*, *save*, and *cave*, but there is also *have*) or wholly arbitrary (e.g., *-olo-* in *colonel*, *-ps* in *corps*).

A second aspect of the writing system the child must learn about concerns the distribution of letter patterns in the lexicon. Only some combinations of letters are possible, and the combinations differ in frequency. These facts about the distribution of letter patterns give written English its characteristic redundancy. Many aspects of orthographic redundancy derive from the fact that the writing system is primarily (though not exclusively) a cipher for spoken language. For example, the fact that letters *gp* never appear in word-initial position derives from a phonotactic constraint on the occurrence of the corresponding phonemes. Of the many possible combinations of 26 letters, only a small percentage yield letter strings that are permissible words in English. An even smaller percentage are realized as actual words in the lexicon. These constraints on the forms of written words may play an important role in the recognition process. The reader must discriminate the input string from other words in his or her vocabulary, a task that might be facilitated by knowledge of the letter combinations that are permissible or realized.

In sum, the child's problem is to learn how English is represented in written form. This task might be facilitated by the systematic aspects of the writing system, i.e., the constraints on possible letter sequences and the correspondences between spelling and sound. However, there are barriers to using these types of information. Facts about orthographic redundancy cannot be utilized until the child is familiar with a large number of words. Acquiring useful generalizations about spelling-sound correspondences is inhibited by the fact that many words have irregular correspondences and the fact that these words are overrepresented among the items the child learns to read first (*give*, *have*, *some*, *does*, *gone*, etc.). The child must nonetheless learn to use knowledge of the orthography in a manner that supports the recognition of words within a fraction of a second.

Our model addresses the acquisition and use of these two types of information: orthographic redundancy and orthographic-phonological correspondences. The model is realized within the connectionist framework being applied to many problems in perception and cognition (see Rumelhart and McClelland 1986; McClelland and Rumelhart 1986a). The model provides an account of how orthographic redundancy and orthographic-phonological correspondences are utilized in recognition and pronunciation. On this account, learn-

ing to read words involves learning facts about the distribution of letter patterns in the lexicon and the correspondences between orthography and phonology. This knowledge can be represented in the terms of the weights on the connections in a distributed memory network that consists of simple processing units. Learning involves modifying the weights through experience in reading and pronouncing words. We will argue that this connectionist approach is ideally suited to accounting for word recognition because of the nature of the task, which is largely determined by characteristics of the orthography.

The model gives a detailed account of a range of empirical issues of continuing interest to reading researchers, including differences between words in terms of processing difficulty, differences between readers in terms of word recognition skill, transitions from beginning to skilled reading, the role of phonology in reading, and differences between silent reading and reading aloud. The model also provides an account of certain forms of dyslexia that are observed developmentally and as a consequence of brain injury.

### *Overview of the Model*

The model represents part of a more general theory of lexical processing. The goal is an integrated theory that accounts for several aspects of lexical processing, including access of meaning and pronunciation from print, access of meaning and spelling from speech, and access of spelling and pronunciation from meaning. The implemented model is concerned with how readers perceive letter strings and pronounce them aloud. Here we give an overview of the model; details are provided in Seidenberg and McClelland, *in press*. The model, which has the general form illustrated in figure 10.1, consists of a network of interconnected processing units. There are 400 units used to code letter strings, 200 hidden units, and 460 units used to code phonemic output. The distributed representation used for coding letter strings is similar to one described by Hinton, McClelland, and Rumelhart (1986). The scheme for coding phonological output, which also makes use of distributed representations, is described by McClelland and Rumelhart (1986b). The main feature of these representations is that a given letter or phoneme in a word is represented by a pattern of activation across a set of nodes, rather than by a single node. The network is a feed-forward system with complete connectivity from input units to hidden units, and from hidden units to output units. Each connection carries a weight that governs the forward spread of activation through the system. Weights on the connections are initially arbitrary and modified during a learning phase using the

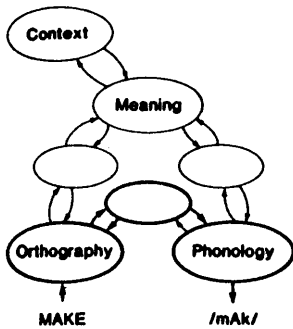


Figure 10.1

A general framework for lexical processing. The implemented model is outlined in bold.

back-propagation algorithm of Rumelhart, Hinton, and Williams (1986).

The system takes letter strings as input, and produces two kinds of output: a pattern of activation across the phonological units, and a recreation of the input pattern across the orthographic nodes. The former can be thought of as the node's computation of the phonological code for a letter string. The latter can be thought of as a computation of an orthographic code that provides an index of the familiarity of the letter string. These codes provide the basis for performing tasks such as naming and lexical decision, as discussed below.

The model was trained on a set of 2,897 monosyllabic English words, a large proportion of the monosyllabic words in the language. The training procedure worked as follows. On each trial the model was given a stimulus pair consisting of a letter string and its pronunciation. The letter string was presented to the model, and the output—a pattern of activation across the phonological nodes and a re-creation of the letter string across the orthographic nodes—was computed using a simple spreading activation procedure, in which the activation of a unit is a function of the sum of the weighted activations along the lines coming into it. The model was initially configured with small random weights on the connections between units. The goal of training was to find a set of weights that enabled the model to produce optimal orthographic and phonological output for any input word, that is, output that closely approximated correct orthographic and phonological patterns. The patterns of activation actually produced by a given input can be compared to those that would be produced if the model performed perfectly. We characterize the

differences between expected and obtained orthographic output in terms of error scores, which are the squared differences between actual and obtained output summed over all of the orthographic units; we derive a similar error score for phonological output. These error terms have two functions. First, they provide measures of the model's performance; error scores for words vary as a function of amount of training and as a function of lexical variables such as orthographic redundancy, orthographic-phonological regularity, and frequency. Second, the error scores were utilized in the learning algorithm, which was used to modify the weights on connections, as described by Seidenberg and McClelland (in press).

The model, then, produces output specifying the orthographic and phonological codes of the input string. These codes are then used in performing tasks such as naming and lexical decision. We characterize the model's performance in terms of error scores calculated for different types of words after different amounts of training and relate these to human performance on these tasks.

Before presenting the results of the simulations in more detail, we shall review the empirical phenomena it addresses. We shall then describe how various aspects of the model's performance relate to human performance. Finally, we shall consider some of the model's implications concerning acquired and developmental dyslexia.

### *Empirical Phenomena*

Words vary in terms of their orthographic properties and in terms of orthographic-phonological correspondences. A large number of studies have investigated the effects of these aspects of word structure on recognition latencies. Five types of words are of primary interest.

- *Regular words* contain spelling patterns that recur in a large number of words, always with the same pronunciation. *Must*, for example, contains the spelling pattern *-ust*; all monosyllabic words that end in this pattern rhyme (*just*, *dust*, etc.). The pool of words sharing this spelling pattern are termed neighbors (Glushko 1979). These words can be pronounced by grapheme-phoneme correspondence rules, such as those proposed by Venezky (1970).
- *Exception words* contain common spelling patterns that are given irregular pronunciations. For example, *-ave* is usually pronounced as in *gave* and *save*, but has an irregular pronunciation in the exception word *have*. In terms of orthographic structure, regular and exception words are similar: both contain spelling

patterns that recur in many words. However, exception words have irregular pronunciations; they cannot be pronounced by rule.

- *Regular inconsistent words* (Glushko 1979) are complements of the exceptions. A word such as *gave* is regular in the sense of containing a common spelling pattern that is given a regular pronunciation in a large pool of words; however, it is inconsistent because there is a similarly-spelled exception word (i.e., *have*). Thus, regular inconsistent words have an exception word in their neighborhoods. These words can also be pronounced by rule.
- *Homographs* are words such as *wind*, *lead*, and *bass* that contain common spelling patterns having two pronunciations, each of which is a word in the language.
- Finally, there are words such as *aisle* and *once*. In contrast to the other classes of words, these items (which Seidenberg et al. [1984] term *strange*) contain unusual spelling patterns that do not recur in a large number of words; hence their neighborhoods are small. Like exception words, they cannot be pronounced by rule.

These types represent clear cases that have been useful in empirical and theoretical work. It should be noted that there are many intermediate cases. For example, a word such as *soap* has an unusual spelling (it is the only monosyllabic word ending in *-oap*), but according to the Venezky (1970) rules, its pronunciation is regular. A spelling pattern such as *-own* has two pronunciations, but in contrast to a pattern such as *-ave*, both pronunciations occur in many words. Empirical research in this area has tended to utilize the clear cases as a way to discover general processing principles. An explicit computational model, however, should be able to account for the entire range of cases.

Contrasts between these types of words have provided a way to investigate the factors that influence word recognition. The comparison between regular and exception words provides information about the computation of the phonological code. These words are similar in terms of orthographic factors; both contain common spelling patterns, and they can be equated in terms of other factors such as length and frequency. Differences between the words in terms of processing difficulty must be attributed to the one dimension along which they differ, namely, regularity of spelling-sound correspondences. The regularity factor might be expected to influence performance of a task such as naming, which requires articulatory output. However, a difference between regular and exception words on a silent reading task such as lexical decision would indicate that readers had computed the phonological code even where overt articulation was not required.

Hence, the contrast between these words provides a way of diagnosing access of this information.

The contrast between regular inconsistent words such as *gave* and regular words such as *must* has been thought to provide a way to test the claim that spelling-sound correspondence rules are used in reading words (Glushko 1979; Henderson 1982). Spelling-sound rules (e.g., a rule governing the pronunciation of *-ave* or a rule that lengthens a vowel when it is followed by a consonant and terminal *e*) correctly specify the pronunciations of regular inconsistent words. However, Glushko (1979) found longer naming latencies for regular inconsistent words than regular words. This result—which indicated that the pronunciation of *gave* was somehow influenced by the irregular *have*—is difficult to reconcile with the rule account. As with exception words, a difference between regular and regular inconsistent words on a silent reading task would indicate that phonological information has been computed.

Homographs provide another basis for distinguishing between orthographic and phonological effects in word recognition. These words are unremarkable from the point of view of orthography; they contain common spelling patterns. Like the spelling patterns in exception and regular inconsistent words, they are associated with two (or more) pronunciations. However, in a homograph both pronunciations yield actual words. Finally, the strange words such as *aisle* provide a way to investigate effects of orthographic redundancy. If the frequency with which a spelling pattern occurs in the lexicon influences processing, strange words should differ from regular words. Moreover, because the strange words contain unusual spelling patterns, their pronunciations may also be difficult to derive.

A large number of studies examining the recognition of such words have yielded a fairly precise set of results, which will be summarized briefly. There are two main findings. First, even among skilled readers of the language, words differ in terms of processing difficulty. Second, different results obtain when the task is lexical decision (a silent reading task) than when the stimuli are named aloud (a pronunciation task).

### *Naming Aloud*

In regard to the effects of irregular spelling-sound correspondences, exception words produce longer naming latencies than regular words only when the stimuli are relatively low in frequency (Andrews 1982; Seidenberg et al. 1984; Seidenberg 1985a; Waters and Seidenberg 1985; Taraban and McClelland 1987). For a large pool of higher-frequency words, irregular spelling-sound correspondences have no



Table 10.1  
Mean naming latencies (in msec) from Seidenberg 1985a, experiment 1

Word type	Example	Subject group			
		Fastest	Medium	Slowest	All
High frequency, regular	Nine	475	523	621	540 (0.4)
High frequency, exception	Lose	475	517	631	541 (0.9)
Difference		0	-6	+10	+3
Low frequency, regular	Mode	500	530	641	556 (2.3)
Low frequency, exception	Deaf	502	562	685	583 (5.1)
Difference		+2	+32	+44	+17

Standard errors in parentheses.

effect on naming latencies. A typical result (from Seidenberg 1985a) is presented in table 10.1.

Although Glushko (1979) found longer naming latencies for regular inconsistent words compared to regular words, later studies have qualified these results. Seidenberg et al. (1984) found that Glushko's results were due in part to repetition of matched regular inconsistent and exception word pairs; for example, the stimuli included *none*, *bone*, *done*, *gone*, and *lone*; *love*, *prove*, *cove*, and *shove*; etc. These repetitions result in priming effects that artifactually increase the magnitude of the regular inconsistent effect (see also Meyer, Schvaneveldt, and Ruddy 1974). Intralist priming will also produce an exception effect for higher frequency words (Treiman, Freyd, and Baron 1984; Taraban and McClelland 1987). When Seidenberg et al. (1984) presented each spelling pattern only once, longer latencies for regular inconsistent words were only obtained for words in the lower frequency range. In an experiment with a larger set of stimuli that also controlled for repetition of spelling patterns, Taraban and McClelland (1987) failed to obtain a statistically significant regular-inconsistent effect for either high or low frequency words. These results suggest that the mere presence of an exception-word neighbor is not sufficient to slow naming latencies for regular inconsistent words, an issue we consider again below.

Effects of the orthographically irregular, strange words also depend on frequency (Seidenberg et al. 1984; Waters and Seidenberg 1985). For higher-frequency words, naming latencies are similar for regular and strange words. For lower-frequency words, however, naming latencies for strange words are the longest of all word classes, including exceptions. Finally, homographs also produce longer naming latencies than regular words (Seidenberg et al. 1984); the effects of frequency have not been investigated with this word class.

In sum, the naming results indicate that for a large pool of high-frequency words, factors such as orthographic redundancy and regularity of orthographic-phonological correspondences have little discernible effect on processing. This pool of words is likely to be quite large because of the type-token facts about English (Seidenberg 1985a). A relatively small number of word types account for a large number of the tokens that a reader encounters. In the Kucera and Francis (1967) count, for example, the 133 most frequent words in the corpus account for about half of the total number of tokens. Hence, a small number of words recur with very high frequency, and these are the words for which the structural variables have little effect. Moreover, Seidenberg (1985a) found that the size of this pool varies as a function of reading skill. Faster readers recognize a larger pool of items without interference from irregular spelling or spelling-sound correspondences. In effect, they treat more words as though they were high-frequency items (see table 10.1).

For lower-frequency, more slowly recognized words, naming latencies depend on structural properties that reflect characteristics of written English. The words with common spelling patterns and regular pronunciations yield the best performance; words with common spelling patterns but irregular pronunciations (exceptions) yield somewhat poorer performance; and strange words, which have irregular spellings and pronunciations, produce the longest latencies and most errors. Note that the fact that effects of word structure are modulated by frequency may go some way towards explaining the inconsistent results of previous studies, most of which failed to consider this factor.

### *Lexical Decision*

The lexical decision task provides important information because it does not require articulatory output. It might be the case that phonological information is accessed only when overt pronunciation is required. As in naming, the effects of the structural variables on lexical decisions are limited to lower-frequency words (Waters and Seidenberg 1985; Seidenberg et al. 1984). For higher-frequency words, all of the word types yield similar lexical decision latencies and numbers of errors. The pattern of results for lower-frequency words in lexical-decision studies differs in important ways from that obtained in naming. In many lexical-decision experiments orthographic-phonological regularity yielded no effects, yet in other experiments it did (see Henderson 1982 and Seidenberg 1985b for reviews). These inconsistent effects have been interpreted as indicating that the recognition of words in silent reading involves both "direct" (visually based) and

"mediated" (phonologically based) processes. Where there were no phonology effects, it was inferred that recognition is direct; where there were such effects, it was thought to be mediated. This interpretation of the lexical-decision results was consistent with the dual-route model. In contrast, the consistent effects of phonological regularity in naming were thought to be a consequence of the task, which requires computation of the phonological code.

While this account has the virtue of reconciling inconsistent experimental results, it did not explain the factors that determined which recognition process would be used in any given case. Note that the inconsistent results that led to this view involved the same types of stimuli (regular and exception words) used in different experiments. Hence, it cannot be the case that direct access is used for one type of word (e.g., exceptions) and mediated access for the other (e.g., regular).

Waters and Seidenberg (1985) provide an alternative account of these seemingly inconsistent results. Their evidence suggests that lexical-decision results depend on the types of words and nonwords included in a stimulus set. Consider first the exception effect. When the stimuli in an experiment contain only regular words, exception words, and pronounceable nonwords, there is no exception effect for lower-frequency words, in contrast to the results in naming (Waters and Seidenberg 1985). Under these conditions the effects of irregular spelling-sound correspondences obtained with the naming task are eliminated.

Consider now an experiment in which the subject sees regular words, exception words, strange words, and pronounceable nonwords. Again, there are no effects of word type among the higher-frequency items. However, for lower-frequency items, an exception effect now obtains (Waters and Seidenberg 1985). That is, with the same regular and exception words and even the same subjects used in the study mentioned in the previous paragraph, exception words produced longer lexical decision latencies than regular words. Strange words produced the longest latencies of all three types. In effect, when regular, exception, and strange words are included in the stimulus set, the results mimic those obtained in naming. The Waters and Seidenberg results for lower frequency words are summarized in table 10.2.

Thus, phonological effects in the lexical decision task depend upon the composition of the stimuli in the experiment. This factor explains the seemingly inconsistent results of previous lexical decision studies. Importantly, the results on the naming task are not affected by this factor; there are robust exception effects for lower-frequency

Table 10.2  
Lexical decision and naming latencies (in msec) as a function of presence or absence of strange words

Word Type	Lexical decision task		Naming task	
	Include str	Exclude str	Include str	Exclude str
LF, regular	606	647	518	549
LF, exception	632	643	559	580
Difference	+26	-4	+41	+31

"Include str" = stimuli included strange words; "Exclude str" = stimuli did not include strange words. "LF" = low frequency.

words whether or not strange words are included (Waters and Seidenberg, 1985).

Three aspects of these results are important. First, the effects of structural variables are limited to lower-frequency words. Second, there are the differences between naming and lexical decision, particularly the fact that the properties of the stimulus set influence only the latter. Finally, the results indicate that both orthographic redundancy and orthographic-phonological correspondences influence recognition. The orthographically irregular strange words produced longer latencies than regular or exception words in both tasks. In contrast, effects of orthographic-phonological regularity were robust in naming but depended on the presence of strange words in lexical decision. Moreover, the latencies for exception words were intermediate between those of regular and strange words.

### Developmental Trends

Studies of children's acquisition of word-recognition skills (e.g., Backman et al. 1984; Barron 1981) have addressed the question of how children reach the steady state observed in adults; they have also sought to identify the bases of failures to acquire age-appropriate reading skills and specific reading disability (dyslexia). Children in the earliest stages of learning to read typically recognize words by "sounding out," that is, they attempt to derive the pronunciation of a written word and match it to a known phonological form. However, word recognition processes rapidly change during the first few years of schooling. The study by Backman et al. (1984) examined the acquisition of the naming skill. Children named written regular, exception, and regular inconsistent words and nonwords derived from these items; the stimuli also included a fourth type of word, termed *ambiguous*. These words (e.g., *love*, *town*) include spelling patterns associated with two or more pronunciations, each of which occurs in

many words (e.g., *love, glove, shove, cove, dove, rove; town, clown, brown, flown, known, blown*). All of the stimuli were words that are high-frequency items in adult vocabularies. The subjects were children in grades 2, 3, 4 and high school reading at or above age-appropriate levels ("good readers") and children in grades 3 and 4 reading below age-appropriate levels ("poor readers"). Some of the main results from the study are presented in figure 10.2. Differences between word classes were manifested in number of mispronunciation errors.

The developmental trend exhibited in these data is clear: younger, less-skilled readers have more difficulty with the words associated with multiple pronunciations (exception, regular inconsistent, ambiguous); they show larger regularity effects. As children acquire reading skills, the differences between word classes are eliminated. The less-skilled readers have weaker knowledge of spelling-sound correspondences; this lack of knowledge is a liability when words have irregular, inconsistent, or ambiguous spelling-sound correspondences. Older children and adults are able to compute the pronunciations of high-frequency exemplars of all word classes about equally well; differences between word classes persist only for lower-frequency items. In effect, the unskilled readers' performance in nam-

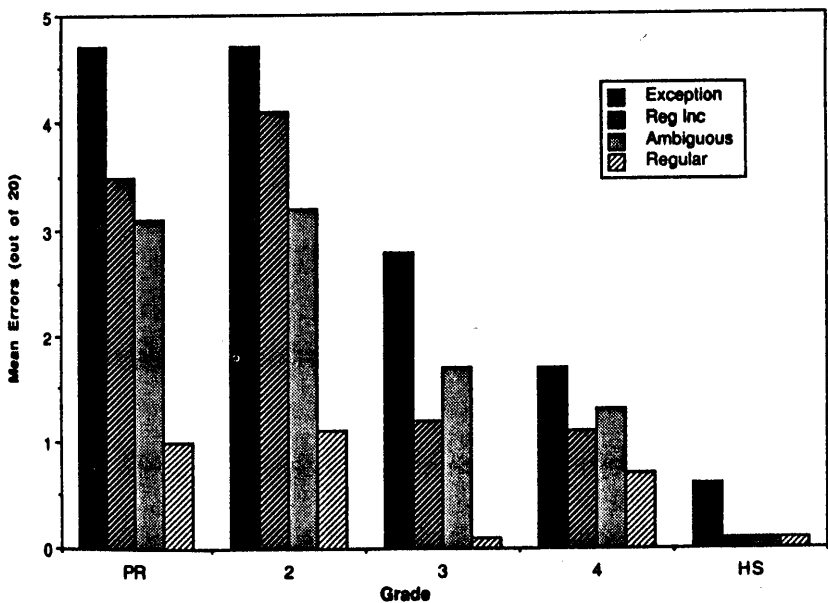


Figure 10.2

Results of the Backman et al. 1984 naming study. PR = poor readers in grades 3 and 4; HS = high school students.

ing higher-frequency words is similar to that of skilled readers in naming of lower-frequency words. At the same time that children are achieving the ability to name different types of words equally well, their knowledge of spelling-sound correspondences is expanding, as evidenced by the older readers' superior performance in reading nonwords.

It is known that acquiring knowledge of spelling-sound correspondences is a key component of learning to read; disorders in phonological analysis skills are thought to be a primary source of reading disability, and children who are backward readers (Backman et al. 1984) or developmental dyslexics (Seidenberg et al. 1986) exhibit relatively poor performance in naming words and nonwords aloud (see Jorm 1983 and Stanovich 1986 for reviews). One of the primary developmental trends observed in studies such as Backman et al. 1984 is that although children acquiring age-expected reading skills initially have more difficulty naming higher-frequency exception words (and other items containing spelling patterns associated with multiple pronunciations) than regular words, this deficit is eliminated by about grade 5 (10 years of age). During the first few years of instruction children learn to name common exception words as efficiently as regular words. Even among skilled adult readers, however, lower-frequency exception words continue to produce longer naming latencies and more errors than lower-frequency regular words. Note that the differences between good and poor younger readers in terms of the number of words read without interference from irregular spelling-sound correspondences are seen at a higher level of performance in skilled readers (Seidenberg 1985a). In both groups the number of words in this pool is related to reading skill. In the model these effects simply derive from the amount of training experience.

Both poor readers who are reading below age-expected levels and children who have been diagnosed as developmental dyslexics fail to show this improvement in naming higher-frequency exception words. For example, the naming performance of the poor readers in grades 3 and 4 in the study of Backman et al. was like that of good readers in grade 2. Both the younger and poorer readers made more errors on exception words and other items containing spelling patterns associated with multiple pronunciations.

### *Simulations*

The network was trained on a list of 2,897 monosyllabic words. The list included all of the uninflected, monosyllabic words from Kucera and Francis 1967 plus additional words not appearing in that corpus

minus a few foreign words, acronyms, and abbreviations. Stimuli ranged in length from three to seven letters and in frequency from 0 to 69,971 on the Kucera and Francis (1967) count. The primary data reported below concern the performance of the model on a subset of 240 words after varying amounts of training. Of these words 192 were used in a behavioral study by Taraban and McClelland (1987). These included 24 words from each of four categories created by crossing the factors of frequency (high and low) and type (exception and regular inconsistent). Each of these words was paired with a regular word matched in terms of frequency, initial phoneme, and length. The test items also include 24 high-frequency and 24 low-frequency strange words similar to those used in the experiments of Seidenberg et al. (1984) and Waters and Seidenberg (1985). Examples of the test stimuli are given in table 10.3.

The learning phase consisted of a series of epochs. For each epoch a subset of the 2,897 items were probabilistically selected for learning trials on the basis of their frequencies; 450 to 550 items were presented per epoch. Connection weights were modified according to the learn-

Table 10.3  
Examples of stimuli used in simulations and experiments

Type	High frequency	Low frequency
Exception	<i>come</i>	<i>gross</i>
	<i>done</i>	<i>pint</i>
	<i>foot</i>	<i>pear</i>
	<i>are</i>	<i>spook</i>
Regular control	<i>came</i>	<i>grape</i>
	<i>dark</i>	<i>peel</i>
	<i>fact</i>	<i>pump</i>
	<i>out</i>	<i>stunt</i>
Regular inconsistent	<i>base</i>	<i>brood</i>
	<i>catch</i>	<i>cove</i>
	<i>cool</i>	<i>harm</i>
	<i>dear</i>	<i>rave</i>
Regular inconsistent control	<i>bird</i>	<i>brisk</i>
	<i>clean</i>	<i>cope</i>
	<i>corn</i>	<i>hide</i>
	<i>dust</i>	<i>reef</i>
Strange	<i>sign</i>	<i>aisle</i>
	<i>film</i>	<i>debt</i>
	<i>jazz</i>	<i>myth</i>
	<i>corps</i>	<i>soap</i>

ing procedure described by Seidenberg and McClelland (in press). Data are presented for 250 learning epochs. The performance of the model on all 2,897 items was tested after 5 epochs, after 5 additional epochs, and after subsequent intervals of 10 epochs. The main results concern the 240 test items. The dependent measures in these analyses are the orthographic and phonological error terms for each item after a given number of learning epochs. These data will be related to results from behavioral studies by Taraban and McClelland (1987), Seidenberg et al. (1984), Waters and Seidenberg (1985) and others. Results for orthographic and phonological output will be considered separately.

### *Results*

*Orthographic-phonological regularity and naming* We assume that overt naming involves three cascaded processes. The phonological code for the input letter string must be computed; the computed phonological code must be translated into a set of articulatory-motor commands; and the articulatory motor code must be executed, resulting in the overt response. Only the first of these processes is implemented in our model. In practice, the phonological output computed by the model is closely related to observed naming latencies.

A word is named by recoding the computed phonological output into a set of articulatory-motor commands, which are then executed. Differences in naming latencies primarily derive from differences in the quality of the computed phonological output. Informally, a word that the model "knows" well produces phonological output that more clearly specifies its articulatory-motor program than a word that is known less well. Thus, naming latencies are a function of phonological error scores, which index differences between observed and expected output.

Differences in naming latencies could also be associated with the execution of the compiled articulatory-motor programs. The distributions of phonemes in high- and low-frequency words differ; some phonemes and phoneme sequences occur more often in high-frequency words than low-frequency words, and vice versa. Phonemes also differ in terms of ease of articulation (Locke 1972); higher-frequency words may contain more of the phonemes that are easier to pronounce, or it may be that the phonemes characteristic of high-frequency words are easier to pronounce because they are used more often. Thus, naming latencies could differ for high- and low-frequency words not because frequency influences the computation of phonological output or the translation of this output into an articulatory code but because words in the two groups contain phonemes



that differ in terms of ease of articulation. We will ignore this aspect of the naming process for two reasons. First, we have not implemented procedures for producing articulatory output. More importantly, existing studies indicate that the lexical variables of interest—frequency, orthographic-phonological regularity, orthographic redundancy, syllabic structure, etc.—have their primary effects on the computation of phonological output. These effects obtain even when articulatory factors are carefully controlled (see Theios and Muise 1977; McRae, Jared, and Seidenberg, in press).

To illustrate, consider a pair of high- and low-frequency homophones such as *main* and *mane*. The model is trained to produce the same phonological output for both words. After a sufficient amount of training, both words produce output that resembles the correct phonological output more closely than it resembles the phonological output for any other string of phonemes. Thus, the model produces the correct phonological codes of both words, from which pronunciations are assumed to be derived. Because they differ in frequency, however, *main* produces a smaller error score than *mane*; in general, the model performs better on words to which it has been exposed more often. It will be easier, then, to compile the pronunciation of the high-frequency item than the low-frequency item, resulting in faster naming latencies for *main* (McRae, Jared, and Seidenberg, in press). Because the words involve the same pronunciation, the same articulatory-motor program is used to produce overt responses. Hence, if the words are named after a one-second delay, the compilation stage will have been completed, and they should produce identical naming latencies, which they do (McRae, Jared, et al., in press). If the words had differed in terms of ease of articulation, latency differences would be observed in both immediate and delayed naming (Balota and Chumbley 1985).

**Exception effects** Figure 10.3 presents the results of the simulation for the exception words and matched regular-word controls. Each data point represents the average phonological error score for the 24 items of each type used in the Taraban and McClelland (1987) experiments. The learning sequence is characterized by the following trends. Training reduces the error scores for all words in an approximately logarithmic manner. Throughout training, there are frequency effects; the model performs better on words to which it is exposed more often. Note that although the test stimuli are dichotomized into high- and low-frequency groups, frequency is actually a continuous variable, and it has continuous effects in the model. Early in training, there are large regularity effects for both high- and low-frequency items; in

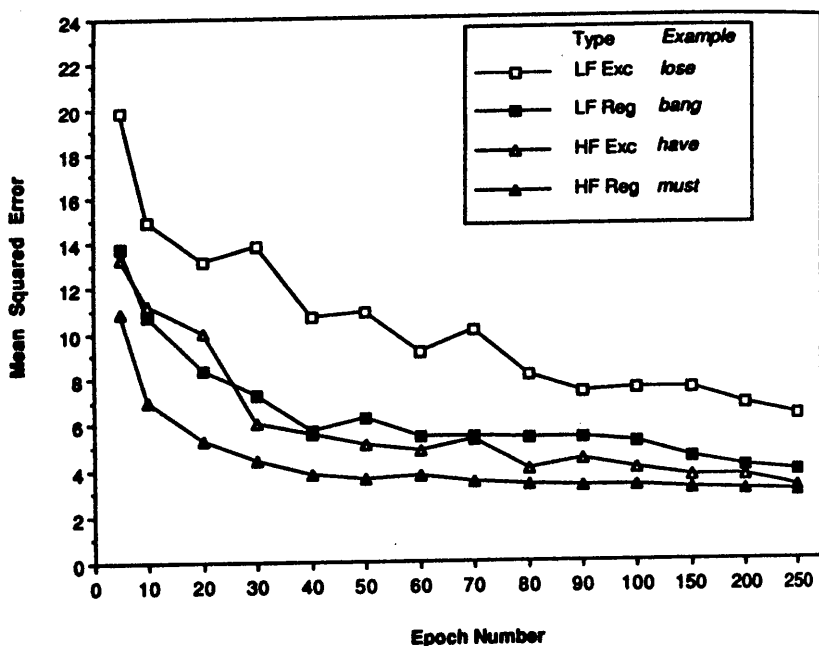


Figure 10.3  
Mean phonological error scores for the stimuli used by Taraban and McClelland (1987).

both frequency classes, regular words produce smaller error terms than exception words. Additional training reduces the exception effect for higher-frequency words to the point where it is eliminated by epoch 250. However, the regularity effect for lower-frequency words remains.

These results relate to behavioral data in the following way. The performance of the model early in training captures a main feature of children's performance early in the acquisition of naming skills. Children in grades 2 to 4 have more difficulty pronouncing exception words than regular words, even when the words are the highest-frequency items in their vocabularies (Backman et al. 1984; Waters, Bruck, et al. 1985). As children acquire reading skills, the exception effect for higher-frequency items is eliminated. By the time children have acquired approximately grade-5 reading skills, they are able to name common exception words as well as regular words. Naming latencies continue to improve with additional experience; however, even among skilled adults there is an interaction between frequency and regularity, because there is a difference between regular and exception words only for relatively infrequent items (Seidenberg et al. 1984; Taraban and McClelland 1987).

The model captures the basic insight that naming performance depends on word structure and familiarity. It also provides a simple account of the differences among skilled adult readers observed by Seidenberg (1985a). That study showed that adults who are good readers nonetheless differ slightly in naming speed. For the fastest readers there was no regularity effect for even for words that occur with low frequency in standard corpora such as Kucera and Francis (1967). The model suggests that the basis for this individual difference is simply the number of times a word is encountered. Skilled readers may read more often, increasing the number of exposures to "low-frequency" items. As in the model, additional exposure tends to decrease the magnitude of the exception effect. It should be noted that word frequencies are unvarying in the model; the probability that a word is sampled on every epoch is determined by its Kucera and Francis frequency, which is fixed. For skilled readers the effect of experience is to move some words from "low" to "high" frequency status, with a resulting decline in the magnitude of the exception effect for such items. In effect, the standard estimates of frequency are not valid for these readers. With additional training, the magnitude of the lower-frequency exception effect in the model would also continue to decrease.

*Regular inconsistent effects* In a well-known study Glushko (1979) reported longer naming latencies for regular inconsistent words such as *gave* compared to regular words such as *must*. As noted above, this effect has not proven to be robust. Seidenberg et al. (1984) obtained a small effect only for lower-frequency words, whereas Taraban and McClelland (1987) did not. The simulation results, presented in figure 10.4, suggest a reason for these inconsistent results. After 250 training epochs, error scores for higher-frequency, regular inconsistent words do not differ from the scores for matched regular words. In the lower-frequency range regular inconsistent words produce larger error scores than regular words, but the difference is very small. These effects may be too small to detect reliably in naming experiments. At best they should be limited to lower-frequency words, as in the Seidenberg et al. 1984 experiment. They should also be larger for less-skilled readers, because the lower-frequency, regular inconsistent effect is larger with fewer training epochs.<sup>1</sup>

Figure 10.5 provides a comparison between the results of the Taraban and McClelland 1987 naming study and the simulation using the same words. Difference scores were obtained by subtracting means for regular words from the means for matched exception and regular,

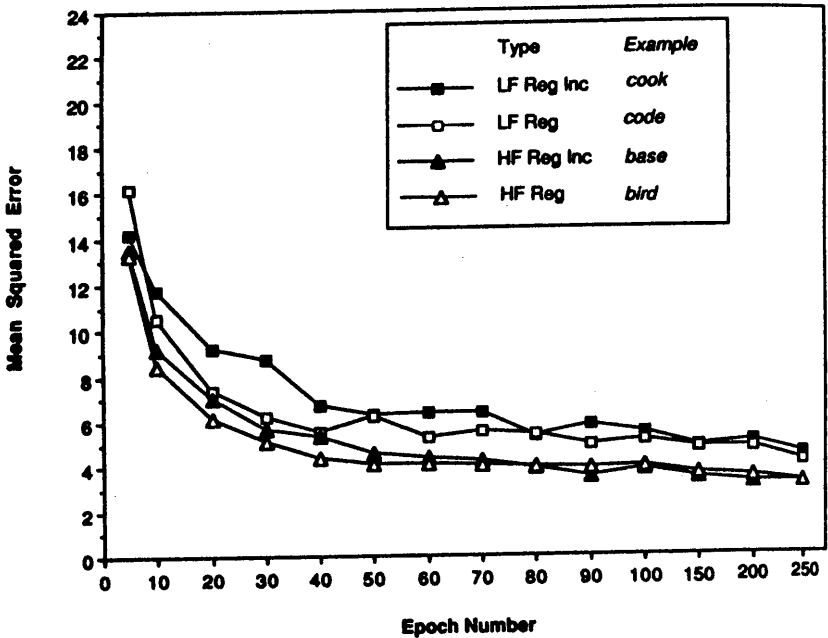


Figure 10.4

The model's performance on regular inconsistent and regular words used in the Taran and McClelland 1987 study.

inconsistent words, respectively. As the graphs indicate, the simulation and naming data closely agree.

**Strange words** Figure 10.6 shows the data for the strange words, which contain unusual spelling patterns, compared to regular words of similar frequency and length (these regular words were also used as controls for the regular inconsistent items). The results are similar to those obtained with other types of words. Throughout the course of learning the model performs better on higher-frequency words than on the lower-frequency words. During the early epochs there is also a main effect of word type; the model performs more poorly on strange words than regular. By 250 epochs, the difference between high-frequency regular and strange words is nearly eliminated, while the difference between low-frequency regular and strange words remains.

The data from skilled human readers also yield this interaction. The 48 strange items used in the simulation have not been employed in any behavioral study, ruling out direct comparisons using the same

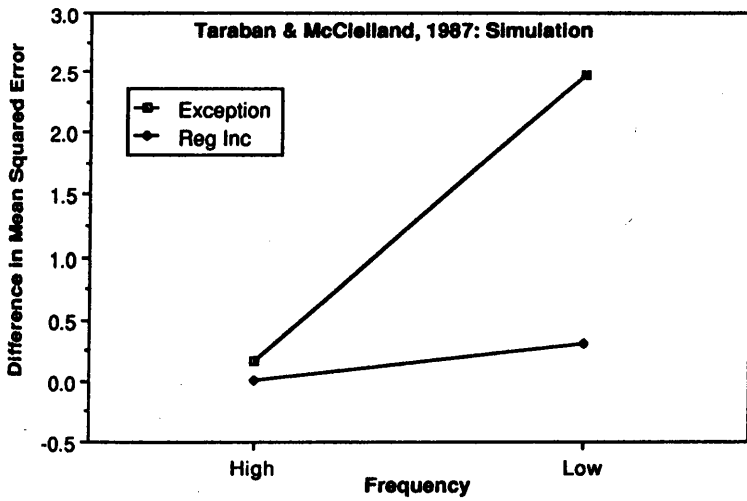
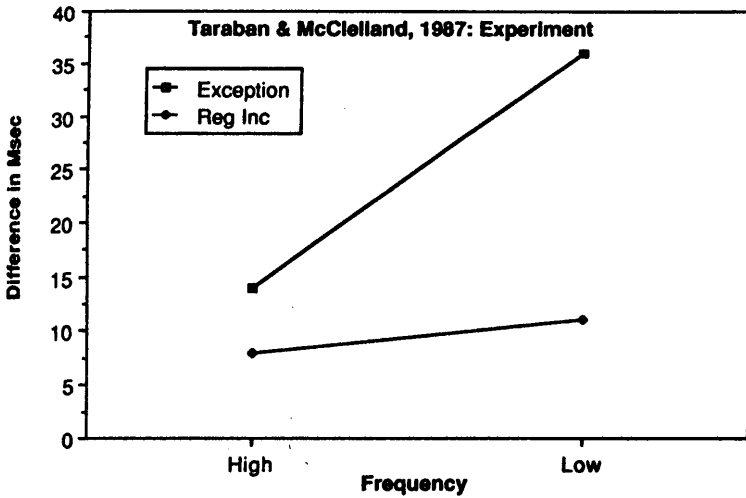


Figure 10.5 Results of the Taraban and McClelland 1987 study: experiment (upper graph) and simulation (lower graph).

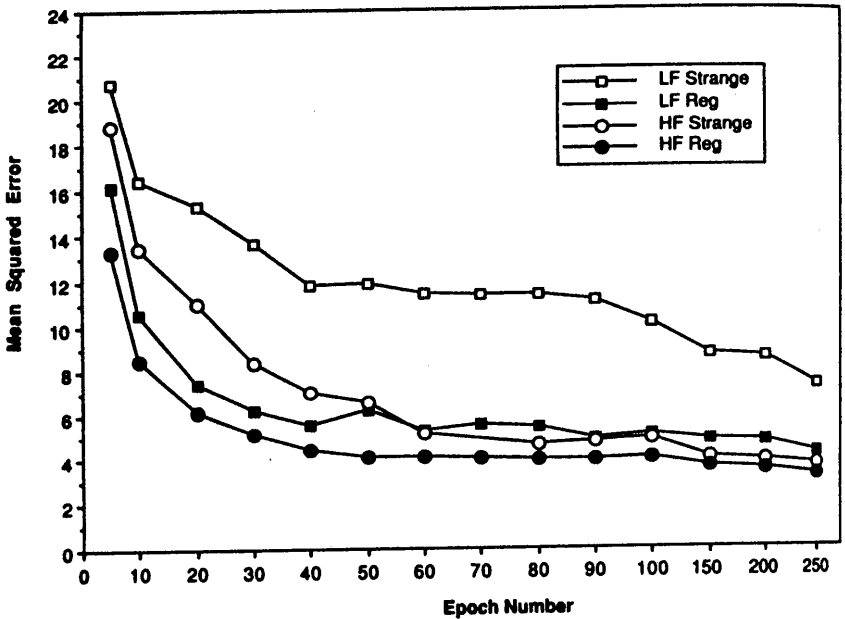


Figure 10.6  
The model's performance on strange words. There were 24 items of each type.

items. However, Waters and Seidenberg (1985) used very similar words in a naming study. Figure 10.7 summarizes the Waters and Seidenberg (1985) data and the model's performance on the same items. In the simulation, there are no differences among exception, strange, and regular high-frequency words; in the lower-frequency range strange items produce the highest error scores, followed by exception and then regular words. Waters and Seidenberg (1985) also found no differences among naming latencies for the three types of higher-frequency words and in the lower-frequency range the same order of relative difficulty: strange > exception > regular.

**Generalizations to novel stimuli** After training, the model has encoded facts about orthographic-phonological correspondences in the weights on the connections from hidden units to phonological-output units. Although the model performs best on the training stimuli, it will compute pronunciations for novel stimuli. In this respect it simulates the performance of subjects asked to pronounce nonwords such as *rone* or *bist*. The model was tested on a set of nonwords derived from the exception words used in the above analysis (table 10.4). For example, *mave* was derived from *have*. These nonwords can

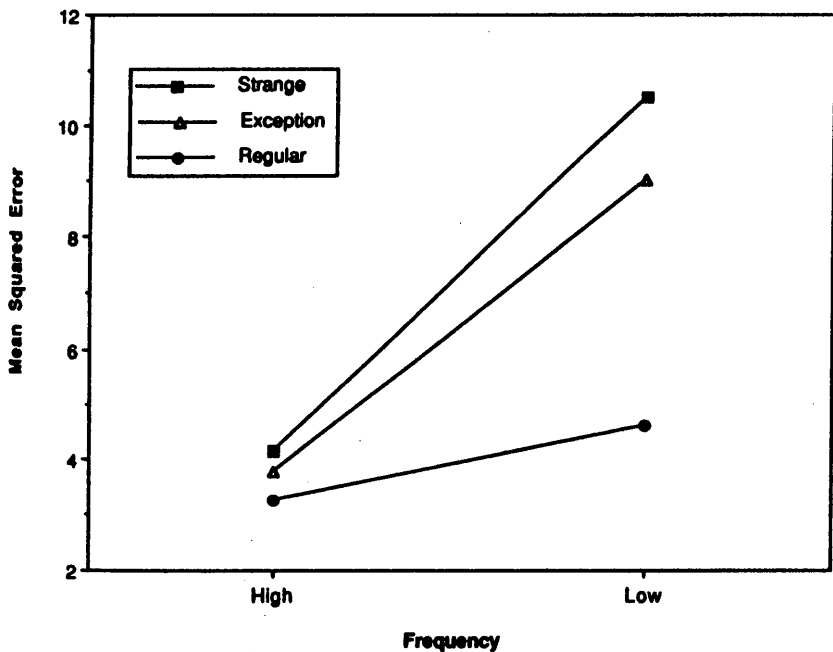
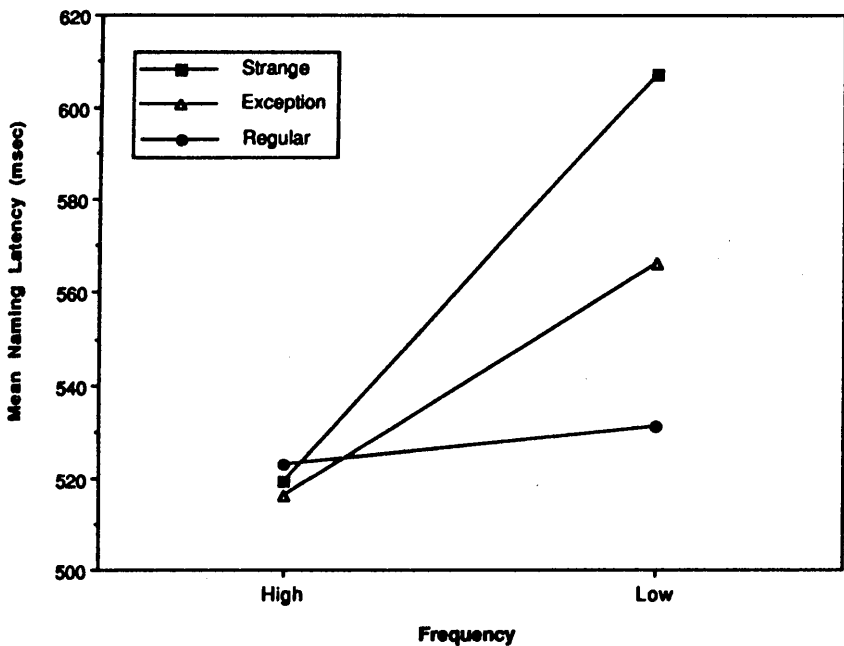


Figure 10.7  
Results of the Waters and Seidenberg 1985 study: experiment (upper graph) and simulation (lower graph).

Table 10.4  
Examples of stimuli used in nonword simulations

Type	Example	Pronunciation	Status of pronunciation
<b>Exception words and derived nonwords</b>			
HF word	<i>have</i>	/hav/	correct exception
	<i>have</i>	/hAv/	regularization error
LF word	<i>lose</i>	/Uz/	correct exception
	<i>lose</i>	/Oz/	regularization error
HF nonword	<i>bave</i>	/bAv/	regular pronunciation
	<i>bave</i>	/bav/	"exception" pronunciation
LF nonword	<i>yose</i>	/yOz/	regular pronunciation
	<i>yose</i>	/yUz/	"exception" pronunciation
<b>Regular words and derived nonwords</b>			
HF word	<i>came</i>	/kAm/	correct regular
	<i>came</i>	/kam/	incorrect regular
LF word	<i>deed</i>	/dEd/	correct regular
	<i>deed</i>	/ded/	incorrect regular
HF nonword	<i>pame</i>	/pAm/	regular pronunciation
	<i>pame</i>	/pam/	incorrect regular
LF nonword	<i>meed</i>	/mEd/	regular pronunciation
	<i>meed</i>	/med/	incorrect regular

Pronunciation key: a = a in *bat*; A = ai in *bait*; e = e in *bet*; E = ee in *beet*; O = o in *rope*; U = oo in *boot*. HF = high frequency; LF = low frequency.

be pronounced in two ways, either by analogy to the exception word (*mave* pronounced to rhyme with *have*) or by analogy to regular inconsistent word (*mave* rhymed with *gave*). Using the weights from 250 epochs, the model was tested to determine which pronunciation would be preferred. For each item, two phonological error scores were calculated: one using the regular pronunciation as target and one using the exception pronunciation as target. We calculated analogous scores for alternative pronunciations of the exception words themselves, e.g., *have* pronounced correctly and pronounced to rhyme with *gave*. This regularization error is sometimes produced by young children and by surface dyslexics (Patterson et al., 1985). Finally, we examined the pronunciations of regular words and nonwords derived from them. Here the alternatives were the "regular"



pronunciation (*came* pronounced correctly; *pame* rhymed with *came*) or a plausible incorrect pronunciation (*came* or *pame* pronounced with a short *a*).

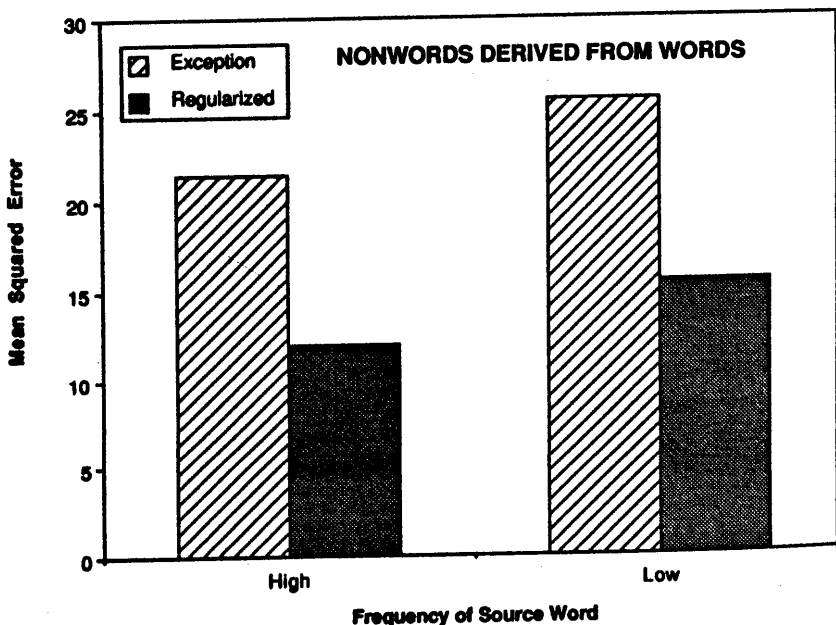
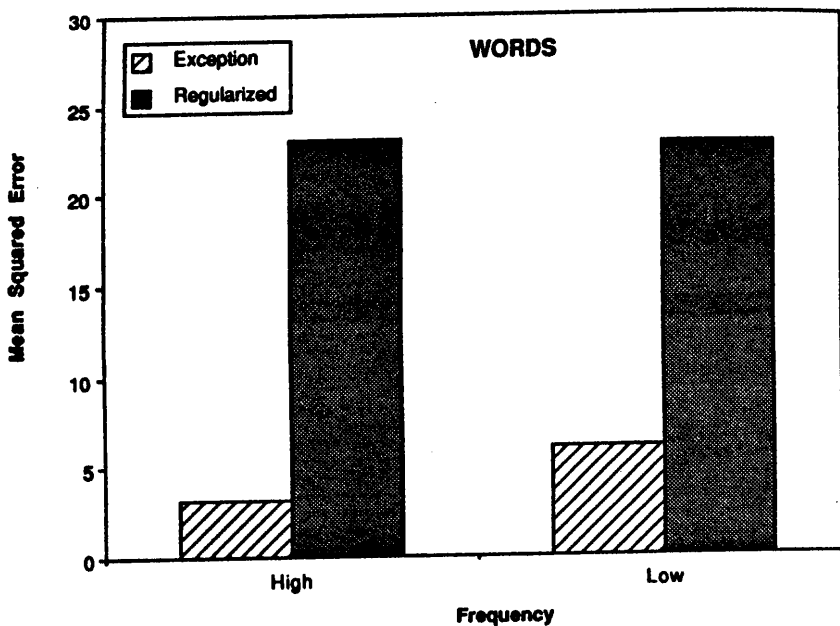
Figure 10.8 presents the mean error scores for the alternative pronunciations of exception words and derived nonwords. For words (upper graph), the correct exceptional pronunciations produce much smaller error scores than the incorrect, regularized pronunciations. Thus, the model's output resembles the patterns associated with the correct pronunciations rather than the regularized pronunciations. This occurs even though the model is exposed to many more words containing the regular pronunciation; that is, it computes correct output for *have* even though it is trained on *gave*, *save*, *pave*, *rave*, etc.

The opposite pattern obtains with the nonword stimuli derived from these words (figure 10.8, lower graph). Here the regular pronunciations are preferred to the pronunciations derived from the matched exception words. Note, however, that the difference between the two pronunciations is much smaller than in the corresponding word data, which suggests that the pronunciation of *mave* is influenced by the fact that the model has been trained on *have*.

Figure 10.9 presents the error scores for the regular pronunciations of nonwords derived from regular and exception words. The error scores are larger for nonwords such as *mave* (derived from an exception word) than *pame* (derived from a regular word). The results again indicate that the pronunciation of novel stimuli such as *mave* is affected by the fact that the model has been trained on both *have* and regular words such as *gave*.

In sum, spelling patterns such as *-ave* are associated with two pronunciations in the lexicon. The model produces output corresponding to the correct pronunciations of exception words such as *have* and regular inconsistent words such as *gave*. As noted above, regular inconsistent words are little affected by training on the corresponding exception word. When presented with novel stimuli containing these ambiguous spelling patterns, the model produces output that corresponds more closely to the regularized pronunciation; however, this output is affected by the fact that the model has been trained on exemplars of both pronunciations. In effect, the model has encoded the fact that *-ave* is typically pronounced as in *gave*; however, its knowledge of exception words such as *have* also influences the computed output.

We have not completed the analysis of the model's performance on novel stimuli. It appears, however, that it produces plausible output. Nonwords containing common spelling patterns with regular pronunciations (e.g., *nust*, *bine*) produce output that specifies a particular



**Figure 10.8**  
 The model's performance on exception words (top graph) and derived nonwords (bottom graph) from Taraban and McClelland 1987. "Exception" pronunciations of nonwords rhymed with exception words (e.g., *mave* is pronounced like *have*); "regularized" pronunciations of nonwords rhymed with regular, inconsistent words (e.g., *mave* is pronounced like *gave*).

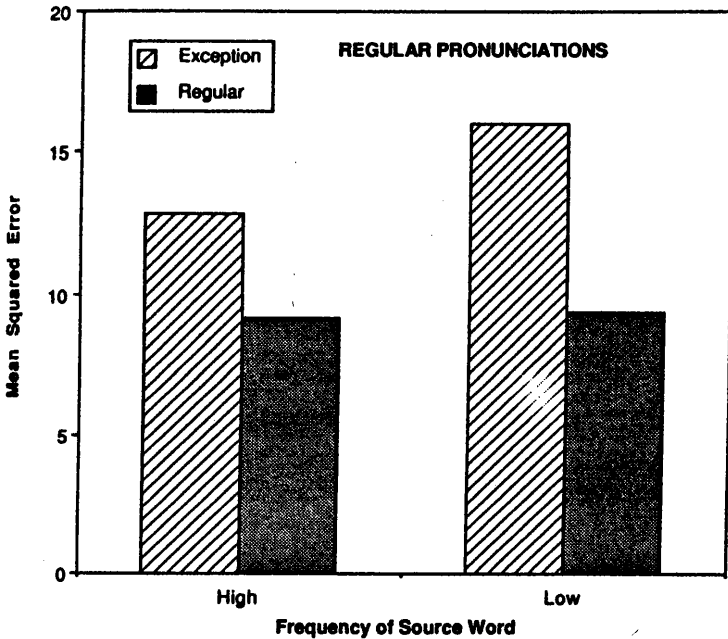


Figure 10.9  
Error scores for regular pronunciations of regular and exception nonwords.

pronunciation. Some of these items actually produce smaller error scores than rare words containing unusual spelling patterns (e.g., *tryst*, *fugue*), leading to the prediction that they should be easier to pronounce. Other nonwords (e.g., those containing unusual spelling patterns or patterns associated with multiple pronunciations) produce output that does not clearly specify a single pronunciation. When human subjects are required to pronounce such nonwords, they may rely on other strategies for formulating a response, such as using explicit analogies to known words or pronunciation rules. In effect, when the computed output is ambiguous, subjects who are nonetheless required to respond may employ other strategies.

**Summary of the naming simulations** The model captures basic facts about naming and the acquisition of the naming skill. Both human subjects and the model compute the pronunciations of different types of higher-frequency words about equally well. For less common words, irregular and inconsistent spelling-sound correspondences produce longer naming latencies and larger phonological error scores than regular words. The worst case is that of the strange items, which

contain unusual spelling patterns. The exception words fare somewhat better, because they contain spelling patterns that recurred in other training stimuli. The model produces the correct output, but with a higher error score than words whose pronunciations are entirely regular. During the learning process the model goes through intermediate stages that closely correspond to the behavior of children learning to read. Finally, the model also performs well in terms of the pronunciation of novel stimuli.

The model simulates a range of empirical phenomena concerning the pronunciations of words and nonwords. Why the model yields this performance can be understood in terms of the effects of training on the set of weights. The values of the weights reflect the aggregate effects of many individual learning trials using the items in the training set. In effect, learning results in the re-creation of significant aspects of the structure of written English within the network. Because the entire set of weights is used in computing the phonological codes for all words, and because all the weights are updated on every learning trial, there is a sense in which the output for a given word is a function of training on all words in the set. Differences between words derive from facts about the writing system distilled during the learning phase. For a given item the magnitude of the effect of training on other words varies as a function of their similarity to the item. For words the main influence on the phonological output is the number of times the model was exposed to the word itself. Number of times the model was exposed to closely related words (e.g., similarly spelled items) exerts secondary effects; there are also small effects due to exposure to other words.

### *Orthographic Output and Lexical Decision*

We turn now to the other type of output computed by the model. Orthographic output represents the retention or recycling of information about the literal form of the input. We assume that this information is relevant to visual-perceptual aspects of reading that have been widely studied by psychologists, including the word-superiority effect (Wheeler 1970; Reicher 1969) and other tachistoscopic-recognition phenomena such as feature-integration errors (Treisman and Schmidt 1982; Seidenberg 1987). This part of the model addresses the phenomena that motivated the interactive-activation model of McClelland and Rumelhart (1981). The computed orthographic code also may play an important role in some forms of deep dyslexia in which patients access semantic information but are unable to report the identities of words. Orthographic output also plays a critical role in

the task that has been most widely employed in studies of word recognition, lexical decision. We shall first describe the computed orthographic output for different types of words and then present an account of the lexical-decision task.

The orthographic output for regular and exception words can be described simply: there are effects of frequency and amount of training but not of word class. The model performs better on high-frequency words than on low-frequency words, and error scores again decrease in an approximately logarithmic manner. However, at no stage in training are there consistent differences between regular and exception words. The absence of any effect of word class confirms the assumption that the regular and exception words differ only in regard to orthographic-phonological regularity, not in regard to orthographic redundancy. The data comparing the orthographic output for regular inconsistent and regular words show a similar pattern: approximately logarithmic learning curves, an effect of frequency, and no effect of word type. Strange words act somewhat differently, however. Again, there were overall effects of frequency, and the model's performance improved with additional training. Through the first 50 epochs the model performs poorly on both high- and low-frequency strange words. Through the next 200 epochs error scores for the high-frequency strange items gradually approached those for high-frequency regular words. But the lower-frequency strange items show much smaller gains: even after 250 epochs they produce the largest error scores. In sum, the regular, regular inconsistent, and exception words act alike in terms of orthographic redundancy; for these items, there are effects of frequency and experience, but not of word class. The strange words differ from the others in terms of orthographic redundancy, as Waters and Seidenberg (1985) suggested. Performance on the higher-frequency strange items improves to the point where they produce output that differs only slightly from that of the regular words; the lower-frequency strange items yield the poorest performance.

When considering both the orthographic and phonological data, one can see that the effects of both orthographic-phonological regularity and orthographic redundancy are largely restricted to words that occur relatively rarely in the language. Whereas lower-frequency exception words produce larger phonological error scores than lower-frequency regular words, these items do not differ in terms of orthographic output. Lower-frequency strange words produce the poorest performance in terms of orthography, because the computation of orthographic output is affected by facts about orthographic redundancy. These words also produce the poorest performance in terms of pho-

nology, because they contain unusual spelling patterns with idiosyncratic pronunciations.

It follows from this analysis that if subjects were able to perform a word-recognition task by addressing only the orthographic output, the results would differ from those obtained with naming. Specifically, there should be no effects of irregular spelling-sound correspondences on such a task, only effects of orthographic redundancy. The lexical-decision task, to which we now turn, yields results consistent with this analysis.

*Lexical decisions* A number of studies have examined the effects of orthographic redundancy and orthographic-phonological correspondences on lexical decisions. The task requires subjects to decide whether a string of letters is a word or not. The key finding from this literature is that subjects' criteria for making this decision vary depending on the kind of stimuli included in an experiment. The model provides a simple account of these varying decision criteria. Recall the Waters and Seidenberg (1985) results discussed above (table 10.2). The naming task produced the familiar frequency-regularity interaction: for higher-frequency stimuli, regular and exception words produced similar naming latencies; for lower frequency stimuli, exception words produced longer latencies than regular. This result obtained whether the stimulus set included strange words or not. In the lexical-decision task the exception effect for lower-frequency words depended upon the presence of strange words in the stimulus list. When the strange words were deleted, no effect of phonological regularity obtained.

Waters and Seidenberg (1985) proposed the following account of these results. When the stimuli consist of regular and exception words and pronounceable nonwords, subjects base their decisions on the results of orthographic analyses. Because the decisions are based on orthographic information, no effects of phonological regularity obtain. Including the strange stimuli increases the difficulty of word/nonword discrimination, however. Subjects are asked to respond "word" when they see an item with an unfamiliar spelling pattern such as *tryst* and to respond "nonword" when they encounter stimuli that contain common spelling patterns but are nonetheless not words (e.g., *rone*). Making this discrimination on the basis of orthographic information is difficult, so subjects change their response strategy: they use phonological information as an additional basis for their decisions. In effect, the subject responds "word" if the stimulus has a familiar pronunciation, and "nonword" if it does not. Under these conditions, the task is much like naming: it requires computing the

phonological code. Thus, results are similar to those in naming, with a regularity effect for lower-frequency words.

Subjects thus vary the criteria by which lexical decisions are made. In contrast, subjects cannot vary their response strategies when the task is to name words aloud. The task does not involve discriminating between words and nonwords; rather, it requires the subject to produce the correct pronunciation, which cannot be accomplished until the phonological code has been computed. The empirical results agree with this analysis; as Waters and Seidenberg (1985) observed, effects of phonological regularity in the naming task do not depend upon the inclusion of strange words.

Consider now how the model accounts for this pattern of results. After the model has been trained, it can be tested on familiar words and on nonwords. These items produce orthographic output, which can be summarized in terms of error scores. In general, the error scores are smaller for words than nonwords; the model performs better on familiar stimuli. However, the magnitudes of these scores vary as a function of factors such as orthographic redundancy and length. The word and nonword stimuli in an experiment will yield distributions of orthographic error scores. If the word stimuli include regular and exception words (items that do not differ in terms of their orthographic properties) and pronounceable nonwords such as *rone* or *bist*, the word and nonword distributions will overlap very little. Lexical decisions can be modeled as the establishment of decision criteria that operate on the orthographic-error scores. As in a signal detection paradigm, the subject can establish a decision criterion such that scores below the cutoff are judged words and those above it, nonwords. Latencies should be a function of distance from the cutoff; words producing very small error scores and nonwords producing very large error scores should yield faster responses than stimuli whose scores are closer to the cutoff. Given these distributions, response criteria can be established that yield low error rates in the range observed in actual experiments.

The effect of including strange items in the stimulus set is to yield distributions of word and nonword scores with greater overlap; many of the lower-frequency strange items produce error scores as high or higher than those for nonwords. Low-frequency strange words such as *tryst* or *fugue* have been encountered very rarely, and they contain spelling patterns that do not recur in other items. Nonwords such as *bist* or *rone* are also encountered very rarely (typically never), but they contain spelling pattern that recur in many other words. In many experiments the nonwords are also somewhat shorter than the word stimuli, which will tend to produce smaller error scores. Hence, the

error scores for some lower-frequency strange words will be larger than those for some nonwords.

Thus, when strange words are included in the stimulus set, subjects can no longer establish response criteria that yield acceptably low error rates. Lexical decisions cannot be based on the discriminability of words and nonwords in terms of orthographic output. Under these circumstances, subjects must utilize other information in making their decisions. One strategy that subjects apparently use is to consult the results of phonological processing, i.e., phonological output. If subjects begin to consult phonological output, their response latencies should be like those obtained in the naming task, which is also based on this output. The empirical prediction, then, is quite clear: if the strange words are excluded, responses are based on orthographic output, with no difference between regular and exception words, only a frequency effect. If strange words are included, blocking responses based on orthographic output alone, there should be longer latencies for lower-frequency exception words than for lower-frequency regular words, because this pattern is observed in the phonological output, which now provides the basis for making the lexical decision. These are exactly the outcomes observed in the Waters and Seidenberg (1985) experiments.

We tested the model on the Waters and Seidenberg word and nonword stimuli, using the weights from 250 learning epochs. Figure 10.10 (top) presents the data for the condition in which the stimuli consist of high- and low-frequency regular and exception words. Figure 10.10 (middle) shows the corresponding data for the nonword stimuli. The distributions of orthographic-error scores are such that a decision criterion can be established that yields a low error rate similar to that observed in the actual experiment. Since the regular and exception words do not differ in terms of orthographic output, no regularity effect is observed. Figure 10.10 (bottom) presents the data for regular, exception, and strange items. There is considerable overlap between the word and nonword distributions, which makes it impossible to establish a decision criterion that yields a low error rate. Under these circumstances, we argue, subjects begin to look to phonological output. Decision latencies now exhibit the pattern associated with the naming task, longer latencies for lower-frequency exception words than regular words.

In sum, the model provides a simple account of observed differences between lexical decision and naming performance. The naming task requires the subject to compute the phonological code for a word. Under most conditions the lexical-decision task can be performed on the basis of the output of orthographic analysis. In these



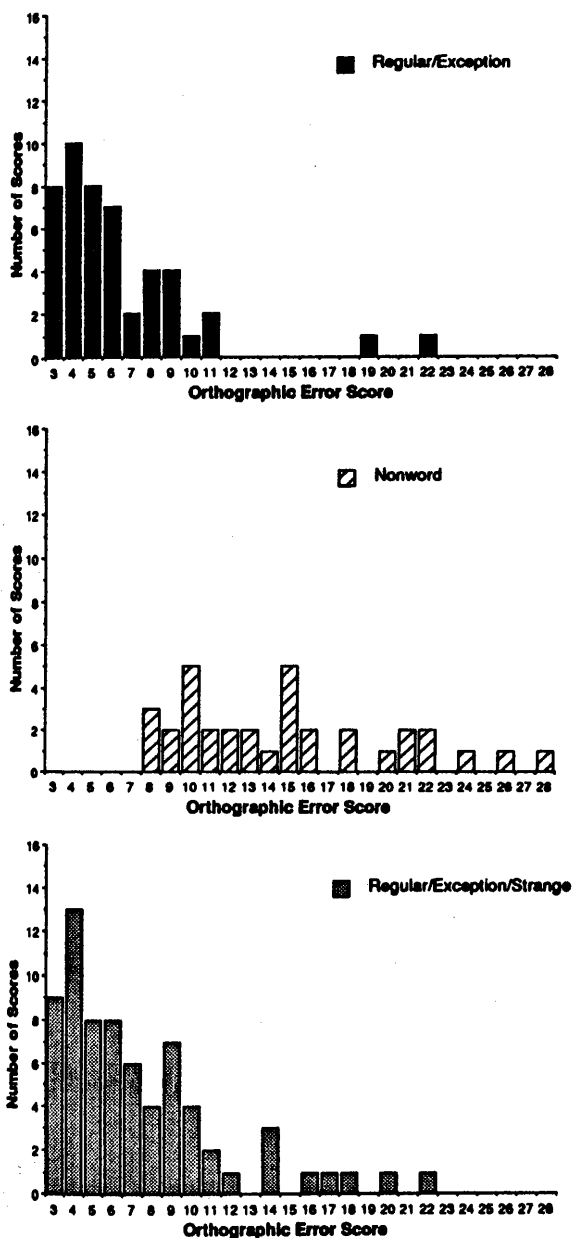


Figure 10.10 Simulation of the Waters and Seidenberg 1985 studies. Orthographic error scores for regular and exception words (top graph); pronounceable nonwords (middle graph); regular, exception and strange words (bottom graph).

cases naming produces effects of phonological regularity for lower-frequency words, but lexical decision does not. Strange words produce longer latencies on both tasks, in lexical decision because they contain unusual spelling patterns and in naming because they are difficult to pronounce. However, if the stimuli in a lexical-decision experiment include very wordlike nonwords and very nonwordlike words, subjects base their decisions on computed phonological codes. Under these conditions lexical decision results are like those obtained in naming, because both responses are based on the same information.

This model of the lexical-decision process will account for other findings in the literature, including the frequency blocking effects observed by Gordon (1983) and Glanzer and Ehrenreich (1979), and the orthographic and phonological priming effects studied by Meyer, Schvaneveldt, and Ruddy (1974), Hillinger (1980), and others (see Seidenberg and McClelland, in press).

*Homographs* Further evidence in support of this analysis of lexical decision and naming is provided by the data for homographs, words like *wind* or *lead* with two pronunciations, each of which corresponds to a different word in English. Seidenberg et al. (1984) found that homographs produce longer naming latencies than regular words; however, latencies for the two types of words did not differ on the lexical-decision task. The corpus used in training the model included 11 homographs. Each homograph was presented with both correct pronunciations: on one trial, for example, the model was given the feedback that the correct pronunciation of *lead* rhymes with *bead*; on another trial it was given the feedback that the pronunciation rhymes with *dead*. The orthographic output of the homographs did not differ from that of regular words matched in terms of frequency and length. However, homographs produce very large phonological-error scores compared to regular words. This occurs, of course, because the model has been exposed to two pronunciations of these words. In the naming task longer latencies result because subjects must choose between two computed pronunciations. Note that a very similar account applies to the naming of strange words (such as *gauge* or *caste*), whose pronunciations are not known with certainty (as discussed above). Readers may encode different pronunciations on different encounters; that is what it means to say that they are uncertain about their pronunciations. Consideration of two different pronunciations for such words makes them effectively act like homographs, yielding long naming latencies. Because only one pronunciation is actually correct, they also produce many errors.

### *Dyslexia*

We have seen that the model provides an account of a broad range of phenomena related to visual word recognition and naming. (The model also accounts for a number of additional phenomena we have not described; see Seidenberg and McClelland, in press, for details). As a learning model, it also speaks to the issue of how these skills are acquired; moreover, it provides an interesting perspective on the kinds of impairments characteristic of developmental and acquired dyslexias. Developmental dyslexia can be seen as a failure to acquire the knowledge that underlies word recognition and naming. Acquired dyslexias naturally correspond to impairments following damage to the normal system. The model clearly suggests a number of possible sources of processing impairment, which we are in the process of fully exploring. In this section we shall summarize our initial efforts, which have focused on naming impairments. We chose this focus for two reasons. First, it is known that acquiring knowledge of spelling-sound correspondences is a key component of learning to read. Disorders in phonological-analysis skills are thought to be a primary source of reading disability, and children who are backward readers (Backman et al. 1984) or developmental dyslexics (Seidenberg et al. 1986) perform relatively poorly in naming words and nonwords aloud. Second, much of the research on various types of acquired dyslexia has been concerned with overt naming. Disorders such as phonological and surface dyslexia (Patterson 1981) are largely associated with different kinds of naming errors. Hence, naming disorders seemed a good place to start.

### *Training with Fewer Hidden Units*

Consider first the results of an experiment in which we retrained the model with half as many hidden units, 100 instead of 200. In all other respects the training procedure was the same as before. The model was initialized with small random weights on the connections between units; the weights were modified during a training phase using the back-propagation learning algorithm. Figure 10.11 gives the mean phonological-error scores for regular and exception words in the test set described above using 100 hidden units. The figure can be compared to the results for the same stimuli using 200 hidden units (presented in figure 10.3). Two main results can be observed. First, from epochs 10 to 250, training with fewer hidden units yields poorer performance for all word types. High-frequency regular words, for example, asymptote at a mean squared error of about 4.8 in the 100-unit model and at about 3 in the 200-unit model; other words yield similar

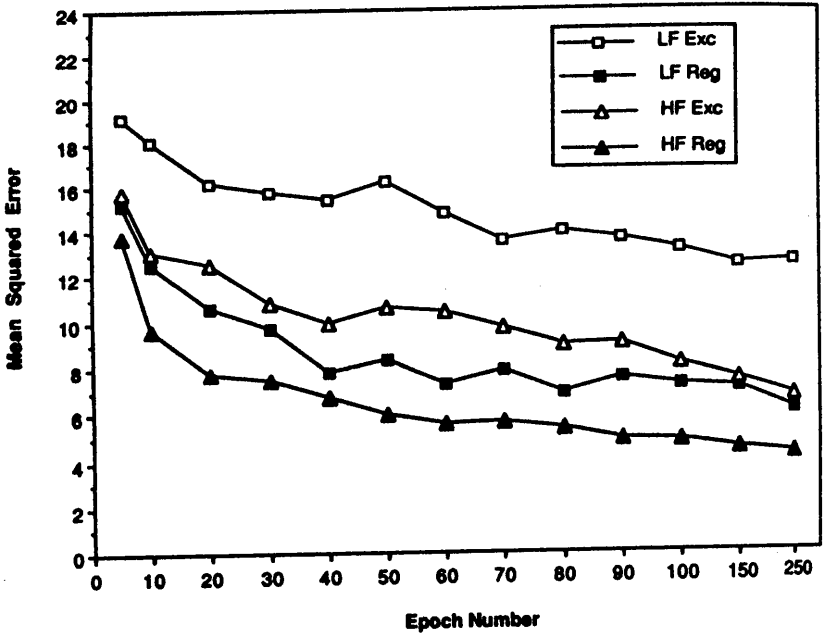


Figure 10.11

The model's performance on regular and exception words from Taraban and McClelland 1987; 100 hidden units.

results. Second, exception words produce significantly poorer output than regular words in both the high- and low-frequency ranges in the 100-unit model even after 250 epochs of training; in the 200-unit model, exception words produce larger error scores only in the lower-frequency range.

As previously noted, one of the primary developmental trends observed in studies such as Backman et al. (1984) is that younger children and poorer readers show larger regularity effects; that is, they perform worse on exception words (and other items with inconsistent spelling-sound correspondences) than on regular words. For most children, this difference between word classes is eliminated for the most common words in their vocabularies by about grade 5. Even among skilled adult readers, however, exception words continue to produce longer naming latencies and more errors than regular words in the lower-frequency range.

Both poor readers who are reading below age-expected levels and children who have been diagnosed as developmental dyslexics fail to show this improvement in naming higher-frequency exception words. We obtained similar results by eliminating half the hidden

units. Like the dyslexic child, the model showed large exception effects for both high- and low-frequency words; similar results were observed for regular, inconsistent and strange words. With 200 hidden units the phonological output for regular inconsistent words did not differ from that for regular words in both high- and low-frequency ranges. With 100 hidden units, lower-frequency regular inconsistent words produced larger error scores than matched regular words. Similarly, training with 200 hidden units eventually produced similar output for high-frequency regular and strange items, but larger error scores for lower-frequency strange items in comparison with regular items. With 100 hidden units the model performs poorest on high- and low-frequency strange words; unlike the 200-unit data, scores for the high-frequency strange words did not converge on those for higher-frequency regular words.

Eliminating half the hidden units, then, produces a general decrement in performance; more importantly, *higher* frequency words produce the patterns associated with *lower* frequency words in the 200-unit simulation, i.e., larger error scores for exception and strange words in comparison to regular words. Even with fewer hidden units the model continues to encode generalizations about the correspondences between spelling and pronunciation; error scores are smaller for regular words than for other types. However, it performs more poorly on words whose pronunciations are not entirely regular. Apparently, including fewer hidden units makes it more difficult to encode item-specific information about pronunciation. Note also that because of this limitation the difference between high-frequency regular and exception words probably would not be completely eliminated with additional training.

These results capture a key feature of the data obtained in studies of dyslexic readers. These children continue to perform poorly in naming even higher-frequency exception words. At the same time, their performance shows that they have learned some generalizations about spelling-sound correspondences; for example, they are able to pronounce many nonwords correctly (Seidenberg et al. 1986). One of the main hallmarks of learning to read English is acquiring knowledge of spelling-sound regularities. Dyslexic readers achieve some success in this regard but cope poorly with the irregular cases. The model performs in a similar manner with too few hidden units; with the resources available it is able to capture crude generalizations about regularity but at the expense of the exception words. The main implication of the simulation, of course, is that children who are not achieving age-expected reading skills may be limited in the computational resources they have available for the task. There is also another

important implication. Apparently, the architecture of the model determines in an important way its ability to behave like humans. If there are too few units, the model can learn generalizations about the regularities in the writing system; however, it does not have the capacity to encode enough of the word-specific information relevant to exception words to perform as well as people. With a sufficient number of units, it is able to cope with both regular and irregular cases, although not equally well on all items. The important point is that human performance seems to reflect rather subtle constraints concerning computational resources. The idea that impaired performance might result from dedicating too few resources to a task is one that could be pursued in future research.

### *Lesioning the Model*

We turn now to a brief discussion of some forms of dyslexia observed following brain injury. Numerous studies have shown that brain injury often results in selective impairment of word-recognition and naming processes (see papers in Patterson, Marshall, and Coltheart 1985 for reviews). Case studies of these acquired forms of dyslexia have provided important clues to the brain bases of reading behavior, as well as a source of evidence bearing on the development of theories of reading. Much of what we know about these patients concerns their impaired performance in reading words aloud. Since our model provides an account of the types of knowledge and processes used in naming, it is reasonable to ask whether damage to this system would yield performance similar to that of different types of dyslexic patients. In this section we describe some experiments that address this question; this work is described in greater detail in Patterson, Seidenberg, and McClelland (in press). We should stress the preliminary nature of these investigations. We present them here primarily to suggest that the model provides an interesting and potentially illuminating perspective on acquired dyslexias.

There are many different ways the model could be "lesioned," possibly yielding very different types of impaired performance. Our initial investigations focused on the main types of errors characteristic of surface (Marshall and Newcombe 1973; Patterson, Marshall, and Coltheart 1985) and phonological (Beauvois and Derouesné 1979; Shallice and Warrington 1975) dyslexias. We were interested in these types of dyslexia because they have been seen as providing evidence for a class of reading models very different from ours (the dual-route model of Coltheart [1978] and variations thereon). We should acknowledge from the outset that categories such as surface and phonological dyslexia are based on characteristic patterns of errors, and it

is widely acknowledged that a given error pattern could arise from more than one type of pathology. Indeed, the performance of patients within each of these categories varies considerably. These observations suggest that it is unlikely that each type of dyslexia is associated with a unique type of damage.

Since it will probably take quite a while to examine all of the implications of the model concerning reading impairments, we began these investigations with a simpler question: Could the model be damaged in such a way as to produce the main types of errors characteristic of surface and phonological dyslexias? If the model simply cannot produce these types of errors, it would indicate that some rethinking was in order; if the model can produce these types of errors, it would provide a source of evidence corroborating the studies of normals and would motivate detailed studies of more subtle aspects of dyslexic performance (concerning, for example, the proportions of errors of different types or particular words that yield errors).

The basic characteristics of these forms of dyslexia are as follows. Differences among patients can be observed by examining their abilities to read regular and exception words and nonwords. Surface dyslexics are able to name regular words and nonwords but show impairments in naming words with irregular or inconsistent spelling-sound correspondences. They make a disproportionate number of errors on the latter types of words, producing characteristic mispronunciations such as *have* → /hAv/, *pint* → /pint/, and *flood* → /flOd/. The first two of these errors are regularization errors; the patient pronounces the word as it would be pronounced if the spelling-sound correspondence were in fact regular. The third example represents another common type of error, in which the pronunciation is phonologically related to the correct pronunciation but is not a strict regularization. In contrast, phonological dyslexics correctly name regular and exception words, but show an impairment in reading nonwords. In some dramatic cases the patient reads words correctly and, when confronted with a nonword such as *nust*, either makes no response (other than "I don't know") or produces a lexicalized response (e.g., *nust* → /must/, or *pind* → /pink/). That is, they respond by naming a word that is orthographically and phonologically similar to the nonword target. In terms of dual-route models of reading (Patterson et al. 1985), surface dyslexia is thought to derive from impairment of the "addressed" or "lexical" naming mechanism, whereas phonological dyslexia derives from impairment of the "assembled" or "sublexical" mechanism. Clearly, since our model admits only one naming route (in the terminology of the earlier model, it suggests that all pronunciations are "assembled"), it would be a challenge to show that it

could produce these distinct patterns of impairment. Presumably these different patterns could be a consequence of damage to different components of the network rather than distinct pronunciation "routes."

According to the standard interpretation of surface dyslexia (within the dual-route model), exception words are read by accessing representations of their pronunciations stored in a phonological lexicon. Errors result because the entries for some words have been damaged. The patient then uses another type of knowledge (rules governing spelling-sound correspondences) to pronounce the exception word, resulting in a regularization error or other systematic mispronunciation. Our model differs from the dual-route conception in two respects. First, there is no phonological lexicon in which the pronunciations of words are stored; rather, pronunciations are computed on the basis of knowledge of spelling-sound correlations encoded by the weights on connections. Second, there are no separate rules governing spelling-sound correspondences. Our hypothesis is that the impaired performance of at least some surface dyslexics results from damage to the lexical network that leaves the processing of regular words and nonwords relatively intact but results in errors on exception words. The basic idea is that the regular spelling-sound correspondences are more robustly encoded by the network than the irregular ones. Damage to the system (e.g., loss of units or connections) is expected to affect performance on all types of stimuli but with greater impact on the exceptions. This produces the characteristic types of errors.

Patterson, Seidenberg, and McClelland (in press) examined this possibility in the following way. We take the weights that were created after 250 epochs of training; these are the weights that provide the basis for our simulations of skilled performance. We can then damage the system in different ways and retest the model's performance on different types of items. As before, we characterize the model's performance on a given word by calculating error scores that index the fit between the computed phonological output and the correct patterns associated with different pronunciations. For example, we compare the computed output for *have* to the pattern for the correct pronunciation /hav/, the regularized pronunciation /hAv/, and other potential pronunciations. In the undamaged mode the best fit to the computed output is almost invariably the correct phonological code. The errors characteristic of surface dyslexia correspond to cases in which the best fit is provided by something other than the correct pronunciation. A regularization error, for example, corresponds to



the case in which the best fit to the computed output is provided by the regularized pronunciation.

The main result of these damage simulations can be summarized as follows. We examined several types of damage to the network, including eliminating connections from input units to hidden units, eliminating connections from hidden units to phonological output units, and eliminating the output from hidden units. We also varied the degree of damage (that is, the proportion of connections eliminated). In all cases damage had a bigger impact on exception words and nonwords than on regular words. Although performance on regular words declines as a consequence of damage (i.e., they produce larger error scores), it is rarely the case that the best fit to the computed output is something other than the regular pronunciation. Exception words fare differently. For a significant proportion of these items (the exact proportion depends on the degree of damage), the best fit to the computed output is not the correct, exceptional pronunciation. In most of these cases the best fit is provided by the regularized pronunciation; for a word such as *deaf*, for example, the computed output more closely approximates /dEɪf/ than /def/. In other cases the best fit is provided by a pronunciation that is systematically related to the correct one but is not a strict regularization. For example, the output for *pint* might correspond to the pronunciation of *pant*.

The simulations indicate that damage to the model can produce the types of errors characteristic of surface dyslexia. This demonstration is suggestive insofar as it appears that accounting for these errors does not require appeal to separate lexical and nonlexical naming mechanisms. To move beyond this demonstration it will be necessary to consider more detailed aspects of surface dyslexics' performance, aspects concerning, for example, the proportions of errors of different types, the latencies of naming responses, and performance on other types of words and on nonwords. This will require addressing the very important differences among patients who have been classified as surface dyslexics. For example, though both were considered to be surface dyslexics, the patients studied by Bub, Cancelliere, and Kertesz (1985) and Shallice, Warrington, and McCarthy (1983) performed in very different ways. The patient studied by Bub et al. made a large number of regularization errors, with a larger proportion of errors among lower-frequency words than among high-frequency words. Moreover her latencies did not differ greatly from those of normal subjects. The patient studied by Shallice et al. produced many errors that were not regularizations, and the errors were not frequency sen-

sitive. It seems unlikely that such different patterns of performance could derive from a common source.

Similar issues arise in connection with phonological dyslexia. We have observed that damage to the system can produce lexicalization errors. We damage the system and test the model's performance on nonwords such as *nust*. In the undamaged state the model produces plausible output for many regular nonwords, with the best fit to the computed output provided by the correct phonological code. A lexicalization error results when the best fit to the output computed by the damaged system is an orthographically and phonologically similar word. The output for *nust*, for example, is closer to /must/ than to /nust/. Damage to the model can produce a number of these errors, but it remains to be determined whether we can obtain a pattern of performance that closely fits the profiles of phonological dyslexic patients.

In sum, because the model in its undamaged mode closely simulates a broad range of behaviors, we expected that damage to the system would produce impaired performance of a recognizable sort. Our preliminary experiments indicate that this is indeed the case. Obviously many questions remain to be addressed, and it is a daunting challenge to accommodate the entire range of phenomena associated with the acquired dyslexias.

### *Conclusions*

We have described a model of lexical processing that illustrates how systematic, "rule-governed" behavior can emerge from a network of simple processing units. According to this account, lexical processing involves computing several types of information—orthographic, phonological, and semantic—in parallel. We have described the computation of the orthographic and phonological codes in some detail and have shown that the model provides a quantitative account of various behavioral phenomena. The model accounts for differences among words in terms of processing difficulty, differences in reading skill, and the course of acquisition and points to some plausible bases for acquired and developmental forms of dyslexia. We characterized lexical decision and naming in terms of how the computed codes are utilized in making these types of responses. A task such as naming focuses on the use of one type of code, phonology; a task such as lexical decision may involve all of the codes. The same types of knowledge representations and processes are involved in the computation of all three codes (although the implemented model is restricted to orthography and phonology). Knowledge is represented

by the weights on connections between units. These weights are primarily determined by the nature of the English orthography that acts as input, in conjunction with feedback during the learning phase. Our claim is that representing knowledge of the orthography in this way is felicitous because of the quasiregular nature of the system; the characteristics of English orthography are more congruent with this type of knowledge representation than with the kinds of pronunciation rules that have previously been proposed. The computation of the orthographic code is affected by the distribution of letter patterns in the lexicon; computation of the phonological code is affected by correlations between orthography and phonology.

The model differs from previous accounts in several respects. First, it makes quantitative, testable predictions about the processing of individual words as well as general types of words. This is simply a consequence of the explicit nature of the knowledge representations and processes, which had to be achieved in order to develop a working simulation model. Second, it differs from previous accounts in terms of the types of knowledge and processes underlying performance. In contrast to the dual-route model, there are no rules specifying the regular spelling-sound correspondences of the language, and there is no phonological lexicon in which the pronunciations of all words are listed. All items—regular and irregular, word and nonword—are pronounced using the knowledge encoded by a single set of connections. The main assumption of the dual-route model is that separate mechanisms are required to account for the capacity to name exception words and nonwords. Exception words cannot be pronounced by rule, only by consulting a stored lexical entry; hence one route is termed "lexical" or "addressed" phonology. Nonwords do not have lexical entries; hence, they can only be pronounced by rule. Accordingly, the second route is termed the "nonlexical" or "subword" process. One of the main contributions of the network model is that it demonstrates that pronunciation of exception words and nonwords can be accomplished by a single mechanism employing weighted connections between units. Our model also differs from proposals by Glushko (1979) and Brown (1987) in that there are no lexical nodes representing individual words and no feedback from neighbors. In fact, phonological output is computed on a single, forward pass through the network, giving the model a very different character from these other accounts, in which behavioral phenomena result from complex interactions among partially activated words and letters.

The model reflects a very basic change in perspective on issues concerning word recognition. The models described by Coltheart (1987b), Monsell (1987), and others contain multiple lexicons, includ-

ing separate orthographic lexicons used in reading and writing and separate phonological lexicons used in listening and speaking. Research within this framework has focused on questions concerning what has been termed lexical access: how the entries for different codes are accessed in reading, the order in which they are accessed, and how access of one code affects access of other codes.

Our model departs from these precursors in a fundamental way: lexical memory does not consist of entries for individual words; there are no logogens (Morton 1969, this volume). Knowledge of words is embedded in a set of weights on connections between processing units encoding letters, phonemes, and the correlations between them. The spellings, pronunciations, and meanings of words are not listed in separate stores; hence, lexical processing does not involve accessing these stored codes. Rather, these codes are computed from the input string using the knowledge stored in the network structure, which results in the activation of distributed representations. Thus, the notion of lexical access does not play a central role in our model, because it is not congruent with the model's representational and processing assumptions.

The view that lexical processing involves *activation* of different types of information rather than *access* to stored lexical codes represents more than a change in terminology. The access view suggests that there is a moment in time at which a letter string is identified as a particular word, which provides immediate access to stored representations of its meanings. In our model, information simply accrues over time, and it would be arbitrary to designate a particular point in this process as the moment of "lexical access." More generally, access to a lexical code is an all-or-none phenomenon, whereas our framework sanctions the notion of partial or graded activation of lexical codes. In an activation model with distributed representations a code is represented as a pattern of activation across a set of nodes. The activations of the nodes can differ in strength. These conceptions raise different questions and generate different empirical predictions. For example, within the access framework it is meaningful to ask how many of the meanings of an ambiguous word are accessed; Swinney (1979) and Onifer and Swinney (1981) has proposed that lexical access make all the meanings of an ambiguous word available. In contrast, the activation approach suggests that meanings may be activated with different strengths; moreover, a network with distributed representations, such as ours, affords the possibility of partial activation of meaning (see Kawamoto 1987; McClelland and Kawamoto 1986; McClelland and Rumelhart 1985; Hinton and Sejnowski 1986). The latter view is consistent with behavioral evidence that different as-

pects of the meaning of a word—or different meanings entirely—are activated in different contexts (Barsalou 1982; Schwanenflugel and Shoben 1985; Tabossi 1988).

Similarly, within the lexical-access framework the well-studied question of phonological mediation concerned whether the phonological code was accessed before or after the semantic code ("prelexically" or "postlexically"). In our framework, phonology is one of several codes computed in parallel, and the primary question concerns how knowledge of the orthography affects this computation. Since the computation of the phonological code is orthogonal to the computation of meaning, questions as to whether frequency effects in naming are due to lexical access or production (Balota and Chumbley 1985; McCann and Besner 1987) do not arise.

In sum, the notion of lexical access carries with it a concern with certain types of theoretical questions. The primary questions concern the number of lexicons, how they are organized and linked, and whether it is orthographic or phonological information that provides access to meaning. The primary processing mechanism is search through one or more ordered lists. In our model the codes are distributed; they are computed on the basis of three orthogonal processes; and the primary processing mechanism is spread of activation. The primary theoretical questions concern the properties of these computations, which are determined by the properties of the writing system that are picked up by the learning algorithm on the basis of experience.

In its present state the model is limited in many respects. It deals only with monosyllabic words; there is no mechanism for actually pronouncing a word based on the computed phonological output; there is neither representation of meaning nor consideration of how contextual information could influence processing; we do not have a solid account of the acquired dyslexias (see Seidenberg and McClelland, *in press*, and Seidenberg, *in press*, for discussion). These are important limitations, and they mean that the model is by no means a complete account of lexical processing in reading. We see the model as having succeeded in addressing a number of basic issues that have preoccupied reading researchers for some time and hope that it represents a step toward developing a genuinely explanatory account of these important phenomena.

### Notes

This research was supported by the National Science Foundation (BNS 8609729), Office of Naval Research (N00014-86-G-0146), Natural Science and Engineering Research

Council (A7924), Quebec Ministry of Education (FCAR EQ-2074), and National Institutes of Health (HD-22271). McClelland was also supported by an NIH career development award (MH00385). We thank Karalyn Patterson, who has collaborated on studies of the implications of the model for acquired forms of dyslexia (Patterson, Seidenberg, and McClelland, in press) and is primarily responsible for the experiments on surface and phonological dyslexia. We should also acknowledge the important work by Sejnowski and Rosenberg (1986), whose NETtalk model was the first application of the algorithm of Rumelhart, Hinton, and Williams (1986) to the problem of learning the spelling-sound correspondences of English. Sejnowski and Rosenberg recognized that this knowledge could be represented within a parallel distributed network rather than as a set of pronunciation rules. Our goal was to explore the adequacy of this approach by developing a model that could be related to a broad range of behavioral phenomena.

1. There is another factor that probably contributed to the inconsistent results in studies comparing regular, inconsistent and regular words. Brown (1987) has noted that the number of times a spelling pattern occurs with a given pronunciation influences naming latency. For example, *-ust* is pronounced /ust/ in a certain number of words, which may differ from the number of times *-ist* is pronounced /ist/. Tests of the effects of the inconsistency of a spelling-sound pattern are valid only if this factor is controlled. Specifically, the number of times a regular, inconsistent pattern is assigned the regular pronunciation (e.g., *-ave* is pronounced /Av/) should be equal to the number of times the spelling pattern in a regular word is assigned the regular pronunciation (e.g., *-ust* is pronounced /ust/). With the stimuli equated in this way, any naming-latency differences could be safely attributed to the inconsistency factor. Studies of regular, inconsistent words did not in fact equate the stimuli in this way. When the stimuli are equated in this way, there is a small but reliable consistency effect for lower-frequency words (Seidenberg and McClelland, in press).

## References

- Andrews, S. 1982. Phonological recoding: Is the regularity effect consistent? *Memory and Cognition* 10:565-575.
- Backman, J., Bruck, M., Hebert, M., and Seidenberg, M. 1984. Acquisition and use of spelling-sound information in reading. *Journal of Experimental Child Psychology* 38:114-133.
- Balota, D. A., and Chumbley, J. I. 1985. The locus of the word frequency effect in the pronunciation task: Lexical access and/or production? *Journal of Memory and Language* 24:89-106.
- Barron, R. W. 1981. Reading skill and reading strategies. In A. M. Lesgold and C. A. Perfetti, eds., *Interactive processes in reading*. Hillsdale, N.J.: Erlbaum.
- Barsalou, L. W. 1982. Context-independent and context-dependent information in concepts. *Memory and Cognition* 10:82-93.
- Beauvois, M. F., and Derouesné, J. 1979. Phonological alexia: Three dissociations. *Journal of Neurology, Neurosurgery, and Psychiatry* 42:1115-1124.
- Beaner, D., Waller, T. G., and MacKinnon, G. E. 1985. *Reading research: Advances in theory and practice*. Vol. 5. New York: Academic Press.
- Brown, G. D. A. 1987. Resolving inconsistency: A computational model of word naming. *Journal of Memory and Language* 26:1-23.
- Bub, D., Cancelliere, A., and Kertesz, A. 1985. Whole-word and analytic translation of spelling to sound in a non-semantic reader. In K. E. Patterson, J. C. Marshall, and M. Coltheart, eds., *Surface dyslexia*. London: Erlbaum.

- Coltheart, M. 1978. Lexical access in simple reading tasks. In G. Underwood, ed., *Strategies of information processing*. New York: Academic Press.
- Coltheart, M., ed. 1987a. *Attention and Performance*. Vol. 12, *Reading*. London: Erlbaum.
- Coltheart, M. 1987b. Functional architecture of the language-processing system. In M. Coltheart, G. Sartori, and R. Job, eds., *The cognitive neuropsychology of language*. London: Erlbaum.
- Glanzer, M., and Ehrenreich, S. L. 1979. Structure and search of the internal lexicon. *Journal of Verbal Learning and Verbal Behavior* 18:381-398.
- Glushko, R. J. 1979. The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance* 5:674-691.
- Goodman, K. S. 1967. Reading: A psycholinguistic guessing game. *Journal of the Reading Specialist* 6:126-135.
- Gordon, B. 1983. Lexical access and lexical decision: Mechanisms of frequency sensitivity. *Journal of Verbal Learning and Verbal Behavior* 22:24-44.
- Henderson, L. 1982. *Orthography and word recognition in reading*. London: Academic Press.
- Hillinger, M. L. 1980. Priming effects with phonemically similar words: The encoding-bias hypothesis reconsidered. *Memory and Cognition* 8:115-123.
- Hinton, G. E., McClelland, J. L., and Rumelhart, D. E. 1986. Distributed representations. In D. E. Rumelhart and J. L. McClelland, eds., *Parallel distributed processing: Explorations in the microstructure of cognition*, vol. 1. Cambridge: MIT Press.
- Hinton, G. E., and Sejnowski, T. J. 1986. Learning and relearning in Boltzman machines. In D. E. Rumelhart and J. L. McClelland, eds., *Parallel distributed processing: Explorations in the microstructure of cognition*, vol. 1. Cambridge: MIT Press.
- Jorn, A. F. 1983. *The psychology of reading and spelling disabilities*. London: Routledge and Kegan Paul.
- Kawamoto, A. 1988. Interactive processes in lexical ambiguity resolution. In S. Small, G. Cottrell, and M. K. Tanenhaus, eds. *Lexical ambiguity resolution: Computational, linguistic, and psychological perspectives*. New York: Morgan Kaufman.
- Kucera, H., and Francis, W. N. 1967. *Computational analysis of present-day American English*. Providence: Brown University Press.
- Locke, J. L. 1972. Ease of articulation. *Journal of Speech and Hearing Research* 15:194-200.
- McCann, R. S., and Besner, D. 1987. Reading pseudohomophones: Implications for models of pronunciation assembly and the locus of word-frequency effects in naming. *Journal of Experimental Psychology: Human Perception and Performance* 13:14-24.
- McClelland, J. L. 1987. The case for interactionism in language processing. In M. Coltheart, ed., *Attention and Performance*, vol. 12, *Reading*. London: Erlbaum.
- McClelland, J. L., and Kawamoto, A. 1986. Mechanisms of sentence processing: Assigning roles to constituents. In J. L. McClelland and D. E. Rumelhart, eds., *Parallel distributed processing: Explorations in the microstructure of cognition*, vol. 2. Cambridge: MIT Press.
- McClelland, J. L., and 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.
- McClelland, J. L., and Rumelhart, D. E. 1985. Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General* 114:159-188.
- McClelland, J. L., and Rumelhart, D. E., eds. 1986a. *Parallel distributed processing: Explorations in the microstructure of cognition*, vol. 2. Cambridge: MIT Press.
- McClelland, J. L., and Rumelhart, D. E. 1986b. A distributed model of human learning

- and memory. In J. L. McClelland and D. E. Rumelhart, eds., *Parallel distributed processing: Explorations in the microstructure of cognition*, vol. 2. Cambridge: MIT Press.
- McRae, K., Jared, D., and Seidenberg, M. S. In press. On the roles of frequency and lexical access in word naming. *Journal of Memory and Language*.
- Marshall, J. C., and Newcombe, F. 1973. Patterns of paralexia: A psycholinguistic approach. *Journal of Psycholinguistic Research* 2:175-199.
- Meyer, D. E., Schvaneveldt, R. W., and Ruddy, M. G. 1974. Functions of graphemic and phonemic codes in visual word recognition. *Memory and Cognition*, 2:309-321.
- Monson, S. 1987. On the relation between lexical input and output pathways for speech. In D. A. Allport, D. G. MacKay, W. Prinz, and E. Scheerer eds., *Language perception and production: Relationships among listening, speaking, reading, and writing*. London: Academic Press.
- Morton, J. 1969. Interaction of information in word recognition. *Psychological Review* 76:165-178.
- Onifer, W., and Swinney, D. 1981. Accessing lexical ambiguities during sentence comprehension: Effects of frequency, meaning, and contextual bias. *Memory and Cognition* 9:225-236.
- Patterson, K. E. 1981. Neuropsychological approaches to the study of reading. *British Journal of Psychology* 72:151-174.
- Patterson, K., Marshall, J. C., and Coltheart, M. 1985. *Surface dyslexia*. London: Erlbaum.
- Patterson, K. E., Seidenberg, M. S., and McClelland, J. L. In press. Reading acquisition and dyslexia: A connectionist perspective. In P. Morris, ed., *Connectionism: The Oxford Symposium*. Oxford University Press.
- Reicher, G. M. 1969. Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology* 81:274-280.
- Rozin, P., and Gleitman, L. 1977. The structure and acquisition of reading, II: The reading process and the acquisition of the alphabetic principle. In A. Reber and D. Scarborough, eds., *Toward a psychology of reading*. Hillsdale, N.J.: Erlbaum.
- Rumelhart, D. E., Hinton, G. E., and Williams, R. J. 1986. Learning internal representations by error propagation. In D. E. Rumelhart and J. L. McClelland, eds., *Parallel distributed processing: Explorations in the microstructure of cognition*, vol. 1. Cambridge: MIT Press.
- Rumelhart, D. E., and McClelland, J. L., eds. 1986. *Parallel distributed processing: Explorations in the microstructure of cognition*, vol. 1. Cambridge: MIT Press.
- Schwanenflugel, P. J., and Shoben, E. J. 1985. The influence of sentence constraint on the scope of facilitation for upcoming words. *Journal of Memory and Language* 24:232-252.
- Seidenberg, M. S. 1985a. The time course of phonological code activation in two writing systems. *Cognition* 19:1-30.
- Seidenberg, M. S. 1985b. The time course of information activation and utilization in visual word recognition. In D. Besner, T. Waller, and G. E. MacKinnon, eds., *Reading research: Advances in theory and practice*, vol. 5. New York: Academic Press.
- Seidenberg, M. S. 1987. Sublexical structures in visual word recognition: Access units or orthographic redundancy? In M. Coltheart, ed., *Attention and Performance*, vol. 12, *Reading*. Hillsdale, N.J.: Erlbaum.
- Seidenberg, M. S. In press. Reading complex words. In G. Carlson and M. K. Tanenhaus, eds., *Language comprehension and linguistic theory*. Amsterdam: Reidel.
- Seidenberg, M. S., Bruck, M., Fornarolo, G., and Backman, J. 1986. Word recognition skills of poor and disabled readers: Do they necessarily differ? *Applied Psycholinguistics* 6:161-180.



- Seidenberg, M. S., and McClelland, J. L. In press. A distributed, developmental model of visual word recognition and pronunciation. *Psychological Review*.
- Seidenberg, M. S., Waters, G. S., Barnes, M. A., and Tannenhaus, M. K. 1984. When does irregular spelling or pronunciation influence word recognition? *Journal of Verbal Learning and Verbal Behavior* 23:383-404.
- Shallice, T., and Warrington, E. 1975. Word recognition in a phonemic dyslexic patient. *Quarterly Journal of Experimental Psychology* 27:189-199.
- Shallice, T., Warrington, E., and McCarthy, R. 1983. Reading without semantics. *Quarterly Journal of Experimental Psychology* 35A:111-138.
- Stanovich, K. E. 1986. Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly* 21:360-407.
- Swinney, D. A. 1979. Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior* 18:645-659.
- Tabossi, P. 1988. Effects of context on the immediate interpretation of unambiguous nouns. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 14:153-162.
- Tannenhaus, M. K., Dell, G., and Carlson, G. 1988. Context effects in lexical processing: A Connectionist approach to modularity. In J. Garfield, ed., *Modularity in knowledge representation and natural language processing*. Cambridge: MIT Press.
- Taraban, R., and McClelland, J. L. 1987. Conspiracy effects in word recognition. *Journal of Memory and Language* 26:608-631.
- Theios, J., and Musie, J. G. 1977. The word identification process in reading. In N. J. Castellan, Jr., D. B. Pisoni, and G. R. Potts, eds., *Cognitive theory*, vol. 2. Hillsdale, N.J.: Erlbaum.
- Treiman, R., Freyd, J., and Baron, J. 1984. Phonological recoding and use of spelling-sound rules in reading of sentences. *Journal of Verbal Learning and Verbal Behavior* 22:682-700.
- Teisman, A., and Schmidt, H. 1982. Illusory conjunctions in the perception of objects. *Cognitive Psychology* 14:107-141.
- Venezky, R. 1970. *The structure of English orthography*. The Hague: Mouton.
- Waters, G. S., Bruck, M., and Seidenberg, M. S. 1985. Do children use similar processes to read and spell words? *Journal of Experimental Child Psychology* 39:511-530.
- Waters, G. S., and Seidenberg, M. S. 1985. Spelling-sound effects in reading: Time course and decision criteria. *Memory and Cognition* 13:557-572.
- Wheeler, D. D. 1970. Process in word recognition. *Cognitive Psychology* 1:59-85.

## Editor's Comments

The chapters by Dudai and by Churchland and Sejnowski in this volume complement the present chapter. Parallel-distributed-processing (PDP) models based on paired associated learning, such as those depicted here, are capable of pronouncing English words of many types and successfully completing lexical decision tasks. For me, the most informative and remarkable aspect of this exercise is that the computer outputs closely resemble humans' performances with real language, at least at the level of single words. (This would not be at all obvious were it not for the fact that experimental linguists have been able to specify a great deal of these data for both normal subjects and individuals with developmental and acquired language disturbances.) However, it remains to be seen what will happen when the models are made to perform at levels involving more than single words, at levels where only sentential and logical theories can explain the empirical data.

To neurologists the discovery that "lesioned" machines can produce errors that resemble those made by our patients is truly awesome. It might be argued that because of the nature of the task, the resemblance will break down in one of a limited number of ways (see Morton in this volume). On the other hand, it is astounding that the nature of the implementation hardware, whether cut and dry silicone chips or wet and sinewy neural networks, does not appear to be very important in this respect.

Two simulations of pathological states are presented here. One considers the effects of developmental anomaly by incomplete development (in the simulation the number of hidden units is intentionally diminished). Several of the difficulties encountered by the device thus altered resemble those met by developmental dyslexics. The other considers the effects of damaged connections. This time the device performs similarly to patients with acquired dyslexia.

A third possibility needs to be simulated: a device with increased numbers of processing units or connections. This may turn out to be a fair characterization of at least part of the neural substrate underlying developmental dyslexia (see Sherman et al. and Galaburda et al. in this volume). Excessive numbers of processing units may lead to difficulty with generalization and sluggish or inadequate emergence of rules from the learning period. Excessive numbers of connections at the initial state may lead to nonrandom initial weighting and subsequent interference with appropriate experience-based weighting. As connections may play a role in the sequential linking of separate

**PDP devices, anomalous connections might conceivably lead to what at the cognitive level may look like abnormal sentence and logic architectures, a fair characterization of the performances of some developmental dyslexics.**