A Connectionist Approach to Word Reading and Acquired Dyslexia: Extension to Sequential Processing

David C. Plaut
Departments of Psychology and Computer Science, Carnegie Mellon University, and the Center for the Neural Basis of Cognition

July 1999
To appear in M. H. Christiansen and N. Chater (Eds.), *Connectionist psycholinguistics.* Norwood, NJ: Ablex.

1 Introduction

Many researchers assume that the most appropriate way to express the systematic aspects of language is in terms of a set of rules. For instance, there is a systematic relationship between the written and spoken forms of most English words (e.g., GAVE ⇒ /geɪv/), and this relationship can be expressed in terms of a fairly concise set of grapheme-phoneme correspondence (GPC) rules (e.g., G ⇒ /g/, A_E ⇒ /eɪ/, V ⇒ /v/). In addition to being able to generate accurate pronunciations of so-called *regular* words, such rules also provide a straightforward account of how skilled readers apply their knowledge to novel items—for example, in pronouncing word-like nonwords (e.g., MAVE ⇒ /meɪv/). Most linguistic domains, however, are only partially systematic. Thus, there are many English words whose pronunciations violate the standard GPC rules (e.g., HAVE ⇒ /hæv/). Given that skilled readers can pronounce such *exception* words correctly, GPC rules alone are insufficient. More generally, skilled language performance at every level of analysis—phonological, morphological, lexical, syntactic—requires both effective handling of exceptional items and the ability to generalize to novel forms.

In the domain of reading, there are three broad responses to this challenge. The first, adopted by "dual-route" theories (e.g., Coltheart, Curtis, Atkins, & Haller, 1993; Zorzi, Houghton, & Butterworth, 1998), is to add to the GPC system a separate, *lexical* system that handles the exceptions. The second response, adopted by "multiple levels" theories (e.g., Norris, 1994; Shallice & McCarthy, 1985), is to augment the GPC rules with more specific, context-sensitive rules, (e.g., OOK ⇒ /ʊk/ as in BOOK), including rules that apply only to individual exceptions (e.g., HAVE ⇒ /hæv/). Both of these approaches retain the general notion that language knowledge takes the form of rules (although such rules may be expressed in terms of connections; see, e.g., Norris, 1994; Reggia, Marsland, & Berndt, 1988; Zorzi et al., 1998).

The third response to the challenge, adopted by distributed connectionist theories (Plaut, McClelland, Seidenberg, & Patterson, 1996; Seidenberg & McClelland, 1989; Van Orden, Pennington, & Stone, 1990) and elaborated in the current paper, is more radical. It eschews the notion that the knowledge supporting online language performance takes the form of explicit rules, and thus denies a strict dichotomy between "regular" items which obey the rules and "exception" items which violate them. Rather, it is claimed that language knowledge is inherently graded, and the language mechanism is a learning device that gradually picks up on the statistical structure among written and spoken words and the contexts in which they occur. In this way, the emphasis is on the degree to which the mappings among the spelling, sound, and meaning of a given word are *consistent* with those of other words (Glushko, 1979).

To make this third perspective concrete, consider the connectionist/parallel distributed processing (PDP) framework for lexical processing depicted in Figure 1 (based on Seidenberg & McClelland, 1989). As the figure makes clear, the approach does not entail a complete lack of structure within the reading system. There is, however, uniformity in the processing mechanisms by which representations are generated and interact, and in this respect the approach is quite different from dual-route accounts. Orthographic, phonological, and semantic information is represented in terms of distributed patterns of activity over groups of simple neuron-like processing units. Within each domain, similar words are represented by similar patterns of activity. Lexical tasks involve transformations between these representations—for example, reading aloud requires the orthographic pattern for a word to generate the appropriate phonological pattern. Such transformations are accomplished via the cooperative and competitive interactions among units, including additional *hidden* units that mediate between the orthographic, phonological, and semantic units. In processing an input, units interact until the network as a

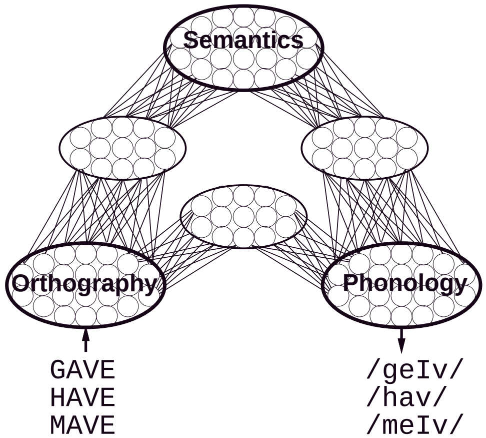

Figure 1: A connectionist framework for lexical processing, based on that of Seidenberg and McClelland (1989). Reprinted from (Plaut, 1997).

whole settles into a stable pattern of activity—termed an *attractor*—corresponding to its interpretation of the input. Unit interactions are governed by weighted connections between them, which collectively encode the system's knowledge about how the different types of information are related. Weights that give rise to the appropriate transformations are learned on the basis of the system's exposure to written words, spoken words, and their meanings.

At a general level, the distributed connectionist approach to word reading is based on three general computational principles:

Distributed representation: Orthography, phonology, and semantics are represented by distributed patterns of activity such that similar words are represented by similar patterns.

Gradual learning of statistical structure: Knowledge of the relationships among orthography, phonology, and semantics is encoded across connection weights that are learned gradually through repeated experience with words in a way that is sensitive to the statistical structure of each mapping.

Interactivity in processing: Mapping among orthography, phonology, and semantics is accomplished through the simultaneous interaction of many units, such that familiar patterns form stable attractors.

Although these principles are general, the challenge is to demonstrate that, when instantiated in a particular domain—single word reading—these principles provide important insights into the patterns of normal and impaired cognitive behavior. The current chapter reviews a series of computational simulations of word reading based on the framework depicted in Figure 1. It then presents a new simulation that address some limitations of this work, relating to sequential processing and effects of orthographic length on the naming latencies of both normal and dyslexic readers. The simulation generates sequential phonological output in response to written input and has the ability to refixate the input when encountering difficulty. The normal model reads both words and nonwords accurately, and exhibits an effect of orthographic length and a frequency-by-consistency interaction in its naming latencies. When subject to peripheral damage, the model exhibits an increased length effect which interacts with word frequency, characteristic of letter-by-letter reading in pure alexia. Although the model is far from a fully adequate account of all the relevant phenomena, it suggests how connectionist models may be extended to provide deeper insight into sequential processes in reading.

2 Background

2.1 Skilled Oral Reading

The distributed connectionist framework for word reading depicted in Figure 1 reflects a radical departure from traditional theorizing about lexical processing, particularly in two ways. First, there is nothing in the structure of the system that corresponds to individual words *per se*, such as a lexical entry, localist word unit (McClelland & Rumelhart, 1981) or "logogen" (Morton, 1969). Rather, words are distinguished from nonwords only by *functional* properties of the system—the way in which particular orthographic, phonological, and semantic patterns of activity interact (also see Plaut, 1997; Van Orden et al., 1990). Second, there are no separate mechanisms for lexical and sublexical processing (cf. Coltheart et al., 1993). Rather, all parts of the system participate in processing all types of input, although, of course, the contributions of different parts may be more or less important for different inputs.

In support of the general framework, Seidenberg and McClelland (1989) trained a connectionist network to map from the orthography of about 3000 monosyllabic English words—both regular and exception—to their phonology. The network corresponded to the bottom portion of the framework in Figure 1 (referred to as the *phonological* pathway). After training, the network pronounced nearly all of the words correctly, including most exception words. It also exhibited the standard empirical pattern of an interaction of frequency and consistency in naming latency (see, e.g., Taraban & McClelland, 1987) when its real-valued accuracy in generating a response was taken as a proxy for response time. However, the model was much worse than skilled readers at pronouncing orthographically legal nonwords (Besner, Twilley, McCann, & Seergobin, 1990) and at lexical decision under some conditions (Besner et al., 1990; Fera & Besner, 1992). Thus, the model failed to refute traditional claims that localist, word-specific representations and separate mechanisms are necessary to account for skilled reading.

More recently, Plaut, McClelland, Seidenberg, and Patterson (1996, also see Seidenberg, Plaut, Petersen, McClelland, & McRae, 1994) have shown that the limitations of the Seidenberg and McClelland model in pronouncing nonwords stem not from any general limitation in the abilities of connectionist networks in quasi-regular domains (as suggested by, e.g., Coltheart et al., 1993), but from its use of poorly structured orthographic and phonological representations. The original simulation used representations based on context-sensitive triples of letters or phonemic features. When more appropriately structured representations are used—based on graphemes and phonemes and embodying phonotactic and graphotactic constraints—network implementations of the phonological pathway can learn to pronounce regular words, exception words, and nonwords as well as skilled readers. Moreover, the networks exhibit the empirical frequency-by-consistency interaction pattern when trained on actual word frequencies. This remains true if naming latencies are modeled directly by the settling time of a recurrent, attractor network (see Figure 2).

Plaut et al. (1996) also offered a mathematical analysis of the critical factors that govern *why* the networks (and, by hypothesis, subjects) behave as they do. The analysis was based on a network that, while simpler than the actual simulations—it had no hidden units and employed Hebbian learning—retained many of the essential characteristics of the more general framework (e.g., distributed representations and structure-sensitive learning). For this simplified network, it was possible to derive an analytic expression for how the response of the network to any input (test) pattern depends on its experience with every pattern on which the network is trained, as a function of its frequency of training, its similarity with the test pattern, and the consistency of its output with that of the test pattern. Specifically, the response $s_j^{[t]}$ of any output unit j to a given test pattern t is given by

$$s_j^{[t]} = \sigma\left(F^{[t]} + \sum_f F^{[f]} O^{[ft]} - \sum_e F^{[e]} O^{[et]}\right) \qquad (1)$$

in which the standard smooth, non-linear sigmoidal input-output function for each unit, $\sigma(\cdot)$, is applied to the sum of three terms: (1) the cumulative frequency of training on the pattern t itself, $F^{[t]}$; (2) the sum of the frequencies $F^{[f]}$ of the *friends* of pattern t (similar patterns trained to produce the same response for unit j), each weighted by its similarity (overlap) with t, $O^{[ft]}$; and (3) minus the sum of the frequencies $F^{[e]}$ of the *enemies* of pattern t (similar

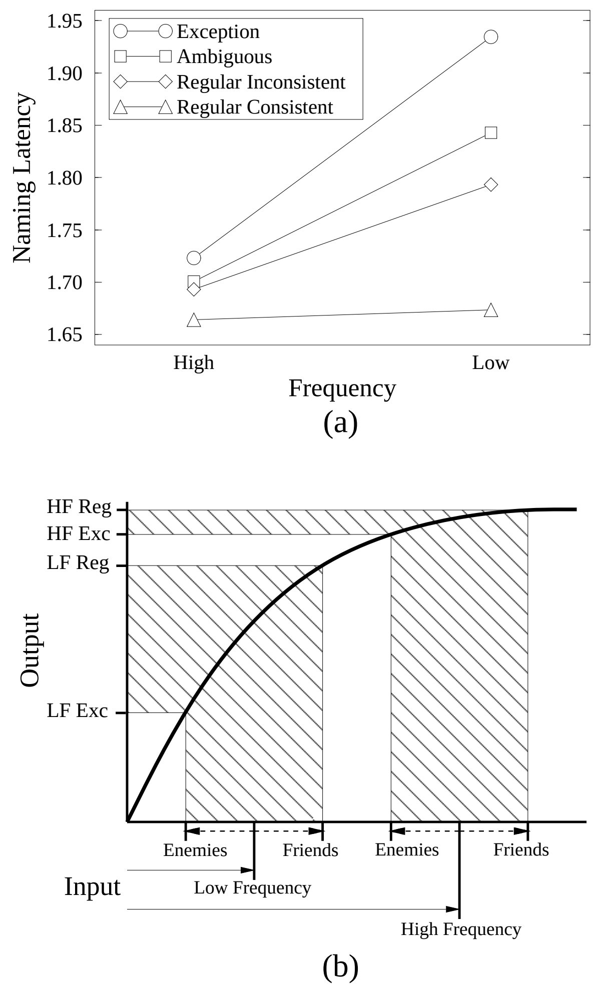

Figure 2: (a) The frequency-by-consistency interaction exhibited in the settling time of an attractor network implementation of the phonological pathway in pronouncing words of varying frequency and spelling-sound consistency (Plaut et al., 1996, Simulation 3); and (b) its explanation in terms of additive contributions of frequency and consistency subject to an asymptotic activation function (only the top of which is shown).

patterns trained to produce the opposite response), each weighted by its similarity to t, $O^{[et]}$.

Many of the basic phenomena in word reading can be seen as natural consequences of adherence to this *frequency-consistency* equation. Factors that increase the summed input to units (e.g., word frequency, spelling-sound consistency) improve performance as measured by naming accuracy and/or latency, but their contributions are subject to "diminishing returns" due to the asymptotic nature of the activation function (see Figure 2b). As a result, performance on stimuli that are strong in one factor is relatively insensitive to variation in other factors. Thus, regular words show little effect of frequency, and high-frequency words show little effect of consistency, giving rise to the standard pattern of interaction between frequency and consistency, in which the naming of low-frequency exception words is disproportionately slow or inaccurate.

2.2 Surface Dyslexia

Although implementations of the phonological pathway on its own can learn to pronounce words and nonwords as well as skilled readers, a central aspect of Plaut et al.'s (1996) general theory is that skilled reading more typically requires the combined support of both the semantic and phonological pathways, and that individuals may differ in the relative competence of each pathway. A consideration of semantics is particularly important in the context of accounting for a pattern of reading impairment known as *surface* dyslexia (see Patterson, Coltheart, & Marshall, 1985), which typically arises from damage to the left temporal lobe. Surface dyslexic patients read nonwords and regular words with normal accuracy and latency, but exhibit an interaction of frequency and consistency in word reading accuracy, such that low-frequency exception words are pronounced disproportionately poorly, often eliciting a pronunciation consistent with more standard spelling-sound correspondences (e.g., SEW read as "sue," termed a *regularization* error).

The framework for lexical processing depicted in Figure 1 (and the associated computational principles) provides an account of surface dyslexia based on the relative contributions of the semantic and phonological pathways in oral reading. At an abstract level, given that phonological units simply sum their inputs from the two pathways, the influence of the semantic pathway can be included in a straightforward manner by adding an additional term, $S^{[t]}$, to the summed input in Equation 1. Furthermore, if this term is assumed to increase with imageability, the equation produces the three-way interaction of frequency, consistency, and imageability found by Strain, Patterson, and Seidenberg (1995). When formulated explicitly in connectionist terms, however, this integration has important implications for the nature of *learning* in the two pathways. To the extent that the semantic pathway reduces performance error during training by contributing to the correct pronunciation of words, the phonological pathway will experience less pressure to learn to pronounce all of the words by itself. Rather, this pathway will tend to learn best those words high in frequency and/or consistency; on its own it may never master low-frequency exception words completely. On this account, the combination of the semantic and phonological pathways is fully competent in normal readers, but brain damage that impairs the semantic pathway reveals the latent limitations of an intact but isolated phonological pathway, giving rise to surface dyslexia.

Plaut et al. (1996) explored the viability of this account by extending their simulations of the phonological pathway to include influences from a putative semantic pathway. They approximated the contribution that a semantic pathway would make to oral reading by providing the output (phoneme) units of the phonological pathway with external input that pushed the activations of these units towards the correct pronunciation of each word during training. Plaut and colleagues found that, indeed, a phonological pathway trained in the context of support from semantics exhibited the central phenomena of surface dyslexia when the contribution of semantics was removed (see Figure 3). Moreover, individual differences in the severity of surface dyslexia could arise, not only from differences in the amount of semantic damage, but also from *premorbid* differences in the division of labor between the semantic and phonological pathways (Plaut, 1997). Thus, the few patients exhibiting mild to moderate semantic impairments without concomitant regularization errors (DRN, Cipolotti & Warrington, 1995; DC, Lambon Ralph, Ellis, & Franklin, 1995) may have, for various reasons, reading systems with relatively weak reliance on the semantic pathway.

2.3 Deep and Phonological Dyslexia

Patients with *deep* dyslexia (see Coltheart, Patterson, & Marshall, 1980) have reading impairments that are in many ways opposite to those with surface dyslexia, in that they appear to read almost entirely via semantics. Deep dyslexic patients are thought to have severe damage to the phonological pathway, as evidenced by their virtual inability to read even the simplest of pronounceable nonwords. They also have impairments in reading words that suggest additional partial damage to the semantic pathway. In particular, the hallmark symptom of deep dyslexia is the occurrence of *semantic* errors in oral reading (e.g., reading CAT as "dog"). Interestingly, these semantic errors co-occur with pure *visual* errors (e.g., CAT ⇒ "cot"), mixed *visual-and-semantic* er-

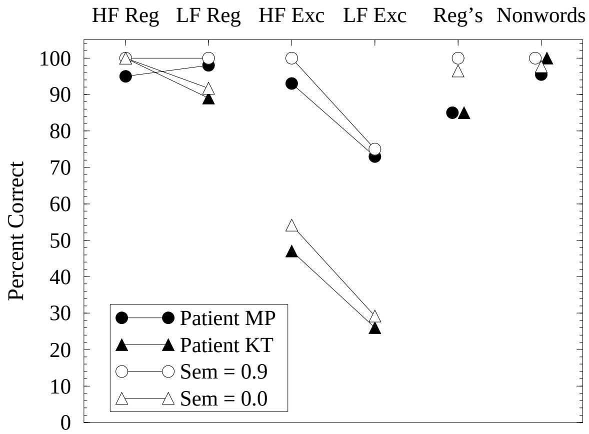

Figure 3: Performance of two surface dyslexic patients (MP, Behrmann & Bub, 1992; Bub, Cancelliere, & Kertesz, 1985; and KT, McCarthy & Warrington, 1986) and the Plaut et al. (1996) network for two levels of semantic impairment. Correct performance is given for Taraban and McClelland's (1987) high-frequency (HF) and low-frequency (LF) regular consistent words (Reg) and exception words (Exc), and for Glushko's (1979) nonwords. "Reg's" is the approximate percentage of errors on the exception words that are regularizations. Adapted from Plaut et al. (1996).

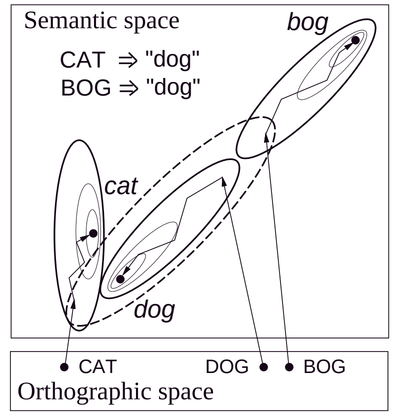

Figure 4: A depiction of the attractor landscape for a network that maps orthography to semantics, and how damage to the network can distort the attractors (dashed oval) in a way that gives rise to both semantic errors (e.g., CAT ⇒ "dog") and visual errors (e.g., BOG ⇒ "dog"). Adapted from Plaut and Shallice (1993).

rors (e.g., CAT ⇒ "rat"), and even mediated *visual-then-semantic* errors (e.g., SYMPATHY ⇒ "orchestra", presumably via *symphony*). Furthermore, correct performance depends on part-of-speech (nouns > adjectives > verbs > function words) and concreteness or imageability (concrete, imageable words > abstract, less imageable words). Finally, differences across patients in written and spoken comprehension, and in the distribution of error types, suggests that the secondary damage to the semantic pathway may occur before, within, or after semantics (Shallice & Warrington, 1980).

Deep dyslexia is closely related to another type of acquired dyslexia—so-called *phonological* dyslexia (Beauvois & Derouesné, 1979), involving a selective impairment in reading nonwords compared with words (without concomitant semantic errors). Indeed, some authors (Friedman, 1996; Glosser & Friedman, 1990) have argued that deep dyslexia is only the most severe form of phonological dyslexia.

Hinton and Shallice (1991) reproduced the co-occurrence of visual, semantic, and mixed visual-and-semantic errors in deep dyslexia by damaging a connectionist network that mapped orthography to semantics. During training, the network learned to form *attractors* for 40 word meanings across five categories, such that patterns of semantic features that were similar to a known word meaning were pulled to that exact meaning over the course of settling. When the network was damaged, the initial semantic activity caused by an input would occasionally fall within a neighboring attractor basin, giving rise to an error response. These errors were often semantically related to the stimulus because words with similar meanings correspond to nearby attractors in semantic space. The damaged network also produced visual errors due to its inherent bias towards similarity: visually similar words tend to produce similar initial semantic patterns, which can lead to a visual error if the basins are distorted by damage (see Figure 4).

Plaut and Shallice (1993) extended these initial findings in a number of ways. They established the generality of the co-occurrence of error types across a wide range of simulations, showing that it does not depend on specific characteristics of the network architecture, the learning procedure, or the way responses are generated from semantic activity. A particularly relevant simulation in this regard involved an implementation of the full semantic pathway—mapping orthography to phonology via semantics—using a deterministic Boltzmann Machine (Hinton, 1989b; Peterson & Anderson, 1987). Lesions throughout the network gave rise to both visual and semantic errors, with lesions prior to semantics producing a bias towards visual errors and lesions after semantics producing a bias towards semantic errors. Thus, the network replicated both the qualitative similarity and quantitative differences among deep dyslexic patients. The network also exhibited a number of other characteristics of deep dyslexia not considered by Hinton and Shallice (1991), including the occurrence of visual-then-semantic errors, greater confidence in visual as compared with semantic errors, and relatively preserved lexical decision with impaired naming.

Plaut and Shallice carried out additional simulations to address the influence of concreteness on the reading performance of deep dyslexic patients. Another full implementation of the semantic pathway was trained to pronounce a new set of words consisting of both concrete and abstract words. Concrete words were assigned far more semantic features than were abstract words, under the assumption that the semantic representations of concrete words are less dependent on the contexts in which they occur (Jones, 1985; Saffran, Bogyo, Schwartz, & Marin, 1980; Schwanenflugel, 1991). As a result, the network developed stronger attractors for concrete than abstract words during training, giving rise to better performance in reading concrete words under most types of damage, as observed in deep dyslexia (see Figure 5a). Surprisingly, severe damage to connections implementing the attractors at the semantic level produced the opposite pattern, in which the network read *abstract* words better than concrete words (see Figure 5b). This pattern of performance is reminiscent of CAV, the single, enigmatic patient with *concrete word dyslexia* (Warrington, 1981). The double dissociation between reading concrete versus abstract words in patients is often interpreted as implying that there are separate modules within the cognitive system for concrete and abstract words. The Plaut and Shallice simulation demonstrates that such a radical interpretation is unnecessary: the double dissociation can arise from damage to different parts of a distributed network, in which parts process both types of items but develop somewhat different functional specializations through learning (see Plaut, 1995, for further results and discussion).

Taken together, the modeling work described above provides strong support for a connectionist approach to normal and impaired word reading, embodying the computational principles outlined in the Introduction: distributed representation, gradual learning of statistical structure, and interactivity in processing. There have, however, been recent empirical challenges to the specific models in particular, and the framework in general, which ultimately need to be addressed if the approach is to remain viable as an account of human performance. A number of these relate to the influence of orthographic length on the naming latencies of both normal and dyslexic readers.

3 Current Challenges: Length Effects

An aspect of the Seidenberg and McClelland (1989) and Plaut et al. (1996) models that has contributed substantially to their theoretical impact is that, because they were

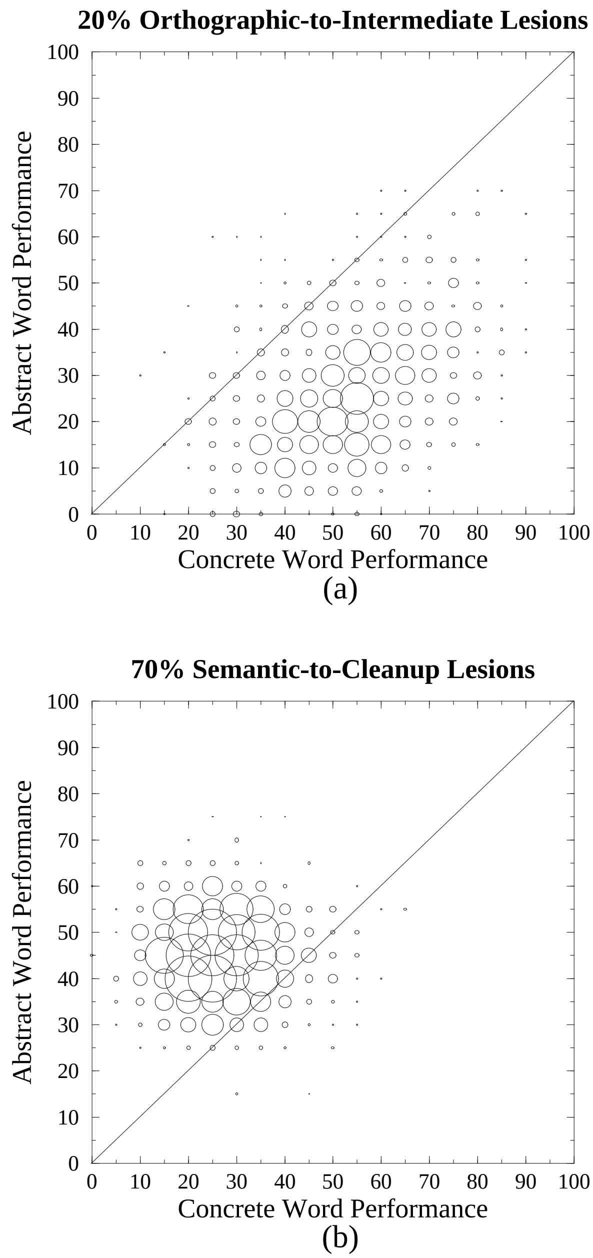

Figure 5: Percent correct performance on concrete versus abstract words of the Plaut and Shallice (1993) simulation after (a) 1000 lesions of 20% of orthographic-to-intermediate connections and (b) 1000 lesions of 70% of semantic-to-cleanup connections. The radius of each circle is proportional to the number of lesions yielding the performance levels indicated by the position of the circle. The diagonal lines correspond to equal levels of performance on concrete and abstract words. The advantage for concrete words in (a) corresponds to the findings for deep dyslexia (Coltheart et al., 1980), whereas the advantage for abstract words in (b) corresponds to the findings for concrete-word dyslexia (Warrington, 1981).

trained on a sufficiently extensive corpus of words, their performance can be compared directly with that of human subjects on the very same stimuli. These comparisons have largely been successfully at the level of accounting for the effects of factorial manipulations (e.g., word frequency, spelling-sound consistency). More recently, however, the models have been found to be lacking when compared with human performance on an item-by-item basis. For instance, Spieler and Balota (1997) correlated the mean naming latencies of 31 subjects naming 2820 words with the models' latencies for the same words, and found that the models accounted for only about 3–10% of the variance associated with individual items. By contrast, the combination of the traditional measures of log frequency, orthographic length, and orthographic neighborhood size (Coltheart's N) collectively accounted for 21.7% of the variance; including an encoding of phonetic properties of the onset phoneme increased this figure to 43.1%.

In response, Seidenberg and Plaut (1998) carried out additional analyses with the Spieler and Balota (1997) dataset as well as another large naming dataset (Seidenberg & Waters, 1989). They found that the models did not account well for effects of orthographic length, but when the model measures and length were entered first in a stepwise regression, there was little remaining variance accounted for by log frequency and orthographic neighborhood. Specifically, each traditional variable accounted for less than 1.7% of the remaining variance in all conditions, except that log frequency still accounted for 4.8% of the variance in the Spieler and Balota dataset (but only 0.25% in the other dataset) after length and the Plaut et al. (1996) RTs were partialed out. Thus the models provide a reasonably good (as well as mechanistic) account of the influence of these traditional factors on naming performance. With regard to orthographic length, Seidenberg and Plaut argued that the effects of this factor were due largely to visual and articulatory factors outside the domain of the existing models.[1]

More recently, Chris Kello (personal communication, January 1998) has provided some support for this claim. He hypothesized that some of the observed length effect might be due to the fact that longer monosyllabic words are more likely to have complex onset consonant clusters (e.g., /pr/, /str/), and the reduced acoustic amplitude at the beginning of such clusters introduces delay in tripping a standard voice key. For example, a voice key might register the /r/ in both RING and STRING, yielding an overly long RT in the latter case (extended by roughly the duration of the /st/). Kello repeated the Spieler and Balota (1997) stepwise regression analysis but used a more sophisticated encoding of the phonetic properties of word onsets, including the presence of certain consonant clusters. He found that, compared with the use of Spieler and Balota's encoding, the new encoding reduced the amount of residual variance accounted for by orthographic length by well over half, from 7.5% to 3.3%. These results indicate that a sizable amount of the effects of orthographic length can be accounted for by articulatory onset characteristics.

Although articulatory factors may contribute substantially to length effects, they cannot be the whole story. Recently, Weekes (1997) has demonstrated differential effects of length for words versus nonwords matched for onset characteristics. Specifically, using 3–6 letter words and nonwords, Weekes found reliable length effects for nonwords and for low- but not high-frequency words. When he partialed out orthographic neighborhood size, the length effect was eliminated for words but not for nonwords. Weekes argued that these findings pose problems for any account in which words and nonwords are processed by a single mechanism.

Finally, length effects also play a prominent role in the analysis of acquired reading impairments, particularly in the context of the letter-by-letter (LBL) reading of pure alexic patients (Dejerine, 1892) and some nonfluent surface dyslexic patients (e.g., Patterson & Kay, 1982). Although the accuracy of these patients can be quite high, their naming latencies show an abnormally large word length effect, sometimes on the order of 1–3 seconds per letter (cf. 5–50 msec/letter for normal readers; Henderson, 1982). One account of such patients (Patterson & Kay, 1982) is that they have a peripheral deficit that prevents adequate activation of letter representations in parallel; they thus must resort to a compensatory strategy of recognizing letters sequentially.

There is, in fact, considerable independent evidence for peripheral impairments in LBL readers (see Behrmann, Nelson, & Sekuler, 1998a, for review). On the other hand, there is also evidence for the influence of lexical/semantic factors on LBL reading performance. There are two forms of this latter influence. First, when presented with words too briefly to allow overt naming, some LBL readers can nonetheless perform lexical decision and semantic categorizations tasks above chance (Coslett & Saffran, 1989; Shallice & Saffran, 1986). Quite apart from this type of "covert" reading, LBL readers also show lexical effects on their letter-by-letter reading latencies. For example, Behrmann, Plaut, and Nelson (1998b) present data on seven LBL readers of varying severity, showing that the magnitudes of their length effects interacted both with frequency and with imageability. Moreover, these inter-

[1] In their reply to Seidenberg and Plaut (1998), Balota and Spieler (1998) question whether length effects fall outside the scope of the models given that Plaut et al. (1996, p. 85) actually demonstrated a small but reliable effect of length on the settling times of their attractor model. However, the fact that the model shows some sensitivity to length does not entail that it should be expected to account for all or even most of the effects of length on performance; the underlying theory may still ascribe length effects to other (unimplemented) parts of the reading system.

actions were modulated by severity of the impairment, such that the most severe patients showed the strongest lexical/semantic effects. Behrmann and colleagues argue that these higher-level effects in LBL reading are consistent with a peripheral impairment given the interactive nature of processing with the reading system: weakened (sequential) letter activation supports partial lexical/semantic activation that accumulates over time and feeds back to facilitate subsequent letter processing. They also propose that the sequential processing in LBL reading is not an abnormal strategy employed only following brain damage, but is the manifestation of the normal reading strategy of making additional fixations when encountering difficulty in reading text (Just & Carpenter, 1987; Reichle, Pollatsek, & Rayner, 1998). For example, in order to enhance stimulus quality, normal subjects make more fixations within long compared with short words. LBL readers also fixate more frequently; in fact, given the very poor quality of the visual input, they fixate almost every letter (Behrmann, Barton, Shomstein, & Black, 1999).

In summary, the effects of orthographic length on naming latency, both in normal and brain-damaged subjects, place important constraints on theories of word reading, and existing distributed models do not provide an adequate account of these effects. A fully adequate model of length effects in reading would need to incorporate considerably detailed perceptual and articulatory processes in addition to the more central processes relating orthography, phonology, and semantics. The intent of the simulation described in the following section is not so much to attempt such a comprehensive account, but rather to begin an exploration of the kinds of networks and processes that might provide deeper insight into length effects.

4 Simulation

4.1 Method

A simple recurrent network (Elman, 1990) was trained to produce a sequence of phonemes as output when given a string of position-specific letters as input. The training corpus consisted of the 2998 monosyllabic words in the Plaut et al. (1996) corpus. The architecture of the network is shown in Figure 6. There are 26 letter units and a "blank" unit at each of 10 positions. The third position from the left, indicated by the dark rectangle in the figure, corresponds to the point of fixation. These 270 letter units are fully connected to 100 hidden units which, in turn, are fully connected to 36 phoneme units.[2] The hidden units also receive input from the previous states of phoneme units. In addition, there is a fourth group of *position* units, with connections both to and from the hidden units, that the network uses to keep track of where it is in the letter string as it is producing the appropriate sequence of phonemes, analogous to a focus of attention. Two copies of the position units and the phoneme units are shown in the figure simply to illustrate their behavior over time. Finally, there is a "done" output unit that the network uses to indicate that a pronunciation is complete. Including bias connections (equivalent to connections from an additional unit with a fixed state of 1), the network had a total of 45,945 connections that were randomized uniformly between ± 1.0 before training.[3]

In understanding how the network was trained, it will help to consider first its operation after it has achieved a reasonable level of proficiency at its task. First, a word is selected from the training corpus according to a logarithmic function of its frequency of occurrence (Kučera & Francis, 1967). Its string of letters is presented with the first letter at fixation,[4] by activating the appropriate letter unit at each corresponding position, and the blank unit at all other positions. Position information for internal letters is assumed to be somewhat inaccurate (see, e.g., Mozer, 1983), so that the same letter units at neighboring internal positions are also activated slightly (to 0.3). In Figure 6, the grey regions for letter units indicate the activations for the word BAY when fixating the B. Initially, the position unit corresponding to fixation (numbered 0 by convention) is active and all others are inactive, and all phoneme units are inactive. (In the figure, the states of position and phoneme units show the network attempting AY ⇒ /A/ after having generated B ⇒ /b/.) Hidden unit states are initialized to 0.2 at the beginning of processing the word.

The network then computes new states for the hidden units, phoneme units, and position units. The network has two tasks: 1) to activate the phoneme corresponding to the current grapheme, and 2) to activate the position of the *next* grapheme in the string (or, if the end of the string is reached, the position of the adjacent blank). For example, when attending to the letter B at fixation in BAY, the network must activate the /b/ unit and position unit 1 (the position of AY in the input). Specifically, the target activations for the phoneme units consist of a one for the correct current phoneme and zeros elsewhere, and the targets for

[2]The encoding of words and nonwords as sequences of phonemes was based on the phonological representation employed by Plaut and McClelland (1993), which differs slightly from that used by Plaut et al. (1996).

[3]Given the composition of the training corpus and all possible refixations, 62 of the letter units would never be activated during training. Therefore, to reduce the computational demands of the simulation slightly, all 6200 outgoing connections from these units were removed, leaving an actual total of 39,745 connections in the network.

[4]A more empirically accurate positioning would have placed the string so that fixation falls at or just to the left of the center of the word, corresponding to the "optimal" or "convenient" viewing position (see O'Regan, 1981). This distinction has no functional consequences for the current model, however, as it does not incorporate variation in visual acuity with eccentricity.

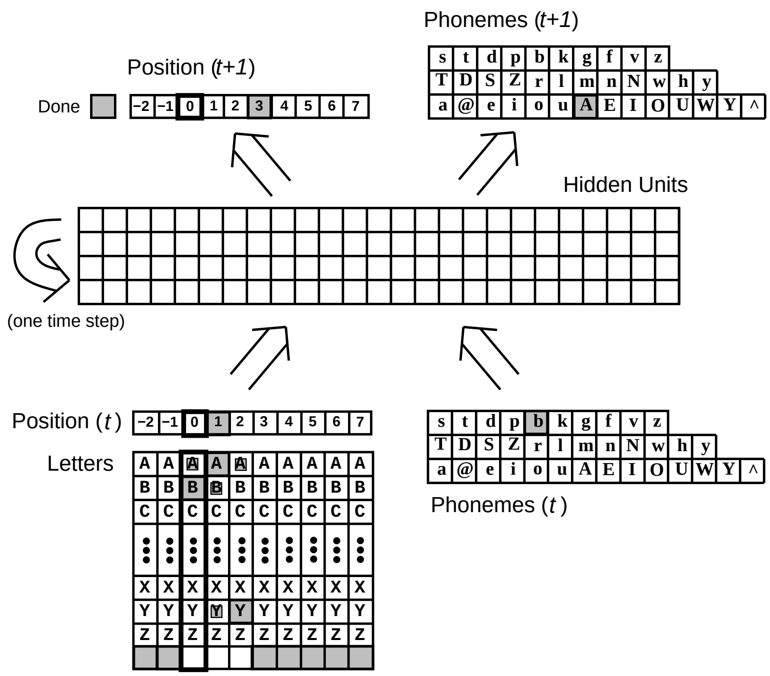

Figure 6: The network architecture for the refixation network. The arrows indicated full connectivity between groups of units. The recurrent connections among the hidden units only convey information about the last time step. The grey areas in the input and output units are intended to depict their activities at an intermediate point in processing the word BAY, after the B ⇒ /b/ has been pronounced (with no refixation) and the AY ⇒ /A/ is being attempted.

the position units consist of a one for the position of the next grapheme/blank in the string and zeros elsewhere. To the extent that the activations over the phoneme and position units are inaccurate (i.e., not within 0.2 of their target values), error is injected and back-propagated through the network. Performance error was measured by the *cross-entropy* (see Hinton, 1989a) between the correct and target activations.

Assuming that the network succeeds at generating the correct phoneme and position, this information is then used to guide the production of the next phoneme and position. For this purpose, the correct phoneme unit had to be activated above 0.7 and all others had to be below 0.3, and the correct position unit had to be more active than any other position unit. (During testing, this criterion applies to the most active phoneme unit rather than to the "correct" unit.) As shown in Figure 6 for BAY, position unit 1 and the phoneme /b/ are now active, the letter input remains the same, and the network must activate /A/, the phoneme corresponding to the indicated grapheme AY); position unit 3, corresponding to the blank following the string; and the "done" unit, indicating a complete pronunciation. In general, when pronouncing a letter string, the network is trained to activate the sequence of phonemes corresponding to its pronunciation, while simultaneously keeping track of the position of the grapheme it is currently working on.

If, in pronouncing a letter string, every phoneme and position is generated correctly, the activations over the letter units remain fixed. If, however, the network fails at generating the correct phoneme or next position at some point, it *refixates* the input string and tries again. It does this by making the equivalent of a rightward saccade to fixate the problematic grapheme, using the position units as a specification of its position relative to fixation. This position information was generated over the position units on the previous time step, and thus is available to guide the appropriate saccade.[5] The actual saccade is implemented by shifting the input activation of the letter units to the left by the specified amount, and resetting position unit 0 to be active. Following this, the network tries again to pronounce the (now fixated) grapheme, and then the remainder of the input string.

In general, the network pronounces as much of the static input as it can until it runs into trouble, then saccades to that part of the input and continues. Note that,

[5]If the network fails on the first grapheme of a string, or immediately after refixating, the target for the position units is used during training as the location of the next fixation; during testing, the most active position unit is used. Also note that the network's rightward saccades are different than the regressive (leftward) saccades that subjects sometimes make when encountering difficult text (see Just & Carpenter, 1987). The current network cannot make regressive saccades.

early on in training, the network repeatedly fails at generating correct output, and so is constantly refixating. This means that essentially all of its training experience consists of pronouncing graphemes (in context) at fixation. As the network learns to pronounce these correctly, it begins to attempt to pronounce the graphemes in the near (right) periphery without refixating. If it fails, it will make a saccade and use its more extensive experience at fixation. Gradually, however, it will learn to pronounce these adjacent graphemes correctly, and will go on to attempt even more peripheral ones. In this way, the network's competence extends gradually from fixation rightward to larger and larger portions of input strings, making fewer and fewer fixations per word as a result. However, the network can always fall back on its more extensive experience at fixation whenever it encounters difficulty. It is perhaps worth noting in this context that, although the network was trained only on monosyllabic words for convenience, it would be entirely straightforward to apply it to pronouncing polysyllabic words of arbitrary length.

To summarize, as the network is trained to produce the appropriate sequence of phonemes for a letter string, it is also trained to maintain a representation of its current position within the string. The network uses this position signal to refixate a peripheral portion of the input when it finds that portion difficult to pronounce. This repositions the input string so that the peripheral portion now falls at the point of fixation, where the network has had more experience in generating pronunciations. In this way, the network can apply the knowledge tied to the units at the point of fixation to any portion of the string that is difficult for the network to read.

4.2 Results and Discussion

Normal Performance. The network was trained on 400,000 word presentations with a learning rate of 0.01, momentum of 0.9, and weight decay of 0.000001. The learning rate was then reduced to 0.001 and the network was trained on an additional 50,000 word presentations, in order to minimize the noise in the final weight values due to sampling error among training examples. The total number of presentations per word ranged from about 40 to 600, with a median of 130. Figure 7 shows, over the course of training, both the overall level of accuracy in pronouncing words as well as the mean number of fixations required. At the end of training, the network read 2978/2998 (99.3%) of the words correctly (where homographs were considered correct if they elicited either appropriate pronunciation). The network made an average of 1.32 fixations per word in generating correct pronunciations, with 2290 (76.9%) involving a single fixation. Just under half (8/20) of the errors were regularizations of low-frequency exception words (e.g., BROOCH ⇒ "brewch", SIEVE ⇒ "seeve").

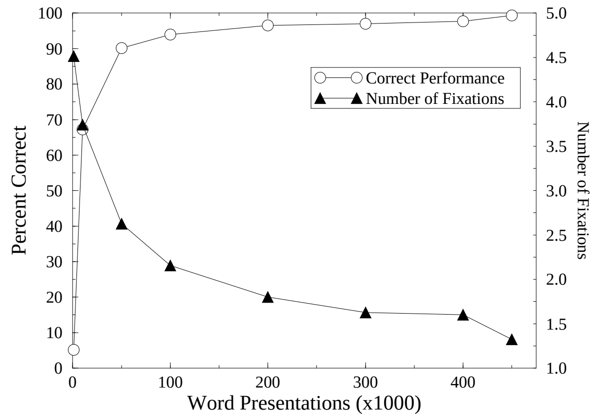

Figure 7: Percentage of words pronounced correctly by the network (top curve; left axis) and the mean number of fixations required (bottom curve, right axis) as a function of the number of words presented during training. The improvement in performance from 400,000 to 450,000 is due to a reduction in learning rate (see text).

Given that the network essentially has a feedforward architecture and outputs only a single phoneme at a time, it is not entirely clear what an appropriate measure of naming latency should be. The most natural analogue to the onset of acoustic energy that would trip a voice key in a standard empirical study would be the real-valued error on the first phoneme. This measure, however, fails to take into account the coarticulatory constraints on executing a fluent pronunciation that apply for subjects but not for the model. A more appropriate, albeit coarse measure in the current context is simply the number of fixations required to generate a correct pronunciation. This measure directly reflects the degree of difficulty that the system experiences in constructing a complete pronunciation.[6]

Figure 8 shows the mean number of fixations made by the model in generating correct pronunciations for words in the training corpus as a function of their length in letters. Using this measure as an analogue to naming latency, the model shows no latencies differences between 3- and 4-letter words ($F < 1$), but a steady increase in latency for 4–6 letter words and an overall length effect ($F_{3,2932} = 76.7$, $p < .001$) with a slope of 0.18 fixations per letter.

The network was tested for its ability to account for two sets of recent findings concerning length effects in normal readers. First, as mentioned earlier, Weekes (1997) found reliable effects of orthographic length in the naming laten-

[6]There is emerging evidence that subjects can initiate their articulation prior to computing the entire pronunciation of a word (Kawamoto, Kello, Jones, & Bame, 1998). Note, however, that the most difficult aspect of mapping orthography to phonology in English relates to inconsistency in vowel pronunciations, and the fixation measure used in the current simulation is sufficiently sensitive to reflect this property.

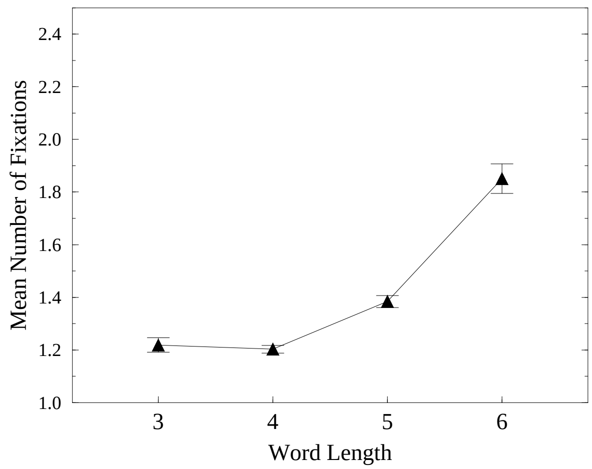

Figure 8: Mean number of fixations made by the network in pronouncing 3–6 letter words. The y-axis scale is the same as that in Figure 11 for ease of comparison.

cies for both words and nonwords, but only the nonword effect remained reliable when orthographic neighborhood size was partialed out. In applying the current model to Weekes' stimuli, 24 of the words had to be eliminated because they are not in the model's training corpus; most of these are inflected forms (e.g., BOARDS, CALLED). Of the remaining items, the model correctly pronounced 86/86 of the high-frequency words, 89/90 of the low-frequency words, and 90/100 of the nonwords (where 4 of the 10 errors were on pseudo-inflected forms; e.g., BRANKS, LOAKED). A nonword pronunciation was scored as correct if it matched the pronunciation of some word in the training corpus (e.g., GROOK pronounced to rhyme with BOOK; see Plaut et al., 1996, for details).

Comparing 4- versus 6-letter stimuli, there was a reliable length effect in the mean number of fixations made by the model in correctly pronouncing high-frequency words (1.00 vs. 1.25; $F_{1,34}$ = 7.56, $p < .01$), low-frequency words (1.38 vs. 1.79; $F_{1,41}$ = 1.82, $p < .05$), and nonwords (1.61 vs. 2.38; $F_{1,42}$ = 6.55, $p < .01$). When orthographic neighborhood size (calculated over the training corpus) was first partialed out of the data, the length effects for both high- and low-frequency words were eliminated ($F_{1,34} < 1$ and $F_{1,41}$ = 1.43, $p > .2$, respectively) whereas the length effect for nonwords remained reliable ($F_{1,42}$ = 6.43, $p < .05$). The only discrepancy between these finding and those of Weekes (1997) is that the small length effect for high-frequency words was reliable for the model but not for the human subjects.

The second length effect to which the model was applied was the recent finding of Rastle and Coltheart (1998) that, among 5-letter nonwords, those with 3-phoneme pronunciations (e.g., FOOPH) produce longer naming latencies than those with 5-phoneme pronunciations (e.g., FROLP); note that this is an effect of *phonological* rather than orthographic length. Certain aspects of Rastle and Coltheart's stimuli are problematic in the current context—namely, 5 of the 24 5-phoneme nonwords are pseudo-inflected (e.g., FRULS). If these and the matched 3-phoneme nonwords are removed from the analysis, the mean number of fixations made by the model in pronouncing the 3-phoneme nonwords is numerically larger than that for the 5-phoneme nonwords, but the difference is not reliable (2.95 vs. 2.79, respectively; paired $t_{17} < 1$). The null result may stem in part from the small number of comparisons but also from the fact that, under the model's phonological encoding, the stimuli that Rastle and Coltheart considered to have 3 phonemes actually had a mean phonological length of 3.58, as a number of the nonwords have 4 or even 5 phonemes (e.g., BARCH ⇒ /bartS/).

The network was also tested for the standard effects of word frequency and spelling-sound consistency in its number of fixations, using a list of 126 matched pairs of regular and exception words falling into three frequency bands (Patterson & Hodges, 1992). The network mispronounced five of the words, producing regularization errors to four low-frequency exception words—BROOCH, SIEVE, SOOT, and SUEDE—and an *irregularization* error to a low-frequency regular word—SOUR to rhyme with POUR (see Patterson, Plaut, McClelland, Seidenberg, Behrmann, & Hodges, 1996, for empirical evidence supporting the occasional occurrence of such errors). Figure 9 shows the mean number of fixations required to correctly pronounce the remaining words, as a function of their frequency and consistency. Overall, there was a main effect of frequency (means: high 1.04, medium 1.35, low 1.62; $F_{2,241}$ = 22.4, $p < .001$) and a main effect of consistency (means: regular 1.14, exception 1.52; $F_{1,241}$ = 27.5, $p < .001$), as well as a frequency-by-consistency interaction, with low-frequency exception words requiring disproportionately more fixations ($F_{2,241}$ = 7.67, $p < .001$). These results are in accord with the relevant empirical findings on the naming latencies of skilled readers.

At the item level, the numbers of fixations made by the model was regressed against the mean naming latencies of Spieler and Balota's (1997) 31 subjects. Over the 2812/2820 words that the model pronounced correctly, its number of fixations accounted for 8.8% of the variance in the latency data (t_{2810} = 16.5, $p < .001$). This value is much better than that of the Plaut et al. (1996) model (3.3%) but not quite as good as the Seidenberg and McClelland (1989) model (10.1%).

Finally, the network was tested for its accuracy in pronouncing three sets of nonwords from two empirical studies: 1) 43 nonwords derived from regular words (Glushko, 1979); 2) 43 nonwords derived from exception words (Glushko, 1979); and 3) 80 nonwords used as controls for

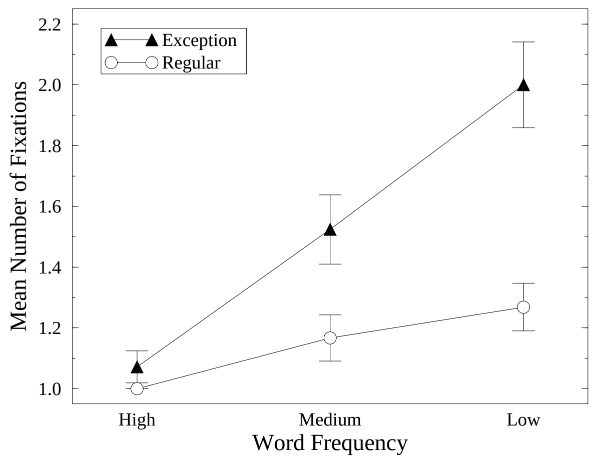

Figure 9: Mean number of fixation required to produce correct pronunciations for words (Patterson & Hodges, 1992) as a function of their frequency and spelling-sound consistency.

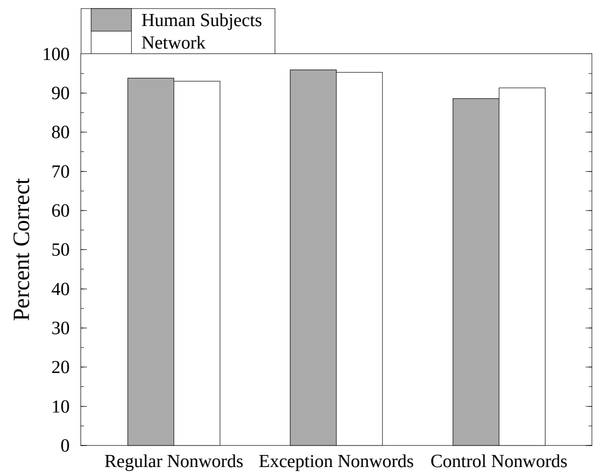

Figure 10: Percent correct performance of the network and of human subjects in pronouncing three sets of nonwords: regular and exception nonwords (*N* = 43 each) from Glushko (1979), and control nonwords (*N* = 80) from McCann and Besner (1987).

a set of pseudohomophones (McCann & Besner, 1987). As before, a nonword pronunciation was considered correct if it was consistent with some word in the training corpus. Figure 10 shows the performance of the network on this criterion, as well as the corresponding data for human subjects. The network was correct on 40/43 (93.0%) of the regular nonwords, 41/43 (95.3%) of the exception nonwords, and 73/80 (91.3%) of the control nonwords. By comparison, the corresponding levels of performance reported for human subjects were 93.8% on regular nonwords and 95.9% on exception nonwords (Glushko, 1979), and 88.6% on the control nonwords (McCann & Besner, 1987). Moreover, in pronouncing these nonwords, the mean number of fixations produced by the network for correct pronunciations was 1.63 for the regular nonwords, 2.27 for the exception nonwords, and 1.92 for the control nonwords. The overall mean for nonwords, 1.94, is comparable to the value for low-frequency exception words (2.00; see Figure 9). Thus, the network's nonword reading accuracy and latency is comparable to that of skilled readers.

Performance Under a Peripheral Impairment. In order to model a peripheral deficit in letter perception of the sort postulated by Behrmann, Plaut, and Nelson (1998b) to produce LBL reading, input letter activations were corrupted by Gaussian noise (*SD* = 0.055). When this was done, correct performance dropped from 99.3% to 90.0% correct (averaged across 10 runs through the training corpus). Using a median split on frequency, accuracy was greater on high- versus low-frequency words (91.7% vs. 88.7%, respectively; $F_{1,2983} = 18.0$, $p < .001$) and on short versus long words (e.g., 91.6% for 4-letter words vs. 86.8% for 6-letter words; $F_{1,1523} = 14.1$, $p < .001$).

It was argued above that number of fixations can be used as a coarse approximation to naming latency for skilled readers because this measure reflects the degree of difficulty in constructing a coherent articulatory output. The situation is rather different in the context of LBL reading because, in this case, it is more literally true that a pronunciation is constructed incrementally. For this reason, number of fixations in the model can be taken as a more direct analogue of the naming latency of LBL readers. Another plausible measure—the total number of processing steps required by the model in generating a pronunciation, including initial attempts and attempts after refixations—gives qualitatively equivalent results.

Among words pronounced correctly, the average number of fixations per word increased from 1.32 to 2.20 as a result of the introduction of input noise. Not surprisingly, this measure was strongly influenced by the length of the word. For example, the impaired model made an average of 2.00 fixations on 4-letter words but 2.97 fixations on 6-letter words ($F_{1,1522} = 380.1$, $p < .001$), corresponding to a slope of 0.49 fixations per letter. The model also made fewer fixations on high- versus low-frequency words (means 2.10 vs. 2.30, respectively; $F_{1,2973} = 50.5$, $p < .001$). Finally, and most important for the Behrmann, Plaut, and Nelson (1998b) account of LBL reading, there was a clear interaction of frequency and length. This was established by comparing performance on sets of 4- and 6-letter words matched for frequency (*N* = 100 for each cell). The average number of fixations per word for these stimuli is shown in Figure 11. In addition to main effects of frequency ($F_{1,396} = 7.13$, $p < .01$) and

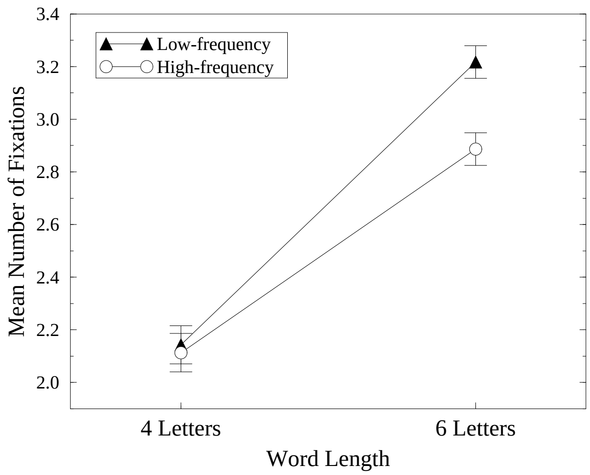

Figure 11: Mean number of fixations made by the model in pronouncing 4- and 6-letter words as a function of their frequency.

length ($F_{1,396}$ = 186.6, $p < .001$), frequency interacted with length such that the effect of frequency was larger for 6- than for 4-letter words ($F_{1,396}$ = 4.96, $p < .05$). Thus, under peripheral damage, the network exhibited the hallmark word length effect characteristic of LBL reading, combined with the appropriate higher-level effects: a word frequency effect which was greater for long compared with short words.[7]

In summary, a simple recurrent network was presented with words as letter strings over position-specific units and was trained to generate the pronunciation of the word in the form of a sequence of phonemes. The model had the ability to refixate the input string when encountering difficulty. The network learned to pronounce correctly virtually all of the 2998-word training corpus, including both regular and exception words, and also was capable of pronouncing nonwords as well as skilled readers. Moreover, if mean number of fixations was taken as an analogue of skilled naming latency, the model exhibited a length effect as well as the standard frequency-by-consistency interaction observed in empirical studies. Finally, peripheral damage to the model, in the form of corrupted letter activations, gave rise to the hallmark characteristics of letter-by-letter reading, including an increased length effect that interacts with lexical variables (e.g., word frequency).

[7]Given that the network contains no semantic representations, it cannot be used to account for the effects of imageability on LBL reading, nor the relatively preserved lexical decision and semantic categorization performance of these patients.

5 General Discussion

Connectionist modeling has made important contributions to a wide range of domains within cognitive science. Word reading, in particular, has received considerable attention because it is a highly learned skill that involves the rapid, online interaction of a number of sources of information in an integrated fashion. There is also a wealth of detailed empirical data on normal reading acquisition and skilled performance, as well as patterns of reading impairments in developmental and acquired dyslexia, that play an essential role in evaluating and constraining explicit computational models. The current chapter contributes to the development of a connectionist theory of normal and impaired word reading based on three general computational principles: distributed representation, gradual learning of statistical structure, and interactivity in processing. This endeavor has led to a number of important insights concerning the nature of the reading system, both in normal operation and when impaired by brain damage. These insights do not typically follow from alternative theoretical frameworks, although versions of them can be incorporated into these frameworks in a post hoc manner. Moreover, many of the insights have implications which extend beyond the specific domain of word reading. Four of these are enumerated and discussed below.

1. The apparent dichotomy between "regular" versus "exception" items is a false one; rather, items vary along a continuum of *consistency* (Glushko, 1979), and a single mechanism can learn to process all types of items and yet also generalize effectively to novel items.

This point was made first by Rumelhart and McClelland (1986) in the domain of inflectional morphology, and later by Seidenberg and McClelland (1989) in the domain of word reading. The impact of these early models was, however, undermined to a certain extent by limitations in the models' performance, particularly with respect to generalization. In the domain of word reading, these limitations were addressed in subsequent modeling work (Plaut et al., 1996) by incorporating more appropriately structured orthographic and phonological representations.

Apart from issues of parsimony, the importance of a single-mechanism account is that it provides insight into why there is so much shared structure between so-called regular and exception items. For instance, the exception word PINT has regular correspondences for the P, N, and T, and even the exceptional I receives a pronunciation that it adopts in many other words (e.g., PINE, DIE). Moreover, nonword pronunciation is influenced by exception as well as regular neighbors (Glushko, 1979). Accounts which invoke separate mechanisms for the regular versus exceptional aspects of language fail to explain or capitalize on this shared structure.

2. Skilled performance is supported by the integration of multiple sources of information; impaired performance following brain damage can reflect the underlying division-of-labor among these sources in the premorbid system.

Patients with fluent surface dyslexia exhibit relatively normal reading of regular words and nonwords but produce "regularization" errors to many exception words, particularly those of low frequency. Dual-route theories explain surface dyslexia as partial damage to the lexical (non-semantic) route that impairs low- more than high-frequency words, with the spared regular and nonword reading supported by the undamaged nonlexical route. There is, however, no explanation for why the lexical damage is always *partial*—the architecture provides equally well for complete elimination of the lexical route with complete sparing of the nonlexical route. This would yield an inability to pronounce any exception words with complete sparing of regular words and nonwords—a pattern that has never been observed empirically. As exception word reading becomes very severely impaired, regular word (and nonword) reading invariably begins to suffer (see Patterson et al., 1996).

By contrast, Plaut et al. (1996) provide an account of surface dyslexia in which it is impossible to eliminate exception word reading without also impairing performance on regular words and nonwords. The reason is that normal performance is supported by the combination of both the phonological and semantic pathways. The contribution from semantics relieves the phonological pathway of having to learn to pronounce all types of words by itself. Rather, it becomes fully adequate only at those aspects of the task for which it is well suited: processing items which are either high in frequency and/or in spelling-sound consistency (see Equation 1). Low-frequency exception words are processed to some degree but typically require additional support from the semantic pathway to be pronounced correctly. Semantic damage, then, reveals the limitations of the undamaged phonological pathway, which manifest as surface dyslexia. Even complete elimination of semantics spares many exception words, particularly those with high frequency. The only way to completely eliminate exception word reading is to damage the phonological pathway as well, but this also impairs regular word and nonword reading (as observed empirically).

3. The co-occurrence of different types of errors can arise from single lesions within a distributed system that learns to map among the different types of information.

The error patterns of brain-damaged patients can place strong constraints on theoretical accounts of cognitive processes. The traditional account of the co-occurrence of visual and semantic errors in deep dyslexia (Morton & Patterson, 1980) assumes an impairment to visual access of (abstract) semantics to explain the visual errors, and a second impairment to semantic access of phonology to explain the semantic errors. The problem is that this account explains the occurrence of visual errors and of semantic errors, but not their *co-occurrence*: it is perfectly feasible within the framework to introduce only one of the lesions—say, the second—and predict patients who produce only semantic errors. While such cases have been reported (e.g., KE; Hillis, Rapp, Romani, & Caramazza, 1990), the vast majority of deep dyslexic patients make both visual and semantic errors (see Coltheart, Patterson, & Marshall, 1987), and the traditional account fails to explain this. An appeal to chance anatomic proximity of the related brain structures fails because the co-occurrence is not symmetric; many dyslexic patients make visual errors but no semantic errors.

On the connectionist account (Hinton & Shallice, 1991; Plaut & Shallice, 1993), the co-occurrence of visual errors with semantic errors is a natural consequence of the nature of learning within a distributed attractor network that maps orthography to semantics. Essentially, the layout of attractor basins must be sensitive to both visual and semantic similarity, and so these metrics are reflected in the types of errors that occur as a result of damage.

4. A double dissociation in performing two tasks does not implicate separate modules dedicated to performing each of the tasks, but can arise from graded functional specialization with a distributed system that performs both tasks.

Cognitive neuropsychologists have traditionally used double dissociations as a means of inferring the structure of the cognitive system (Teuber, 1955). If each of two tasks can be selectively impaired by brain damage while leaving the other relatively intact, it seems reasonable to assume that the two tasks are subserved by separate mechanisms. Unfortunately, this logic is often applied under the assumption that the cognitive system is composed of a set of distinct modules, but various types of nonmodular systems can also give rise to double dissociations (for discussion, see Farah, 1994; Shallice, 1988).

As a case in point, deep dyslexic patients are much worse at reading aloud abstract words compared with concrete words, whereas concrete-word dyslexic CAV (Warrington, 1981) showed the reverse pattern. This double dissociation prompted Warrington and others (e.g., Morton & Patterson, 1980) to assume that the semantics for abstract words was represented separately from those for concrete words. By contrast, Plaut and Shallice (1993, also see Plaut, 1995) developed an extension of the Hinton and Shallice (1991) deep dyslexia simulation in which there is no separation of the representations and processes subserving abstract and concrete word reading. The network does, however, develop stronger attractors for concrete words because they have much richer semantic representations (i.e., many more semantic features). This difference leads to a degree of functional specialization in the system. Damage between orthography and phonology produces a greater impairment on abstract words because these items benefit much less from the clean-up provided by the semantic attractors. Severe damage to sets of connections that implement these attractors, by contrast, impairs concrete words the most because they have come to rely on the clean-up, whereas many abstract words can be read without this support. Thus, the abstract-concrete double dissociation does reveal something important about the underlying organization of the system, but this organization does not correspond directly to the empirically manipulated stimulus dimension (concreteness).

The above four points illustrate ways in which a distributed connectionist approach has provided new insights both normal and impaired word reading. It must be acknowledged, however, that the existing implemented models have a number of basic limitations that ultimately prevent them from collectively constituting a comprehensive account of the domain. These limitations stem largely from the fact that all of them have very restricted temporal behavior: Single static monosyllabic words are presented as input, and a single, static semantic and/or phonological pattern is generated as output. Naturalistic reading is, of course, a far more fluid and temporally complex activity, involving sequences of attentional shifts and eye movements over lines of text as input, sequences of articulatory gestures as spoken output, and interactions among multiple levels of linguistic structure in both comprehension and production (see Just & Carpenter, 1987).

The current chapter presents a simulation which can be seen as a first step towards incorporating some of these complexities into connectionist models of reading. The model is still applied only to single monosyllabic words, but this limitation reflects more the choice of training corpus than any intrinsic limitation of the architecture. The network generates sequences of phonemes as output in response to letter strings as input. Critically, it maintains a focus of attention within the word as it is being pronounced; this focus is used to refixate the input string when the network encounters difficulty in generating a pronunciation. The model learned to pronounce virtually all of the 2998-word training corpus, and pronounced nonwords as well as skilled readers. It also exhibited a length effect and the standard interaction of word frequency and spelling-sound consistency if the number of fixations it makes in pronouncing a word was taken to reflect its naming latency.

Consideration of sequential processing for both visual input and articulatory output is critical for a full account of a number of empirical phenomena, particularly those related to the effects of the length of the input string. The current model is applied only to a small subset of these effects, relating to differential effects for words versus nonwords (Weekes, 1997), and the exaggerated length effect of letter-by-letter readers and its interaction with lexical variables (Behrmann et al., 1998b). In the latter case, the empirical adequacy of the model is somewhat limited in that the magnitude of the length effects, relative to normal performance, are much smaller than for most letter-by-letter readers. Nonetheless, the model illustrates how letter-by-letter reading can be interpreted as reflecting the operation of the normal reading system following peripheral damage (see Behrmann, Plaut, & Nelson, 1998b, for discussion).

Given that the current model is, in many respects, very different from previous models (Plaut et al., 1996; Seidenberg & McClelland, 1989), it is important to consider how they are related. With regard to the orthographic input, the models are relatively similar in that all of them are presented with an entire word as input. The current model differs in the use of position-specific letter units and a refixation mechanism. However, most words are processed in a single fixation in skilled performance, which corresponds to the static presentation of input in the previous models. In this way, even though the current model produces a single phoneme at a time, the fact that it does so based on the entire orthographic input at every step makes it fully consistent with evidence suggesting a considerable degree of parallel visual processing during word reading (see, e.g., Reichle et al., 1998). This property also distinguishes it from other sequential models in which the orthographic input is shifted leftward one letter each time a phoneme is generated (e.g., Bullinaria, 1997; Sejnowski & Rosenberg, 1987). In fact, these models are very similar to the current model when it is refixating every grapheme.

The more substantial difference between the model and the previous parallel ones concerns the generation of phonological output. The previous models generated a static representation of the pronunciation of an entire (monosyllabic) word, whereas the current model generates a pronunciation phoneme-by-phoneme. An intermediate case would be a model which derived a representation of an entire word (or at least a syllable) and then used this representation as input to generate sequential articulatory output. Plaut and Kello (1998) describe such a system in the context of modeling phonological development, although the phonological representation is generated from acoustic rather than orthographic input. A reading model which adopted the current model's treatment of orthographic input but Plaut and Kello's treatment of articulatory output would combine the strengths of the current sequential model and previous parallel models, and should be able to model effects on naming latencies, including those relating to orthographic length, directly in its temporal behavior. While such an approach appears promising for addressing the full range of empirical phenomena in normal and impaired word reading, it remains for future work to bring it to fruition.

6 Further Readings

Sejnowski and Rosenberg's (1987) NETtalk model was one of the first attempts to apply connectionist networks to realistic tasks. The subsequent highly influential modeling work by Seidenberg and McClelland (1989) was more psychologically oriented, making detailed contact with specific patterns of empirical data. Plaut, McClelland, Seidenberg, and Patterson (1996) elaborated the approach taken by Seidenberg and McClelland by developing models that provided a better match to some empirical findings and by providing a more systematic treatment of the computational principles underlying the approach. Other recent connectionist models of word reading include Bullinaria (1997), Zorzi, Houghton, and Butterworth (1998), and Harm (1998). The most influential non-connectionist implementation of word reading is the Dual-Route Cascaded (DRC) model of Coltheart, Curtis, Atkins, and Haller (1993).

The above models focus largely on the mapping from print to sound; Hinton and Shallice (1991) carried out an important investigation of how networks can be applied to the task of mapping print to meaning. This work was followed up extensively by Plaut and Shallice (1993).

For background on some of the relevant empirical phenomena in normal and impaired word reading, see the following: normal skilled reading (Balota, 1994), surface dyslexia (Patterson, Coltheart, & Marshall, 1985), deep dyslexia (Coltheart, Patterson, & Marshall, 1980), phonological dyslexia (Coltheart, 1996), length effects (Henderson, 1982), and pure alexia/letter-by-letter reading (Coltheart, 1998).

7 Acknowledgments

This work was supported financially by the National Institute of Mental Health (Grants MH47566 and MH55628). The simulation was run on a Silicon Graphics 4D/440 Challenge with eight R4400/150MHz CPUs, using an extended version of the Xerion simulator developed by Tony Plate, Drew van Camp, and Geoff Hinton at the University of Toronto. Some basic aspects of the simulation were first reported in Plaut, McClelland, and Seidenberg (1995). I thank Dave Balota, Marlene Behrmann, John Bullinaria, Jay McClelland, Karalyn Patterson, Mark Seidenberg, and Tim Shallice for helpful comments and discussions.

Correspondence should be addressed to David C. Plaut, Mellon Institute 115–CNBC, 4400 Fifth Avenue, Pittsburgh, PA 15213–2683, plaut@cmu.edu.

References

Balota, D. A. (1994). Visual word recognition: The journey from features to meaning. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 303–358). New York: Academic Press.

Balota, D. A., & Spieler, D. H. (1998). The utility of item-level analyses in model evaluation: A reply to Seidenberg and Plaut. *Psychological Science, 9*, 238–240.

Beauvois, M.-F., & Derouesné, J. (1979). Phonological alexia: Three dissociations. *Journal of Neurology, Neurosurgery, and Psychiatry*, *42*, 1115–1124.

Behrmann, M., Barton, J. J., Shomstein, S., & Black, S. E. (1999). *Eye movements reveal the sequential processing in letter-by-letter reading.* Manuscript submitted for publication.

Behrmann, M., & Bub, D. (1992). Surface dyslexia and dysgraphia: Dual routes, a single lexicon. *Cognitive Neuropsychology*, *9*, 209–258.

Behrmann, M., Nelson, J., & Sekuler, E. (1998a). Visual complexity in letter-by-letter reading: "pure" alexia is not pure. *Neuropsychologia*, *36*, 1115–1132.

Behrmann, M., Plaut, D. C., & Nelson, J. (1998b). A literature review and new data supporting an interactive account of letter-by-letter reading. *Cognitive Neuropsychology*, *15*, 7–51.

Besner, D., Twilley, L., McCann, R. S., & Seergobin, K. (1990). On the connection between connectionism and data: Are a few words necessary? *Psychological Review*, *97*, 432–446.

Bub, D., Cancelliere, A., & Kertesz, A. (1985). Whole-word and analytic translation of spelling-to-sound in a non-semantic reader. In K. Patterson, M. Coltheart, & J. C. Marshall (Eds.), *Surface dyslexia* (pp. 15–34). Hillsdale, NJ: Erlbaum.

Bullinaria, J. A. (1997). Modeling reading, spelling, and past tense learning with artificial neural networks. *Brain and Language*, *59*, 236–266.

Cipolotti, L., & Warrington, E. K. (1995). Semantic memory and reading abilities: A case report. *Journal of the International Neuropsychological Society*, *1*, 104–110.

Coltheart, M. (Ed.). (1996). Special issue on Phonological Dyslexia. *Cognitive Neuropsychology*, *13*, 749–934.

Coltheart, M. (Ed.). (1998). Special issue on Pure Alexia. *Cognitive Neuropsychology*, *15*, 1–???

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review*, *100*, 589–608.

Coltheart, M., Patterson, K., & Marshall, J. C. (Eds.). (1980). *Deep dyslexia.* London: Routledge & Kegan Paul.

Coltheart, M., Patterson, K., & Marshall, J. C. (1987). Deep dyslexia since 1980. In M. Coltheart, K. Patterson, & J. C. Marshall (Eds.), *Deep dyslexia* (pp. 407–451). London: Routledge & Kegan Paul, 2 edition.

Coslett, H. B., & Saffran, E. M. (1989). Evidence for preserved reading in "pure alexia". *Brain*, *112*, 327–359.

Dejerine, J. (1892). Contribution à l'étude anatomoclinique et clinique des differentes variétés de cécité verbale. *Mémoires del la Société de Biologie*, *4*, 61–90.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*, 179–211.

Farah, M. J. (1994). Neuropsychological inference with an interactive brain: A critique of the locality assumption. *Behavioral and Brain Sciences*, *17*, 43–104.

Fera, P., & Besner, D. (1992). The process of lexical decision: More words about a parallel distributed processing model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 749–764.

Friedman, R. B. (1996). Recovery from deep alexia to phonological alexia. *Brain and Language*, *52*, 114–128.

Glosser, G., & Friedman, R. B. (1990). The continuum of deep/phonological alexia. *Cortex*, *26*, 343–359.

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.

Harm, M. W. (1998). *Division of labor in a computational model of visual word recognition.* PhD thesis, Department of Computer Science, University of Southern California, Los Angeles, CA.

Henderson, L. (1982). *Orthography and word recognition in reading.* London: Academic Press.

Hillis, A. E., Rapp, B., Romani, C., & Caramazza, A. (1990). Selective impairments of semantics in lexical processing. *Cognitive Neuropsychology*, *7*, 191–243.

Hinton, G. E. (1989a). Connectionist learning procedures. *Artificial Intelligence*, *40*, 185–234.

Hinton, G. E. (1989b). Deterministic Boltzmann learning performs steepest descent in weight-space. *Neural Computation*, *1*, 143–150.

Hinton, G. E., & Shallice, T. (1991). Lesioning an attractor network: Investigations of acquired dyslexia. *Psychological Review*, *98*, 74–95.

Jones, G. V. (1985). Deep dyslexia, imageability, and ease of predication. *Brain and Language*, *24*, 1–19.

Just, M. A., & Carpenter, P. A. (1987). *The psychology of reading and language comprehension.* Boston, MA: Allyn and Bacon Inc.

Kawamoto, A. H., Kello, C., Jones, R., & Bame, K. (1998). Initial phoneme versus whole word criterion to initiate pronunciation: Evidence based on response latency and initial phoneme duration. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 862–885.

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English.* Providence, RI: Brown University Press.

Lambon Ralph, M., Ellis, A. W., & Franklin, S. (1995). Semantic loss without surface dyslexia. *Neurocase*, *1*, 363–369.

McCann, R. S., & Besner, D. (1987). Reading pseudohomophones: Implications for models of pronunciation and the locus of the word-frequency effects in word naming. *Journal of Experimental Psychology: Human Perception and Performance*, *13*, 14–24.

McCarthy, R., & Warrington, E. K. (1986). Phonological reading: Phenomena and paradoxes. *Cortex*, *22*, 359–380.

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

Morton, J. (1969). The interaction of information in word recognition. *Psychological Review*, *76*, 165–178.

Morton, J., & Patterson, K. (1980). A new attempt at an interpretation, Or, an attempt at a new interpretation. In M. Coltheart, K. Patterson, & J. C. Marshall (Eds.), *Deep dyslexia* (pp. 91–118). London: Routledge & Kegan Paul.

Mozer, M. C. (1983). Letter migration in word perception. *Journal of Experimental Psychology: Human Perception and Performance*, *9*, 531–546.

Norris, D. (1994). A quantitative multiple-levels model of reading aloud. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 1212–1232.

O'Regan, K. (1981). The "convenient viewing position" hypothesis. In D. F. Fisher, R. A. Monty, & J. W. Senders (Eds.), *Eye movements: Cognition and visual perception* (pp. 289–298). Hillsdale, NJ: Erlbaum.

Patterson, K., Coltheart, M., & Marshall, J. C. (Eds.). (1985). *Surface dyslexia.* Hillsdale, NJ: Erlbaum.

Patterson, K., & Hodges, J. R. (1992). Deterioration of word meaning: Implications for reading. *Neuropsychologia*, *30*, 1025–1040.

Patterson, K., & Kay, J. (1982). Letter-by-letter reading: Psychological descriptions of a neurological syndrome. *Quarterly Journal of Experimental Psychology*, *34A*, 411–441.

Patterson, K., Plaut, D. C., McClelland, J. L., Seidenberg, M. S., Behrmann, M., & Hodges, J. R. (1996). Connections and disconnections: A connectionist account of surface dyslexia. In J. Reggia, R. Berndt, & E. Ruppin (Eds.), *Neural modeling of cognitive and brain disorders* (pp. 177–199). New York: World Scientific.

Peterson, C., & Anderson, J. R. (1987). A mean field theory learning algorithm for neural nets. *Complex Systems*, *1*, 995–1019.

Plaut, D. C. (1995). Double dissociation without modularity: Evidence from connectionist neuropsychology. *Journal of Clinical and Experimental Neuropsychology*, *17*, 291–321.

Plaut, D. C. (1997). Structure and function in the lexical system: Insights from distributed models of naming and lexical decision. *Language and Cognitive Processes*, *12*, 767–808.

Plaut, D. C., & Kello, C. T. (1998). The interplay of speech comprehension and production in phonological development: A forward modeling approach. In B. MacWhinney (Ed.), *The emergence of language*. Mahweh, NJ: Erlbaum.

Plaut, D. C., & McClelland, J. L. (1993). Generalization with componential attractors: Word and nonword reading in an attractor network. In *Proceedings of the 15th Annual Conference of the Cognitive Science Society* (pp. 824–829). Hillsdale, NJ: Erlbaum.

Plaut, D. C., McClelland, J. L., & Seidenberg, M. S. (1995). Reading exception words and pseudowords: Are two routes really necessary? In J. P. Levy, D. Bairaktaris, J. A. Bullinaria, & P. Cairns (Eds.), *Connectionist models of memory and language* (pp. 145–159). London: UCL Press.

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

Plaut, D. C., & Shallice, T. (1993). Deep dyslexia: A case study of connectionist neuropsychology. *Cognitive Neuropsychology*, *10*, 377–500.

Rastle, K., & Coltheart, M. (1998). Whammies and double whammies: The effect of length on nonword reading. *Psychonomic Bulletin & Review*, *5*, 277–282.

Reggia, J. A., Marsland, P. M., & Berndt, R. S. (1988). Competitive dynamics in a dual-route connectionist model of print-to-sound transformation. *Complex Systems*, *2*, 509–547.

Reichle, E. D., Pollatsek, A., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, *105*, 125–157.

Rumelhart, D. E., & McClelland, J. L. (1986). On learning the past tenses of English verbs. In J. L. McClelland, D. E. Rumelhart, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 2: Psychological and biological models* (pp. 216–271). Cambridge, MA: MIT Press.

Saffran, E. M., Bogyo, L. C., Schwartz, M. F., & Marin, O. S. M. (1980). Does deep dyslexia reflect right-hemisphere reading? In M. Coltheart, K. Patterson, & J. C. Marshall (Eds.), *Deep dyslexia* (pp. 381–406). London: Routledge & Kegan Paul.

Schwanenflugel, P. J. (1991). Why are abstract concepts hard to understand? In P. J. Schwanenflugel (Ed.), *The psychology of word meanings.* Hillsdale, NJ: Erlbaum.

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

Seidenberg, M. S., & Plaut, D. C. (1998). Evaluating word reading models at the item level: Matching the grain of theory and data. *Psychological Science*, *9*, 234–237.

Seidenberg, M. S., Plaut, D. C., Petersen, A. S., McClelland, J. L., & McRae, K. (1994). Nonword pronunciation and models of word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 1177–1196.

Seidenberg, M. S., & Waters, G. S. (1989). Word recognition and naming: A mega study [Abstract 30]. *Bulletin of the Psychonomic Society*, *27*, 489.

Sejnowski, T. J., & Rosenberg, C. R. (1987). Parallel networks that learn to pronounce English text. *Complex Systems*, *1*, 145–168.

Shallice, T. (1988). *From neuropsychology to mental structure*. Cambridge: Cambridge University Press.

Shallice, T., & McCarthy, R. (1985). Phonological reading: From patterns of impairment to possible procedures. In K. Patterson, M. Coltheart, & J. C. Marshall (Eds.), *Surface dyslexia* (pp. 361–398). Hillsdale, NJ: Erlbaum.

Shallice, T., & Saffran, E. (1986). Lexical processing in the absence of explicit word identification: Evidence from a letter-by-letter reader. *Cognitive Neuropsychology*, *3*, 429–458.

Shallice, T., & Warrington, E. K. (1980). Single and multiple component central dyslexic syndromes. In M. Coltheart, K. Patterson, & J. C. Marshall (Eds.), *Deep dyslexia* (pp. 119–145). London: Routledge & Kegan Paul.

Spieler, D. H., & Balota, D. A. (1997). Bringing computational models of word naming down to the item level. *Psychological Science*, *8*, 411–416.

Strain, E., Patterson, K., & Seidenberg, M. S. (1995). Semantic effects in single-word naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 1140–1154.

Taraban, R., & McClelland, J. L. (1987). Conspiracy effects in word recognition. *Journal of Memory and Language*, *26*, 608–631.

Teuber, H. L. (1955). Physiological psychology. *Annual Review of Psychology*, *9*, 267–296.

Van Orden, G. C., Pennington, B. F., & Stone, G. O. (1990). Word identification in reading and the promise of subsymbolic psycholinguistics. *Psychological Review*, *97*, 488–522.

Warrington, E. K. (1981). Concrete word dyslexia. *British Journal of Psychology*, *72*, 175–196.

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

Zorzi, M., Houghton, G., & Butterworth, B. (1998). Two routes or one in reading aloud? A connectionist "dual-process" model. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 1131–1161.