Parallel Distributed Processing

Implications for Psychology and Neurobiology

Edited by
R. G. M. MORRIS
University of Edinburgh

CLARENDON PRESS · OXFORD
1989

7

Connections and disconnections: acquired dyslexia in a computational model of reading processes

KARALYN PATTERSON, MARK S. SEIDENBERG, and JAMES L. McCLELLAND

Introduction

In this chapter we describe a new, parallel distributed processing (PDP) model of visual word recognition and pronunciation, the acquisition of these skills, and their breakdown following brain injury. The model consists of a working, computational simulation of the process of learning to recognize and pronounce written words. In developing this model we were motivated by two general concerns. The first is that, since word recognition is a key component of reading, a comprehensive account of word recognition is critical to an understanding of this important human cognitive skill. A basic characteristic of reading comprehension is that it occurs 'on-line', i.e. essentially as the stimulus is perceived. This characteristic derives in part from the fact that words are recognized rapidly and usually effortlessly; a large amount of research has addressed the types of knowledge and processes that support this capacity, the kinds of information that become available as part of the recognition process, and how this information contributes to other aspects of reading. Furthermore, word recognition presents important developmental issues; learning to read words is among the first tasks confronting the beginning reader, and problems in reading acquisition are typically associated with deficits in this skill. Finally, reading impairments that are a consequence of brain injury are often associated with deficits in word recognition; studies of these acquired forms of dyslexia have provided important evidence concerning the reading process and its neurological realization. As reading researchers, one of our primary goals was to develop a computational model that incorporates much of what is known about these aspects of word recognition.

The other primary motivation for this work was the observation that word recognition provides a domain in which to explore the properties of the connectionist or parallel distributed processing (PDP) approach to under-

standing human cognition. This approach represents the modern realization of Hebb's (1949) idea that complex human behaviours emerge from the operation of aggregations of simple neuronal processing units. The approach has generated broad interest among cognitive- and neuro-scientists, and has been applied to a wide range of problems in perception, learning, and cognition (e.g. McClelland and Rumelhart 1986; Rumelhart and McClelland 1986*a*). The first generation of connectionist models illustrated the basic principles and the potential of this approach, but were limited in scope. As a relatively mature area of research, word recognition presented a domain in which to develop a second-generation model capable of simulating a broad range of behavioural phenomena in detail. Such a comprehensive model would provide a basis for assessing the value of the connectionist approach in the development of explanatory theories.

The plan of the paper is as follows. We first provide an overview of the model, describing its basic structure and operation. We then summarize the model's account of the task of naming words aloud. This material is developed in greater detail elsewhere (Seidenberg 1988*a*; Seidenberg and McClelland 1988*a*,*b*), so our treatment of these issues will necessarily be limited. The main focus of this paper concerns our initial explorations of the model's potential to account for certain reading disorders that are observed following brain injury. Although it is by no means a complete theory of word recognition and pronunciation, the model provides a plausible account of some basic phenomena concerning normal performance; we sought to determine whether aspects of pathological performance could be captured in terms of damage to this system. This work represents one of the first attempts to describe and explain pathological performance following brain injury by 'lesioning' a working computational model of normal performance. Although these studies are as yet preliminary in nature, we think that this effort illustrates the utility and potential of the approach.

Overview of the model

We conceive of a lexical processing module with the general form illustrated in Fig. 7.1. The long-term goal is an integrated theory that accounts for various aspects of lexical processing involving orthographic, phonological, and semantic information. Such a theory would specify how these types of information are represented in memory, and how they are used in tasks such as deriving the meaning of a word from its written form, deriving the spelling of a word from its meaning or its pronunciation, and deriving a pronunciation from spelling. The implemented model, represented by that part of Fig. 7.1 in heavy outline, is concerned with how readers recognize letter strings and pronounce them aloud. The model consists of a network of interconnected

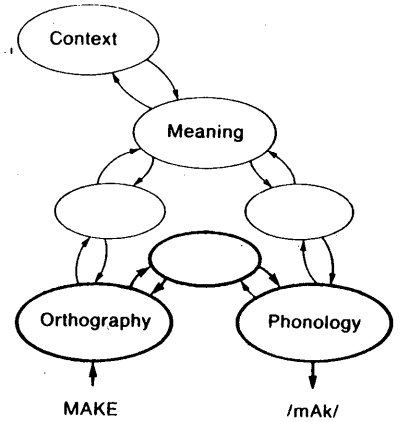

Fig. 7.1 General framework for processing of words in reading: the implemented model is in bold outline.

processing units. There are 400 units used to code orthographic information, 200 hidden units, and 460 units used to code phonological information. There are connections from all orthographic units to all hidden units, and from all hidden units to all phonological units. In addition, there is a set of connections from the hidden units back to the orthographic units. The connections between units carry weights that govern the spread of activation through the system. As will become clear below, these weights encode what the model knows about written English, specifically orthographic redundancy (i.e. the frequency and distribution of letter patterns in the lexicon) and the correspondences between orthography and phonology.

Orthographic and phonological representations

The orthographic and phonological codes for words (and non-words) are represented as patterns of activation distributed over a number of primitive representational units. Each processing unit has an activation value ranging from 0 to 1. The representations of different entities are encoded as different patterns of activity over these units. The details of these representational schemes are described elsewhere (Seidenberg and McClelland 1988*a*,*b*); here we summarize some of their main features.

The phonological representation we employed was the one developed by Rumelhart and McClelland (1986*b*). The phonemes in a letter string are encoded as a set of triples, each specifying a phoneme and its flankers. The

word MAKE, for example, consists of three such triples or 'Wickelphones' (in honor of Wickelgren 1969). The correspondence between Wickelphones and units is one-to-many. Each Wickelphone is encoded as a pattern of activation over a set of units representing phonetic features. Each unit represents a triple of phonetic features, one feature of the first of the three phonemes in each Wickelphone, one feature of the second of the three, and one of the third. For example, there is a unit that represents [vowel, fricative, stop]. This unit should be activated for any word containing a Wickelphone in which this sequence occurs, such as the words POST and SOFT. Word boundaries are also represented in the featural representation, so that there is a unit, for example, that represents [vowel, liquid, word-boundary]; this unit would come on in words like CAR and CALL. In Rumelhart and McClelland's (1986*b*) scheme, there are 460 units and each Wickelphone activates 16 of them (see their paper for discussion).

The representation used at the graphemic level has similarities with that used at the phonological level, but it consists of 400 units set up according to a slightly different scheme. For each unit, there is a table containing a list of 10 possible first letters, 10 possible middle letters, and 10 possible end letters. These tables are generated randomly, except for the constraint that the symbol for beginning/end of word does not occur in the middle position. When the unit is on, it indicates that the string being represented contains one of the 1000 possible triples that could be made by selecting one member from the first list of 10, one from the second, and one from the third. Each letter triple activates about 20 units. Though each unit is highly ambiguous, over the full set of 400 such randomly constructed units, the probability that any two sequences of three letters would activate all and only the same units in common is effectively zero.

In sum, both the phonological and the orthographic representations can be described as coarse-coded, distributed representations of the sort discussed by Hinton, McClelland, and Rumelhart (1986). The representations allow any letter and phoneme sequences to be represented, subject to certain saturation and ambiguity limits that can arise when the strings get too long. Thus, there is a minimum of built-in knowledge of orthographic or phonological structure. The use of a local context-sensitive coding scheme promotes the exploitation of local contextual similarity as a basis for generalization in the model; that is, what the model learns to do for a grapheme in one local context (e.g. the M in MAKE) will tend to transfer to the same grapheme in similar local contexts (e.g. M in MADE and MATE and, to a lesser extent, M in contexts such as MILE and SMALL). Note that we do not claim that these encoding schemes are fully sufficient for representing all of the letter or phoneme sequences that form words (see Pinker and Prince 1988). However, we are presently applying the model only to monosyllables, for which the representation is adequate (see Seidenberg and McClelland, 1988*b*, for discussion).

Processing in the model

The model takes a letter string as input and yields two types of output: (1) a pattern of activation across the phonological units; and (2) a recreation of the input pattern across the orthographic units. The former can be thought of as the model's computation of a phonological code for the input, and will be discussed in some detail because of its relevance to the word naming task. The latter can be considered a representation of the orthographic input in a short-term sensory store and is critical to our account of lexical decision (Seidenberg and McClelland 1988*a*,*b*). Each word-processing trial begins with the presentation of a letter string, which the simulation program then encodes into a pattern of activation over the orthographic units, according to the representational assumptions described above. Next, activations of the hidden units are computed on the basis of the pattern of activation at the orthographic level. For each hidden unit, a quantity termed the net input is computed: this is the activation of each input unit times the weight on the connection from that input unit to the hidden unit, plus a bias term unique to the unit. The bias term may be thought of as an extra weight or connection to the unit from a special unit that always has activation of 1.0. The activation of the hidden unit is then determined from the net input using a non-linear function called the logistic function. The activation function must be non-linear for reasons described in Rumelhart, Hinton, and McClelland (1986). It must be monotonically increasing and have a smooth first derivative for reasons having to do with the learning rule. The logistic function satisfies these constraints.

Once activations over the hidden units have been computed, these are used to compute activations for the phonological units and new activations for the orthographic units based on feedback from the hidden units. These activations are computed following exactly the same procedures already described: first the net input to each unit is calculated, based on the activations of all of the hidden units; then the activation of each of these units is computed, based on the net inputs.

Learning

When the model is first initialized, the connection strengths and biases in the network are assigned random initial values between -0.5 and $+0.5$. This means that each hidden unit computes an entirely arbitrary function of the input it receives from the orthographic units, and sends a random pattern of excitatory and inhibitory signals to the phonological units and back to the orthographic units. This also means that the network has no initial

knowledge of spelling patterns or of correspondences between spelling and sound. Thus, the model is effectively *tabula rasa*; the abilities to re-create the orthographic input and generate its phonological code arise as a result of learning from exposure to letter strings and the corresponding strings of phonemes.

Learning occurs in the model in the following way. An orthographic string is presented and processing takes place as described above, producing first a pattern of activation over the hidden units, then a feedback pattern on the orthographic units and a feedforward pattern on the phonological units. At this point these two output patterns produced by the model are compared to the correct, target patterns that the model should have produced. The target for the orthographic feedback pattern is simply the orthographic input pattern; the target for the phonological output is the pattern representing the correct pronunciation of the presented letter string. A real-world counterpart of this second procedure would be a child seeing a letter string and hearing a teacher or other person say its correct pronunciation.

For each graphemic and phonemic unit, the difference between the correct or target activation of the unit and its actual activation is computed. The learning procedure adjusts the strengths of all of the connections in the network in proportion to the extent to which this change will reduce a measure of the total error, E. This algorithm is the 'back-propagation' learning procedure of Rumelhart, Hinton, and Williams (1986). Readers are referred to Rumelhart *et al.* for an explanation of how the weights are modified. The most important feature is that the rule changes the strength of each weight in proportion to the size of the effect that changing it will have on the error measure. Large changes are made to weights that have a large effect on E, and small changes are made to weights that have a small effect on E.

The training corpus

The model was trained on all of the monosyllabic words consisting of three or more letters in the Kucera and Francis (1967) word count, minus proper nouns, foreign words, abbreviations, and words that are formed by the addition of a final -s or -ed inflection. This is not a complete list of the uninflected monosyllabic words in English; for example, the word FONT is one of many that do not appear in Kucera and Francis. Nevertheless, the corpus provides a reasonable approximation of the set of monosyllables in the vocabulary of an average American reader. To this list we added a number of words that had been used in some of the experiments that we planned to simulate. The resulting corpus contained 2897 words.

The training regime was divided into a series of 250 epochs. In each epoch, each word had a probability of being presented that was a logarithmic function

of its Kucera and Francis frequency. The most frequent word (THE) had a probability of about 0.93; words occurring once per million had probabilities of about 0.05. Thus, the expected value of the number of presentations of a word over 250 epochs ranged from about 230 to about 12. Since the sampling process is in fact random, about 5 per cent of the lowest-frequency items will have occurred less than six times during training.

This sampling method is not intended to mimic the experience of children learning to read in American culture. In the model, all words are available for sampling throughout training, with frequency represented by the probability of selection on a given learning trial. In actual experience, however, frequency derives in part from age at acquisition; words that are higher-frequency for adults tend to be learned earlier by children. Moreover, our treatment of frequency only approximates the differences in familiarity that are relevant to skilled readers, for two reasons. First, there are known inaccuracies in standard frequency norms (Gernsbacher 1984), especially in the lower-frequency range. Second, our encoding of frequency greatly underweights the advantage of higher-frequency words relative to words of lower frequency. In the Kucera and Francis (1967) count, for example, frequencies range from about 70 000 to 1; with the logarithmic compression used in our model, the ratio of highest-frequency word to lowest- is only about 16 to 1.

Characterizing the model's performance

The model produces patterns of activation across the orthographic and phonological units as its output. For word naming, we assume that the pattern over the phonological units serves as the input to a system that constructs an articulatory-motor program, which in turn is executed by the motor system, resulting in an overt pronunciation response. In reality, we believe that these processes operate in a cascaded fashion: the response is triggered when the articulatory-motor program has evolved to the point where it is sufficiently differentiated from other possible motor programs. Thus, activation would begin to build up first at the orthographic units, propagating continuously from there to the hidden and phonological units and from there to the motor system.

The simulation model simplifies this procedure. Activations of the phonological units are computed in a single step, and the construction and execution of articulatory-motor programs are unimplemented. Activations computed in this manner can be shown to correspond to the asymptotic activations that would be achieved in a cascaded process (Cohen, Dunbar, and McClelland 1988). We use the phonological error score—the sum of the squared differences between the target activation value for each phonological unit and the actual activation computed by the network—to relate the model's

performance to experimental data on latency and accuracy of word-naming responses. The error score is a measure of how closely the pattern computed by the net matches the correct pronunciation (or any other specified pronunciation). In general, after training the error score is lower for the correct pronunciation than for any other.

Even though the correct phonological code may be the best match to the pattern of activation over the phonological units, there is still considerable variation in error scores, and we assume that lower error scores are correlated with faster and more accurate responses under time pressure. The rationale for the accuracy assumption is simply that a low error score signifies a pattern produced by the network that is relatively clear and free from noise, providing a better signal on which the articulatory-motor programming and execution processes can operate. The rationale for the speed assumption is that in a cascaded system, patterns that are relatively clear (low in error) at asymptote reach a criterion level of clarity relatively quickly. Simulations demonstrating this point are presented in Cohen, Dunbar, and McClelland (1988).

The error score should not be viewed as a literal measure of the accuracy of an overt response made by the network. The error scores can never actually reach zero, since the logistic function used in setting the activations of units prevents activations from ever reaching their maximum or minimum values. With continued practice, error scores simply get smaller and smaller, as activations of units approximate more and more closely to the target values of 1 and 0. This improvement continues well beyond the point where the correct answer is the best match to the pattern produced by the network.

We also calculate an orthographic error score, analogous to the phonological error score, which provides a measure of the familiarity and redundancy of a letter string. This measure plays an important role in our account of lexical decision performance, but will not be considered further here (see Seidenberg and McClelland 1988*a*,*b*).

In sum, when presented with letter strings, the model produces orthographic and phonological codes which provide the basis for performing tasks such as lexical decision and naming. We characterize the model's performance in terms of error scores calculated for different types of stimuli after different amounts of training, and relate these to human performance on these tasks. Because the model contains such a large pool of words, we can perform very close simulations of many empirical phenomena reported in the literature, often using the identical stimuli as in a particular experiment.

Summary of the model's performance

Seidenberg and McClelland (1988*a*,*b*) describe a broad range of behavioural phenomena simulated by the model. Here we briefly summarize results from

simulations of the task of naming words and non-words aloud. We focus on naming because the acquired forms of dyslexia discussed below are typically associated with impairments on this task. The problem of learning to read single words aloud in English is largely determined by properties of the writing system. The alphabetic writing system for English is a code for representing spoken language; units in the writing system—letters and letter patterns—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). These inconsistencies derive from several sources: there is a competing demand that the orthography preserve morphological information; there are diachronic changes in pronunciation; there is lexical borrowing and historical accident. In fact, the English orthography partially encodes several types of information (orthographic, phonological, syllabic, morphological) simultaneously. Thus, English provides an example of what can be termed a quasiregular system: a body of knowledge that is systematic but admits many exceptions (Seidenberg 1988*a*). In such systems the relationships among entities are statistical rather than categorical.

During the training phase, the model is exposed to a significant fragment of written English. The effect of the learning rule is that the model picks up on facts about orthographic–phonological correspondences and encodes them in terms of the weights on connections between units. Eventually, the weights achieve values that permit the model to produce the correct output for almost any word in the training set, despite the quasiregular character of the writing system. By 'correct' we mean that the error score for the correct pronunciation is typically very much smaller in magnitude than the error score for an incorrect pronuncation. As already mentioned, even when the best fit is the correct phonological code, the size of the error score varies; i.e. the model performs better on some stimuli than on others. How well it performs on a given stimulus depends on factors such as the frequency of the word and its similarity to other words in the corpus. We evaluate the model by comparing its performance on different types of words to that of human subjects.

Consider two classes of words that have been studied in a large number of behavioural experiments. Regular words such as MUST, LIKE, and CANE contain spelling patterns that recur in a large number of words, always with the same pronunciation. MUST, for example, contains the ending -UST; all monosyllabic words that end in this pattern rhyme (JUST, DUST, etc.). The words sharing the critical spelling pattern are termed the neighbours of the input string (Glushko 1979). Neighbours have been primarily defined in terms of word-endings, also termed rimes (Treiman and Chafetz 1987) or bodies

(Patterson and Morton 1985), although other aspects of word structure also matter (Taraban and McClelland 1987; Kay 1987). Exception words such as HAVE, SAID, and LOSE contain a common spelling pattern which in this particular word is pronounced irregularly. That is, since -AVE is usually pronounced as in GAVE and SAVE, the word HAVE is characterized by an exceptional spelling-to-sound correspondence. In terms of orthographic structure, regular and exception words are similar: both contain spelling patterns that recur in many words. Whereas regular words are thought to obey the pronunciation 'rules' of English, exception words do not. Given that these two word classes are similar in orthographic structure, and that they can be equated for other factors such as length and frequency, then differences between them in terms of processing difficulty must be attributed to the one dimension along which they differ, regularity of spelling–sound correspondences.

Studies examining the processing of such words have yielded the following results. First of all, there are frequency effects: higher-frequency words are named more quickly than lower-frequency words. In addition, regularity effects–faster naming latencies for regular words compared to exceptions—are substantial with lower-frequency items, but may be small or non-existent for higher-frequency words (Andrews 1982; Seidenberg *et al.* 1984; Seidenberg 1985*b*; Waters and Seidenberg 1985; Taraban and McClelland 1987). In short, there is a frequency by regularity interaction. In Taraban and McClelland's study, the difference between lower-frequency regular and exception words was a statistically significant 32 ms, while the difference for higher-frequency words was a non-significant 13 ms.

To examine the model's performance on these types of words, we used the identical stimulus set studied by Taraban and McClelland (1987, Experiment 1). Figure 7.2 presents the model's performance on this set of high- and low-frequency regular and exception words after different amounts of training. Each data point represents the mean phonological error score for the 24 items of each type used in the Taraban and McClelland experiment. Training reduces the error scores for all words following a negatively accelerated trajectory. Throughout training, there is a frequency effect: the model performs better on the 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 both frequency classes, regular words produce smaller error scores than exception words. Additional training reduces the regularity effect for higher-frequency words, to the point where it is eliminated by 250 epochs. However, the regularity effect for lower-frequency words remains. Figure 7.3 demonstrates the similarity of results from Taraban and McClelland's adult subjects and from the model.

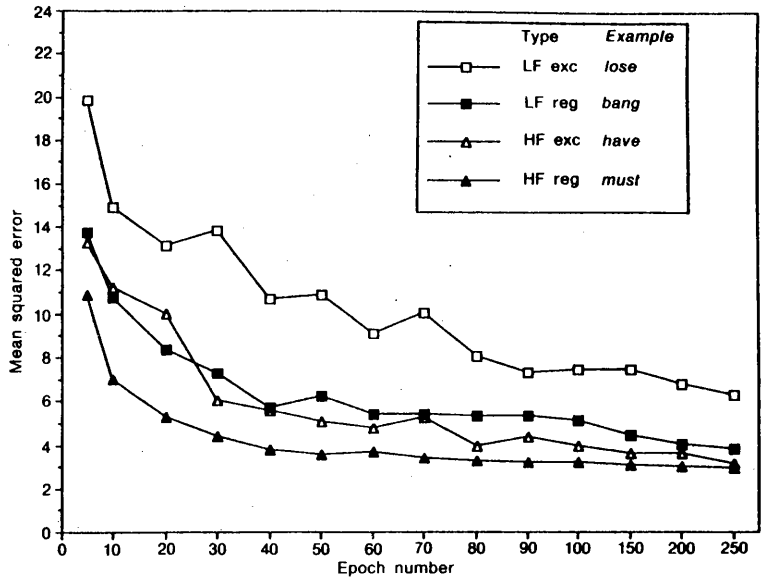

Fig. 7.2 The model's mean phonological error scores at various stages in training for the words used by Taraban and McClelland (1987).

The frequency-by-regularity interactions obtained in two additional studies, with different sets of stimulus words (Seidenberg 1985*b*, Experiment 2; Seidenberg *et al.* 1984*a*, Experiment 3), have been recreated with equal success by the model's performance (see Seidenberg and McClelland 1988*b*). Indeed, following simulations of 14 conditions from eight experiments comparing regular and exception words, Seidenberg and McClelland obtained a correlation of 0.915 between the experimental data (difference in naming latency between regular and exception words) and the model's performance (difference in phonological error score between regular and exception words).

The model is revealing about the behavioural phenomena in two respects. First, it is clear that in the model the frequency by regularity interaction results because the output for both types of higher-frequency words approaches asymptote before the output for the lower-frequency words. Hence the difference between the higher-frequency regular and exception words is eliminated, while the difference between the two types of lower-frequency words remains. This result suggests that the interaction observed in the behavioural data is attributable to a kind of 'floor' effect due to the acquisition of a high level of skill in de-coding common words. In the model, the differences between the two types of lower-frequency words would also

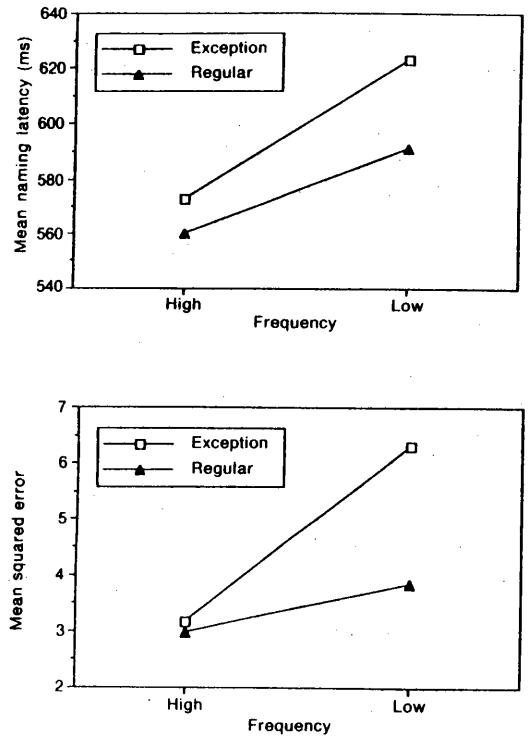

Fig. 7.3 Results of the Taraban and McClelland (1987) study (top panel) and the model's performance at 250 epochs (lower panel)

diminish if training were continued for more epochs. This aspect of the model provides an explanation for Seidenberg's (1985) finding that there are individual differences among skilled readers in terms of regularity effects. The fastest subjects in his study showed no regularity effect, even for words that are 'lower' in frequency according to standard norms. The model suggests that the fastest readers may have encountered lower-frequency words more often than the slower subjects, with the result that these words effectively become 'high-frequency' items.

Second, the model provides a theoretical link between effects of frequency and regularity. Both effects are due to the fact that connections that are required for correct performance have been adjusted more frequently in the required direction for frequent or regular items than for infrequent or irregular items. This holds for frequent words simply because they are presented more often. It holds for regular words because they make use of the same connections as other, neighbouring, regular words. Hence, regularity effects

are frequency effects: both derive from the effects of repeated adjustment of connection weights in the same direction.

Not only in its simulation of frequency and regularity effects but, more generally, the model's performance is determined by the connection weights which reflect the aggregate effects of many individual learning trials with the items in the training set. In effect, learning results in the recreation within the network of significant aspects of the structure of written English. Because the entire set of weights is used in computing the phonological codes for all words, and because all of 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. The main influence on the phonological output is the number of times the model was exposed to the word itself; after a sufficient amount of training, this is the only factor relevant to performance on 'high-frequency' words. Performance on less-frequent words, however, is also affected by exposure to other words. Words that resemble one another in spelling–sound correspondences have mutually beneficial effects on the weights; words that are similar in spelling but dissimilar in pronunciation have mutually inhibitory effects on the weights. Performance is then determined by the cumulative effects of training on the weights.

To see this more clearly, consider the following experiment. We test the model's performance on the low-frequency regular word TINT; with the weights from 250 epochs, it produces an error score of 8.92. We train the model on another word, adjusting the weights according to the learning algorithm, and then re-test TINT. By varying the properties of the training word, we can determine which aspects of the model's experience exert the greatest influence on the weights relative to the target. In effect, we can simulate the phonological priming effects studied by Meyer, Schvaneveldt, and Ruddy (1974), Hillinger (1980), Tanenhaus, Flanigan, and Seidenberg (1980), and others. For example, Meyer *et al.* observed that lexical decision latencies to a target word such as ROUGH were facilitated when preceded by the rhyming prime TOUGH but inhibited when preceded by the similarly spelled non-rhyme COUGH. For the purposes of the simulation, we examined the cumulative effects of a sequence of ten prime (learn)—target (test) trials. The primes were a rhyming orthographic neighbour (MINT), a non-rhyming orthographic neighbour (the exception word, PINT), a word with the same consonants but a different vowel (TENT), and an unrelated control (RASP). The data are presented in Fig. 7.4.

The results indicate, first, that overlap in the ends of words (word-bodies or rimes) has greater impact than overlap in word-beginnings. Thus, priming TINT with MINT has greater impact than priming TINT with TENT (it also has greater impact than priming with a word such as TINS or TILT). The model supports the common assumption that the terminal segments of words

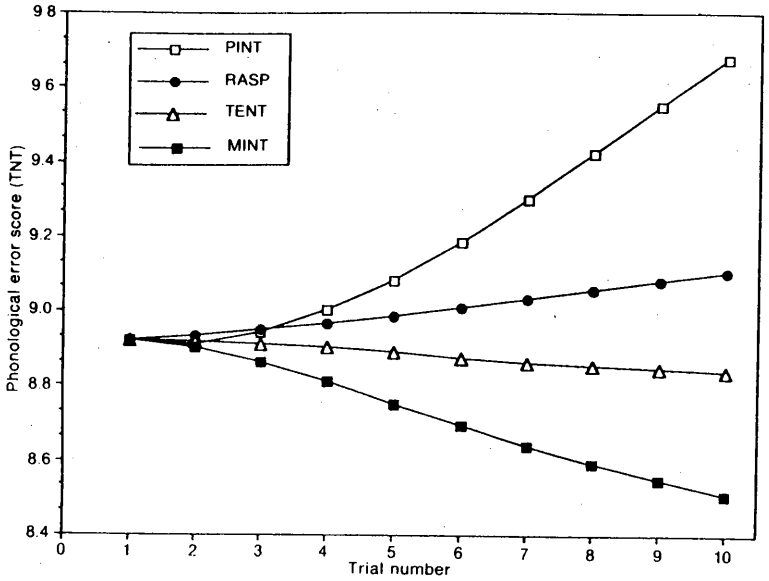

Fig. 7.4 Effects on the phonological error score for TINT of training with MINT, PINT, TENT or RASP.

are especially critical to naming (Glushko 1979; Meyer, Schvaneveldt, and Ruddy 1974; Seidenberg *et al.* 1984*a*; Patterson and Morton 1985; Brown 1987; Treiman and Chafetz 1987). This fact derives from properties of the learning algorithm and the training corpus. Word-bodies turn out to be salient because there is more redundancy at the ends than at the beginnings. The learning algorithm picks up on these regularities, which have a large impact on the weights. Importantly, these same characteristics of the model also dictate that the effective relationships between words are not limited to word-bodies. These units happen to be especially salient, but they are not the only aspects of word structure relevant to processing. Thus, in the priming experiment, both TENT and RASP have small effects on the weights relevant to TINT, as do many other words. Experimental data by Kay (1987) confirm the relevance to naming of neighbourhoods defined over word-initial segments.

The other important point is that the model encodes facts about the consistency of spelling–sound correspondences. Thus, priming TINT with MINT has a large positive effect on the weights, but priming with PINT has complementary negative effects. It is clear, then, why the model performs better on regular words than on exceptions. The model's training on MINT and HINT and LINT and PRINT, matching in both spelling pattern and pronunciation, pushes the values of the weights in the same direction. The

exception word PINT suffers because the weights come to reflect the fact that words ending in -INT typically rhyme with MINT. Having been exposed to PINT and its pronunciation, the model produces the correct phonological code for PINT; however, this code yields a larger error score than for a comparable regular word, owing to the impact of training on the gang of words like MINT.

The impact of the model's experiences during training can be evaluated not only by presenting words from the training vocabulary but by presenting novel words. Non-words have played an important role in experimental considerations of how people convert print to sound, because such stimulus items can be constructed along any dimensions that the experimenter fancies. In a widely cited study, Glushko (1979) demonstrated that readers are quicker to pronounce non-words (like TIFE) derived from a word-body whose neighbourhood has a regular, consistent pronunciation (LIFE, KNIFE, WIFE, etc.) than to pronounce nonwords (like TIVE) with an inconsistent neighbourhood (FIVE v. GIVE). As Seidenberg and McClelland (1988*b*) have shown, the significant 22 ms difference obtained in Glushko's Experiment 2 is mirrored by a significant difference of over two points in the model's phonological error scores for these two types of non-word.

The foregoing description of the model's performance in word naming has focused primarily on regularity effects. This is partly due to the prominence of this issue in the literature on reading over the past decade or so, and partly because regularity effects are particularly germane to certain types of reading disorders, to which we now turn. Before doing so, however, we wish to emphasize that evaluations of the model's naming performance are by no means restricted to the contrast between regular and exception words. Seidenberg and McClelland (1988*a*,*b*) and Seidenberg (1988*a*) present successful simulations of experiments on many other characteristics of words, and the reader is referred to these papers for a picture of the full scope of the model.

Acquired dyslexia

We turn now to questions concerning the impairments of word naming characteristic of certain forms of acquired dyslexia. We have suggested that the model provides a good characterization of a broad range of phenomena related to the naming performance of skilled readers, and that it provides an integrated explanation for these phenomena in terms of the consequences of learning. As a learning model, it also speaks to the issue of how these skills are acquired. Furthermore, the model provides an interesting perspective on the kinds of impairments characteristic of developmental and acquired dyslexias. Developmental dyslexia, which could be seen as a failure to acquire the

knowledge that underlies word recognition and naming, is discussed in Seidenberg and McClelland (1988). Acquired dyslexia, arising from damage to a fully developed normal system, is discussed here.

Acquired dyslexia refers to impairments in reading processes observed following brain injury in people who were previously normal readers. Several different types of acquired dyslexia have been identified, each characterized by impairments to selected aspects of processing (for recent reviews, see Coltheart 1985; Ellis and Young 1988). Many of these impairments relate to the process of naming words aloud; in fact, word-naming performance has provided the primary basis for distinguishing among different varieties of acquired dyslexia. These impairments presumably reflect damage to part(s) of the neural machinery responsible for word recognition and pronunciation. Since our model provides a computational account of some of this machinery, it should be possible to simulate word-naming impairments by selectively damaging the model. In this section we report some preliminary experiments of this sort.

Acquired forms of dyslexia have primarily been discussed in the context of a class of 'dual-route' models. As the name implies, these accounts emphasize the idea that two different procedures or mechanisms are required in order to account for naming performance. The mechanisms are distinguished in terms of the types of knowledge representations involved and the types of letter strings to which these are suited. One mechanism involves rules encoding the reader's knowledge of the correspondences between spelling and pronunciation characteristic of written English. These mapping rules can be used to construct a correct pronunciation of any letter string that obeys them—specifically, regular words such as MUST and regular non-words such as NUST; the rule-based procedure will generate incorrect pronunciations for words that violate the rules (e.g. exceptions such as HAVE). This mechanism has been termed a 'non-lexical' or 'subword-level' process because the rules involve generalizations concerning spelling sound correspondences rather than knowledge of whole specific words. The other mechanism involves stored representations of the pronunciations of known words. The idea here is that the reader identifies a familiar word (directly on the basis of its spelling, and possibly further by consulting its meaning) and then accesses a stored representation of its pronunciation. This mechanism could apply to all known words, but would fail in the case of novel strings such as non-words, which lack representations in memory. This mechanism has been termed a 'lexical' or 'word-level' process because the relevant knowledge representations concern the pronunciations of individual words. Further descriptions of dual-routine accounts of word naming can be found in Patterson, Marshall, and Coltheart (1985).[1]

The major theoretical alternative to the dual-routine model, analogy theory, carved things up slightly differently. Analogy theories proposed by Glushko (1979), Marcel (1980), Humphreys and Evett (1985) and others

contain a single type of knowledge representation relevant to pronunciation: they eliminated the separate rule-based knowledge about correspondences between graphemes and phonemes, leaving only the lexical representations, which were thought to be employed in naming both words and non-words. As Patterson and Coltheart (1987) noted, however, these models implicitly preserved the distinction between two phonological procedures in naming: while a word could be named by accessing a stored phonological representation, the pronunciation of a non-word still had to be created by segmenting known words and cobbling together the phonology of the individual segments comprising the non-word.

In summary, most theorizing about how readers (of English) translate orthography to phonology has assumed that different naming mechanisms are required for the correct pronunciation of exception words on the one hand and non-words on the other. One of the main contributions of the Seidenberg and McClelland (1988*a*,*b*) model is that it accomplishes this translation process with a single mechanism employing weighted connections between units. All items—regular and irregular, word and non-word—are pronounced using the knowledge encoded in the same sets of connections. This model also differs from dual-routine accounts in that there are no rules specifying the regular spelling–sound correspondences of the language, and there is no lexicon in which the pronunciations of words are listed. The model also differs from proposals by Glushko (1979) and Brown (1987) in that there are no lexical nodes representing individual words and no influences from orthographic neighbours at the time of processing a word. Where the model agrees with these accounts is in regard to the notion that regularity effects result from a conspiracy among known words. In the present model, this conspiracy is realized in the setting of connection strengths. Words with similar spellings and pronunciations produce overlapping, mutually beneficial changes in the connection weights.

Some of the evidence thought to support the distinction between two naming processes came from studies of normal readers pronouncing various types of letter strings. However, this general class of theories perhaps took even greater comfort from the neuropsychological literature. In particular, the patterns of reading performance in two 'varieties' of acquired dyslexia, phonological and surface dyslexia, have been considered to provide crucial evidence. Phonological dyslexic patients (Beauvois and Derouesné 1979; Shallice and Warrington 1980; Patterson 1982) show a dissociation between word and non-word naming; in some cases (e.g. Funnell 1983), the dissociation can be dramatic, with around 90 per cent success on words of any class or length but total failure to read aloud even the simplest non-words. Surface dyslexic patients (Marshall and Newcombe 1973; Shallice and Warrington 1980; Coltheart *et al.* 1983) show a dissociation between regular and exception word naming. Though performance on exception words has

not, in any case thus far recorded, been at zero, once again the dissociation can be substantial: for example, about 90 per cent success on low-frequency regular words compared with 40 per cent on low-frequency exception words (Bub, Cancelliere, and Kertesz 1985) or, distinguishing amongst 'levels' of regularity, around 80 per cent correct on regular words v. 35 per cent on very irregular words (Shallice, Warrington, and McCarthy 1983).

If phonological dyslexia could be considered to reflect almost total disruption of the routine for subword-level translation, and surface dyslexia could be considered to indicate severe disruption of the routine for word-level translation, it is easy to see why these neuropsychological dissociations have been emphasized, nay treasured, by dual-routine theories. Accordingly, they represent a challenge to any theory proposing a single process by which all letter strings, whether regular words, exception words or non-words, are converted from orthography to phonology. Failure to account for these patterns would weaken this proposal, while a demonstration that such dissociations could arise within a truly single-routine theory like the model outlined here would constitute a powerful argument against the need for postulating multiple routines.

We shall have nothing to say about phonological dyslexia because we have only just begun to consider how the model might account for it. The pattern of reading performance observed in surface dyslexia, on the other hand, seems ideally suited to one kind of evaluation of the model. Some of the earliest studied cases of surface dyslexia (e.g. Marshall and Newcombe 1973) appeared to use their impaired oral reading skill to make sense of the printed word, yielding slow responses, multiple responses, and generally poor performance. More interestingly, three recent studies of patients with impaired comprehension (for both speech and reading) reveal (1) a high degree of accuracy in naming regular words and non-words; and (2) word-naming latencies at least within the range of age-matched controls. The description of 'reading without semantics' has been offered for the first of these cases (Shallice, Warrington, and McCarthy 1983) and is equally appropriate for the other two cases (Bub, Cancelliere, and Kertesz 1985; McCarthy and Warrington 1986). Reading without semantics is of course precisely what the Seidenberg and McClelland model does. Therefore it seems highly relevant to an evaluation of the model to ask the following question: after the model has been trained to the high level of successful 'oral reading' performance described earlier, if it is now damaged in various ways, will we observe the characteristics of surface dyslexic reading?

The remainder of this chapter is largely devoted to exploring this question. Before we begin, it may be helpful to have a slightly more expanded description of reading performance by surface dyslexic patients. As emphasized for neuropsychological impairments in general (Caramazza 1986), and for this pattern of impaired reading in particular (Patterson, Marshall, and Coltheart 1985), no two patients are identical and so a syndrome label should

be taken as a loose descriptive device rather than a precise classification. In fact, the naming performance of patients described as surface dyslexics varies greatly. With these caveats in mind, we began by assuming that the following features, observed in what are perhaps the most 'pure' and certainly the best-studied surface dyslexic patients, provided a starting point for our explorations of damage to the model:

1. Most central, and already mentioned, is the patients' significantly greater success in naming of words (like PINE) that have regular, typical spelling-to-sound correspondences than of words (like PINT) with exceptional or atypical spelling-to-sound correspondences.
2. Accuracy in naming non-words can be relatively intact, or at least within the (widely varying: Masterson 1985) range of non-word reading skill shown by normal subjects.
3. At least some surface dyslexic patients' accuracy in word naming mimics a characteristic, discussed earlier, of normal subjects' latencies in word naming: an interaction between frequency and regularity. In the best demonstration of this interaction (by Bub, Cancelliere, and Kertesz 1985), the patient showed an advantage on regular over exception words of 15 per cent for high-frequency words but 50 per cent for low-frequency words.
4. The most common type of reading error is the regularization of an exception word, e.g. PINT—/pint/ rhyming with HINT, COME—/kOm/ rhyming with DOME, etc. (Note: pronunciations will be rendered here not in the international phonetic alphabet but, rather, in terms of the phonemic encoding scheme used in the model, taken from Rumelhart and McClelland (1986*b*), and reproduced here as Table 7.1.) All patients thus far reported do make errors of some other types, such as occasional errors on regular words (e.g. HORSE—/hWs/, BASE—/pAs/) and non-regularization errors on exception words (e.g. FLOOD—/flOd/, LOSE—/lUs/) (all examples from Shallice, Warrington, and McCarthy 1983). The majority of errors, however, are strict regularizations.
5. Finally, as already mentioned, the patients' reading speed can be roughly normal.

In the following sections, we report experiments in which we observed the effects of different types of damage to the simulation model on performance with different types of words and non-words. We shall only be considering the oral reading performance (not lexical decision) of surface dyslexic patients; accordingly, when we talk about output from the model, this will always refer to error scores calculated over the phonological output units. The primary goal of the experiments was exploratory: how would the model perform when different components were damaged? A second goal was to determine whether damage to the model would produce the types of errors characteristic of

Table 7.1 *The phonemic categorization system used in the model, plus a pronunciation key*

		Place					
		Front		Middle		Back	
		V/L	U/S	V/L	U/S	V/L	U/S
Interrupted	Stop	b	p	d	t	g	k
	Nasal	m	—	n	—	N	—
Continuous consonant	Fric.	v/D	f/T	z	s	Z/j	S/C
	Liq/SV	w/l	—	r	—	y	h
Vowel	High	E	i	O	^	U	u
	Low	A	e	I	a/α	W	*/o

Key: N = ng in *sing*; D = th in *the*; T = th in *with*; Z = z in *azure*; S = sh in *ship*; C = ch in *chip*; E = ee in *beet*; i = i in *bit*; O = oa in *boat*; ^ = u in *but* or schwa; U = oo in *boot*; u = oo in *book*; A = ai in *bait*; e = e in *bet*; I = i_e in *bite*; a = a in *bat*; α = a in *father*; W = ow in *cow*; * = aw in *saw*; o = o in *hot*.

Reproduced from Rumelhart and McClelland (1986*b* Table 5, p. 235).

surface dyslexic patients. Ultimately, we would like to achieve simulations of specific cases of surface dyslexia, capturing both their qualitative and quantitative aspects, but we have not done so as yet. In the final section of the paper we discuss some of the issues that need to be addressed if additional research is to achieve this ultimate goal.

One final introductory comment: the model was developed on the basis of, and has been extensively evaluated relative to, data from normal readers. By comparison, these explorations of the neuropsychological implications of the model are at an embryonic stage. At many points in what follows, we shall have to say (or, rather, to save the reader from boredom, we shall hope that it is generally understood) that much more work on this approach is needed before a comprehensive story can be told. Our justification for offering this somewhat premature account is that, as we have already suggested, neuropsychological dissociations could be considered a major challenge to this kind of model which eschews separate routines, rule-based systems and other notions that have played a central role in cognitive neuropsychology. Even if premature, then, it seems useful to indicate some of the ways in which the model might respond to this challenge.

Overview of methods

The general procedure involved in these explorations was as follows. All lesion studies were done using the weights created after 250 epochs of training, when

the model had reached the nearly asymptotic level of performance illustrated earlier. All experiments were concerned with effects of damage ('lesions') to the system; in a later section we discuss other types of pathology that might be simulated. Lesions were made at the three different locations within the model where changes take place as a result of learning: weights on the connections from orthographic input units to hidden units ('early' weights); the hidden units themselves; and weights on the connections from hidden units to phonological output units ('late' weights). Damage was inflicted by zeroing a proportion of the connections or units at the lesion site. In the first, parametric, study to be reported, the proportions of damaged connections or units tested were 0.1, 0.2, 0.4 and 0.6. Damage was introduced probabilistically; with a damage value of 0.4, for example, the output from a random 40 per cent of the specified connections or units was zeroed. For any given lesion experiment, then, we shall be looking at performance for a particular combination of location and level of damage.

Although all representations and processes within the model are distributed (such that the model never, for example, assigns a single hidden unit to a particular word), it is by no means the case that all units are activated. In fact, in the processing of any given word, the majority of hidden units are not activated by the input; on average, about 24 of the 200 hidden units will be activated for any word. The result of this, when combined with probabilistic damage, is high variability from one lesion test to the next. In order to produce a larger pool of data yielding a more stable and less idiosyncratic picture of the model's behaviour when damaged, all lesion experiments for any stimulus set at any particular location and level of damage consisted of ten tests.

The data from a lesion experiment will consist of two measures. The first is the phonological error scores (means and standard deviations) for different pronunciations of the stimulus words being tested. Typically only two pronunciations of each word were tested; for the exception word PINT, for example, the pronunciations of primary interest were the correct one /pInt/ and the regularization /pint/. These mean phonological error scores will represent averages both over words within the set being tested (N to be specified for each experiment) and over tests ($N = 10$). The second measure concerns the relative error scores for the two pronunciations of a given word on a given test. If the model is performing correctly, then it should of course 'prefer' the pronunciation /pInt/ to the pronunciation /pint/. If, when damaged, it yields a lower score for the alternative pronunciation, we call this a reversal. We counted as reversals any cases where the alternative score was at least one full point lower than the score for the correct pronunciation. This measure will be given as reversal rate, meaning the percentage of occasions in a particular lesion experiment where an alternative pronunciation was preferred.

Arguments have been offered elsewhere (earlier in this chapter and in Seidenberg and McClelland 1988) for the adequacy of the phonological error score as a measure of naming performance by normal readers; and the impressive similarity between functions derived from the model and those from real subjects (as illustrated above) supports this rationale. There are, however, two aspects of this error score which make it less than ideal for lesion studies. The first is a substantive issue: the error score combines accuracy and latency in a single measure. While this may offer a reasonable characterization of normal readers, operating at relatively full efficiency with a minimum of errors and a maximum of speed, it is not necessarily so satisfactory for pathological data. If, for example, one patient reads slowly and makes many errors while another reads quickly with (approximately) the same error rate, then no single speed–accuracy trade-off function will characterize both. In the model, a high error score could represent a fast, wrong reading, a slow, correct reading, or a slow, wrong reading; when trying to relate the model's performance to patients, it would be informative if we could discriminate among those alternative interpretations. With the present measure, we cannot.

The other problem is a purely practical, procedural one: the error score represents the degree to which the pattern of activation over the phonological output units differs from the ideal pattern for the specified phonological code. In general, for simulations of normal data (but note that it may not always be safe to assume that the model in its normal, undamaged state 'knows' the correct pronunciation for all words; in fact, it does not), it will be adequate to specify only the correct pronunciation. But as soon as we wish to simulate error-prone patients, then in order to discover what sorts of pronunciation errors the model may make when it has been lesioned, we are required to specify every pronunciation of interest in order to identify the pronunciation yielding the lowest error score. This nuisance is admittedly of concern to the authors rather than to the readers of this chapter. We mention it here only to explain why, for many of the lesion studies to be reported below, we test two pronunciations of each word rather than many.

Experiment 1: the effect of different locations and levels of damage

To provide a basic introduction to the way in which the model's performance degrades under conditions of damage, we begin with a parametric exploration of the three locations and four levels of damage mentioned above. The stimulus items for this experiment, all four-letter words from the model's vocabulary, were the 16 regular words and 16 exception words listed in Table 7.2. The two sets were approximately balanced for both Kucera and Francis frequency and orthographic error scores; thus we can be reasonably

Table 7.2 *The 16 regular and 16 exception words used in Experiment 1, with their correct and alternative pronunciations, mean frequencies* and mean orthographic and phonological error scores from performance of the undamaged model after 250 epochs of training*

Regular	COR	OTH	Exception	COR	REG
COVE	/kOv/	/kUv/	MOVE	/mUv/	/mOv/
TINT	/tint/	/tInt/	PINT	/pInt/	/pint/
BEAD	/bEd/	/bed/	DEAD	/ded/	/dEd/
FOUL	/fWl/	/fOl/	SOUL	/SOl/	/sWl/
HOWL	/hWl/	/hOl/	BOWL	/bOl/	/bWl/
LEAF	/lEf/	/lef/	DEAF	/def/	/dEf/
HOOP	/hUp/	/hup/	HOOD	/hud/	/hUd/
LAKE	/lAk/	/lak/	LOSE	/lUz/	/lOz/
FILE	/fIl/	/fil/	FOOT	/fut/	/fUt/
PASS	/pas/	/pos/	POST	/pOst/	/p*st/
DAMP	/damp/	/domp/	COUP	/kU/	/kWp/
PINE	/pIn/	/pin/	POUR	/pOr/	/pWr/
BEND	/bend/	/bEnd/	PEAR	/pAr/	/pEr/
SKIN	/skin/	/skIn/	TOMB	/tUm/	/tom/
MEET	/mEt/	/mAt/	MONK	/m∧nk/	/monk/
DEAL	/dEl/	/del/	AUNT	/ant/	/*nt/
40.1 X̄ frequency			45.6		
5.8 X̄ orthographic error score			6.1		
4.1 X̄ phonological error score			4.8		

*From Kucera and Francis 1967.

confident that any differences in behaviour between the two sets should be genuinely attributable to regularity of spelling-to-sound correspondences. Also shown in Table 7.2 are the two pronunciations tested for each word. For the exception words, the alternative (to the correct, or COR) pronunciation was of course the regularization (REG). For regular words, it is not always obvious what the alternative should be, but we attempted to make these other (OTH) pronunciations as plausible as possible, for example choosing a pronunciation of the vowel or vowel combination which occurs in other words.

The mean phonological error scores for the two pronunciations of the words in the two sets are shown in Fig. 7.5 (regular) and Fig. 7.6 (exception). The abscissa in each graph represents level of damage, from none (performance of the model in its normal state) to proportion of damage = 0.6. For both regular and exception words, COR phonological error scores rise in a monotonic, indeed essentially linear, fashion with increasing level of damage. The effect of level of damage is much more striking than that of

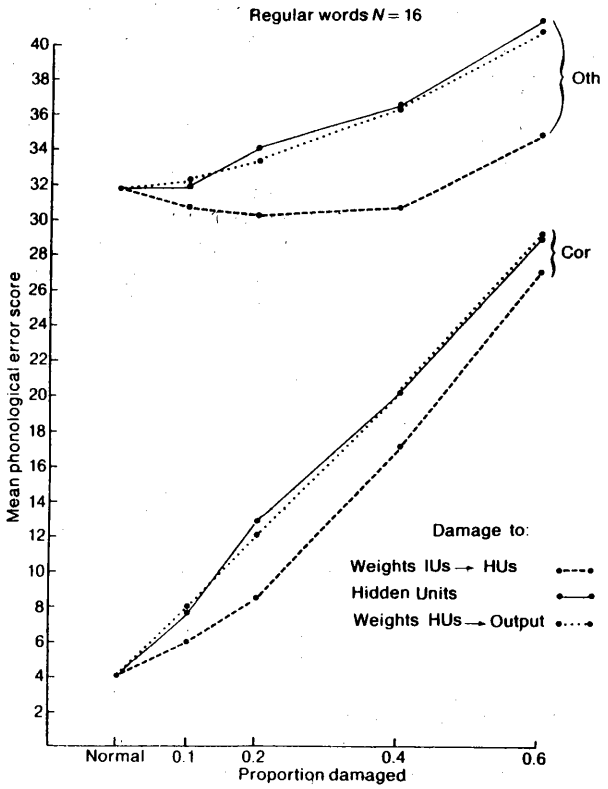

Fig. 7.5 The model's mean phonological error scores for correct and other pronunciations of regular words under normal conditions and with various locations and levels of damage.

location of damage: in fact, considering only the COR means, location appears to be relatively inconsequential. Damage to early weights consistently yields slightly lower scores than damage to either hidden units or late weights; but means for these latter two locations are virtually indistinguishable.

Three further aspects of the results in Figs 7.5 and 7.6 need to be highlighted:

1. Error scores for REG and OTH pronunciations also rise as a function of damage, but very much less dramatically than those for COR. This is probably not a ceiling effect because it is possible to obtain considerably higher error scores (on other types of words or with additional damage). As a result of this difference in rate of increase, the error scores for the correct and for the alternative pronunciations begin to converge at higher levels of damage, especially for exception words.
2. With regard to the two classes of words, COR scores for the regular words are consistently, though only marginally, lower than COR scores for the exception words; OTH scores for the regular words, on the other hand, are quite a bit higher than REG scores for the exception words. (This can be seen more easily in Table 7.3 where mean values are listed.) The net result is that there is a smaller difference between COR and REG scores for exception words than between COR and OTH scores for regular words.
3. Where location of damage has a notable effect is not on the means but on the standard deviations of the COR scores. Of particular interest, because

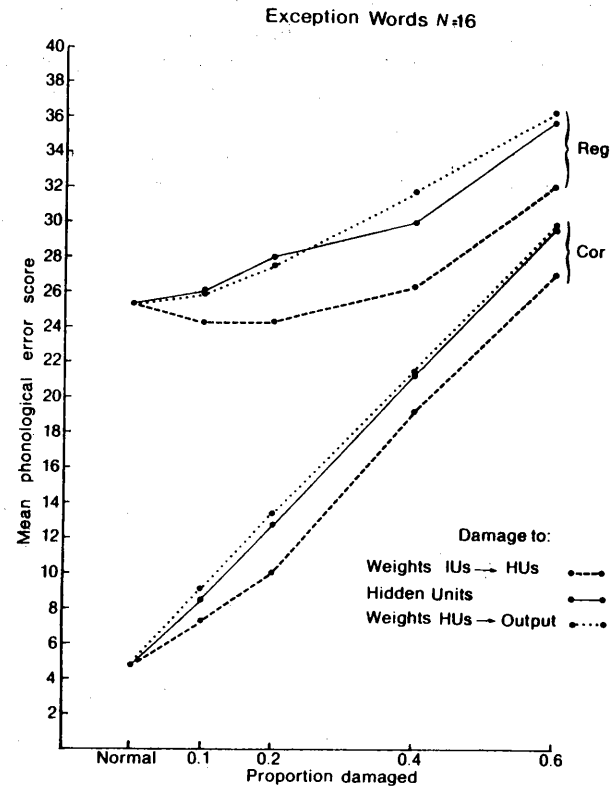

Fig. 7.6 The model's mean phonological error scores for correct and regularized pronunciations of exception words under normal conditions and with various locations and levels of damage.

the corresponding means are so similar, is the fact that damage to hidden units is associated with high variability in COR scores, while damage to connections between hidden units and output units produces very much lower standard deviations. These standard deviations are shown beside their corresponding means in Table 7.3.

Table 7.3 *Results from Experiment 1 for regular and exception words with different proportions of damage to hidden units or to connections from hidden units to output units: means and standard deviations of phonological error scores for the two different pronunciations, and reversal rates—the proportion of occasions on which the REG or OTH pronunciation yielded a lower error score than the COR pronunciation*

	Proportion damaged			
	0.1	0.2	0.4	0.6
Damage to hidden units				
Regular words				
mean (s.d.) error COR	7.6 (3.8)	12.8 (5.9)	20.2 (7.9)	28.9 (8.6)
mean (s.d.) error OTH	31.9 (11.1)	34.0 (10.7)	36.4 (11.3)	41.2 (10.8)
reversal rate	0	0	2.5%	7.5%
Exception words				
mean (s.d.) error COR	8.4 (4.0)	12.7 (5.3)	21.3 (7.6)	29.7 (8.1)
mean (s.d.) error REG	26.0 (8.8)	28.0 (8.7)	30.0 (9.1)	35.8 (10.2)
reversal rate	1.3%	2.5%	16.3%	23.8%
Damage to connections from hidden units to output units				
Regular words				
mean (s.d.) error COR	7.9 (2.5)	12.0 (3.1)	20.2 (3.8)	29.0 (3.8)
mean (s.d.) error OTH	32.2 (10.7)	33.4 (9.8)	36.3 (8.3)	40.6 (6.7)
reversal rate	0	0	0	0
Exception words				
mean (s.d.) error COR	9.2 (2.4)	13.4 (2.7)	21.5 (3.2)	29.9 (3.6)
mean (s.d.) error REG	26.0 (7.9)	27.5 (7.4)	31.8 (7.2)	36.4 (6.7)
reversal rate	0	0	1.3%	8.3%

The importance of these three points is revealed when we turn to the other measure of interest. Table 7.3 shows reversal rates under conditions of damage to hidden units and to late weights, the two locations which yielded virtually identical mean error scores. The first point (convergence between scores for

correct and for alternative pronunciations with increasing level of damage) means that while small amounts of damage yield few if any occasions on which the alternative pronunciation produces a lower score, higher levels of damage produce a number of preferences for the alternative pronunciation. The second point (less distance between COR and REG error scores for exception words than between COR and OTH error scores for regular words) means that any combination of location and level of damage which produces reversals at all produces higher reversal rates for exception than for regular words. The third point (higher variability around the mean for damage to hidden units than for late weights) has the outcome of substantial reversal rates when hidden units are disrupted, but low reversal rates with zeroing of late weights. When the effects of these three points are considered together, the complete picture is a notable number of reversals only for exception words and only given higher levels of damage to hidden units.

The interpretations of both the difference between regular and exception words and the effect of increasing level of damage are reasonably obvious. Even when 60 per cent of the normally encoded information at some level is unavailable, the model generally prefers the correct pronunciation of regular words because the correspondences embodied in regular words are in essence overlearned in the model. Exception words are more vulnerable to damage, but they are sufficiently well learned via distributed representations that their COR pronunciations are typically preferred so long as damage level is low. When about half of the hidden units are inactivated, however, pronunciations reflecting overlearned REG correspondences begin to be more attractive for some exception words on some tests.

The interpretation of the location effect is less obvious. Recall that it is not the case that lesioning late weights results in better performance (i.e. lower mean error scores) than lesioning hidden units: it simply results in more consistent, less variable performance. It is clear why this should have the effect that it has on reversal rates: given that the COR means are always lower than the REG means, it is only when the COR scores vary considerably around the mean that some of them will turn out to be higher than the corresponding REG scores. But why is such variability associated only with damage to hidden units and not with damage to late weights? We suggest that while all representations are distributed, some representations are more distributed than others. There are only 200 hidden units, and on average only about 24 of these are activated for any given word. With 60 per cent probabilistic damage, one infliction of damage could by chance knock out a number of the relevant 24 units, while the next lesion might happen to hit none of the crucial units: thus, high variability from one test to the next. By contrast, the number of connections from hidden units to output units germane to any given word is much larger. Each of the 20-odd hidden units relevant to a word is connected to all 460 phonological units; if the model were performing without error, 16 of

these phonological units would be activated for each phoneme in the correct pronunciation. Zeroing the weights on 60 per cent of these connections will indeed have a deleterious effect on performance; but because the 'knowledge' at this level is distributed over such a large number of connections, any random 60 per cent loss will produce deleterious effects similar to any other.

There are marked differences in reversal rates for individual exception words within the set, but we defer discussion of this until Experiment 4, where we consider the variables that make words prone or resistant to regularization.

In summary of Experiment 1, increasing levels of damage at all three locations produces steady increments in the phonological error scores associated with correct pronunciations of known regular and exception words. Higher levels of damage at the location of the hidden units yield a significant number of tests on which the model 'prefers' the regularized pronunciation of an exception word to the correct pronunciation, just as surface dyslexic patients often do in oral reading. Accordingly, most of the remaining lesion experiments will concentrate on this location and level of damage.

Experiment 2: lesions and novelty

In earlier sections of this chapter we discussed how the model in its normal state performs both on words in its vocabulary and on novel stimuli (non-words). Experiment 1 demonstrated how the model in various damaged states performs on regular and exception strings, but only ones from its premorbid vocabulary. The purpose of Experiment 2 was to examine how the model deals with novelty once it has been damaged. As explained in the introduction to lesioning the model, at least some surface dyslexic patients (e.g., Bub, Cancelliere, and Kertesz 1985; McCarthy and Warrington 1986) show normal accuracy in non-word reading, though other reported cases of surface dyslexia are either at the bottom end of the range of normal performance (e.g. Kay and Lesser 1985) or frankly impaired at non-word naming (e.g. Masterson 1985).

For our initial exploration of novelty, the stimulus items were triplets consisting of an exception word, a regular word, and a non-word, matched for orthographic 'body' or rime: for example, COME, HOME, and NOME. This design enabled us to test the model's performance using the identical two pronunciations of each body within a triplet: the stimulus items ($N=20$ triplets) with their alternative pronunciations are shown in Table 7.4. The regular pronunciation of a body will of course be considered the correct pronunciation for both the regular word and the non-word members of the triplet but the incorrect pronunciation of the exception word; correspondingly, the irregular pronunciation will be correct for the exception word but

Table 7.4 *Stimulus items for Experiment 2: triplets of exception words, regular words and non-words matched for body, with the two alternative pronunciations tested for each triplet*

Exception word	Regular word	Non-word	Pronunciation of body: Regular	Pronunciation of body: Exception
PUT	CUT	DUT	/ʌt/	/ut/
PINT	HINT	RINT	/int/	/Int/
GROSS	CROSS	BROSS	/*s/	/Os/
CASTE	HASTE	NASTE	/Ast/	/ast/
TOUCH	COUCH	BOUCH	/WC/	/ʌC/
BOWL	HOWL	POWL	/Wl/	/Ol/
COME	HOME	NOME	/Om/	/ʌm/
STEAK	BLEAK	SHEAK	/Ek/	/Ak/
GIVE	FIVE	MIVE	/Iv/	/iv/
DEAF	LEAF	NEAF	/Ef/	/ef/
SOUL	FOUL	DOUL	/Wl/	/Ol/
PEAR	GEAR	MEAR	/Er/	/Ar/
GLOVE	GROVE	BLOVE	/Ov/	/ʌv/
BULL	DULL	TULL	/ʌl/	/ul/
SWEAT	TREAT	SNEAT	/Et/	/et/
LOSE	NOSE	BOSE	/Oz/	/Uz/
BLOWN	CROWN	TROWN	/Wn/	/On/
FLOOD	BROOD	FROOD	/Ud/	/ʌd/
POST	LOST	FOST	/*st/	/Ost/
HAVE	GAVE	BAVE	/Av/	/av/

incorrect for both the regular word and non-word. Reversals will then be instances of phonological error scores REG < IRR for exception words and IRR < REG for regular words and non-words. In order to make reversals a meaningful concept for the non-words, it is obviously necessary to ensure that in its normal, undamaged state, the model prefers the REG to the IRR pronunciation of all of the non-words. The original set of triplets ($N=30$) turned out to contain 10 items where this was not the case. These have been eliminated, yielding 20 triplets.

Table 7.5 shows the mean phonological error scores (and standard deviations) for the two pronunciations of the 20 items in each word set, both for undamaged performance and with damage to hidden units, $p=0.6$ (where N/cell = 10 runs × 20 items). With the model in healthy condition, performance differs as a function of word class in two ways. First, mean phonological error scores for the COR pronunciations (which are, remember, REG for regular words and non-words but IRR for exception words) have the ordering of regular words < exception words ≪ non-words; this of course reflects what

Table 7.5 *Results for Experiment 2. Columns correspond to: (1) the mean phonological error scores for the correct pronunciation of the stimulus items; (2) the standard deviations associated with Column 1 means; (3) the mean phonological error scores for the other pronunciation (regular for the exception words, irregular for the regular words and non-words); (4) s.d.'s for Column 3 means; (5) the difference between the means in Columns 3 and 1; (6) the percentage of tests (N=200) on which a particular item had a lower error score for OTH than for COR*

	(1) $\bar{X}$ COR	(2) (s.d.)	(3) $\bar{X}$ OTH	(4) (s.d.)	(5) OTH–COR	(6) Reversals (%)
Undamaged						
regular	4.4	(1.8)	28.9	(9.8)	24.5	—
exception	5.1	(2.5)	26.9	(11.5)	21.8	—
non-word	11.4	(4.2)	24.6	(9.5)	13.2	—
Damage HU's $p=0.6$						
regular	27.1	(8.0)	36.8	(9.2)	9.7	9
exception	28.1	(7.7)	33.7	(9.6)	5.6	24
non-word	29.2	(8.1)	35.2	(9.4)	6.0	21

the model knows about these three types of letter string. Second, mean error scores for the OTH pronunciations have the reverse ordering, non-words < exception words < regular words. Once again, this is to be expected, reflecting as it were the model's confidence in its preferred pronunciation. The net result, also shown in Table 7.5, is that the difference between OTH and COR pronunciations is biggest for regular words and smallest for non-words.

Turning to damaged performance, we see that the ordering of error scores for correct pronunciations of the three word classes is maintained, but only just: the discrepancies among them are now small. In particular, the major advantage (in undamaged performance) for familiar lexical items over unfamiliar strings is all but lost. The difference between OTH and COR means is much reduced in all three conditions; most interestingly, the exception words, which yielded a difference score not very dissimilar to regular words under normal conditions, show a difference score virtually identical to the non-words after lesioning.

Table 7.5 also shows reversal rates (percentage of tests on which OTH < COR) for the three string types. Experiment 1 taught us the importance of variability to reversal rates, and its role can be seen again here, not in the standard deviations *per se* (which are quite constant across word class under damaged conditions) but in terms of the standard deviations

relative to the OTH–COR difference score. For regular words, where this difference score is larger than 1 s.d., the reversal rates are low; for exception words and non-words, where the difference score is less than 1 s.d., the reversal rates are substantially higher.

The reversal rates for the regular and exception words merely replicate (albeit with different stimulus items) those reported for Experiment 1 (Table 7.3). The interesting finding from Experiment 2 is the notable number of reversals for non-words. This aspect of the model's performance seems to constitute a good match for some surface dyslexic patients but not others. The patients studied by Bub, Cancelliere, and Kertesz (1985) and McCarthy and Warrington (1986) were both asked to read aloud Glushko's (1979) list of 43 'exception' pseudowords, which are very similar to the non-words used here. These two patients showed normal accuracy of non-word reading (indistinguishable, in fact, from Glushko's university-student subjects), with few irregular pronunciations (e.g BLEAD—/bled/ rather than /blEd/): 4/43 (9 per cent) and 3/43 (7 per cent), respectively. On the other hand, a surface dyslexic patient studied by Kay and Lesser (1985) made some outright errors in his non-word reading, and his acceptable responses included a somewhat larger proportion (19 per cent) of irregular pronunciations. A question for future exploration is whether other features of the model's lesioned performance have a greater resemblance to Kay and Lesser's patient than to the Bub, Cancelliere, and Kertesz and McCarthy and Warrington patients.

The relatively high and approximately equal reversal rates for exception words and non-words are intriguing in their suggestion that damage can destabilize the model's performance on two different types of items in roughly the same way (at least as assessed by this somewhat gross measure, simple preference for an alternative plausible pronunciation). The exception words are familiar orthographic sequences to the model, but embody letter–sound correspondences that are atypical. The non-words offer correspondences which (at least most often, over the range of known words) are typical; but since the non-words are not familiar orthographic sequences, the model is much less confident about an appropriate pronunciation for them. Only for items that are both familiar and regular does the damaged model retain a reliable preference for the 'correct' pronunciation.

Experiment 3: frequency effects

The most notable feature of surface dyslexic oral reading, the tendency to regularize words with an irregular spelling-to-sound correspondence, is strongly modulated by frequency for at least some reported patients. M.P., the case studied by Bub, Cancelliere, and Kertesz (1985), made few regularization errors or, indeed, errors of any kind in oral reading of high-frequency exception words. As word frequency declined, her error rate increased

steadily, and virtually all her errors were regularizations. Such dramatic frequency effects do not, however, characterize all surface dyslexic patients. For H.T.R. (Shallice, Warrington, and McCarthy 1983), regularity seems to have been such a powerful determinant of reading success that with highly irregular words (e.g. SUEDE, UNIQUE or BUSINESS) she mispronounced the majority of words whatever their frequency. On the other hand, for mildly irregular words (more like the 'exception' words that we have been testing here, e.g. DREAD, CROW), H.T.R.'s success showed some sensitivity to frequency. Experiment 3 was an attempt to determine whether the model's tendency to produce regularization errors (reversals) is modulated by word frequency.

This evaluation was made using three sets of exception words; their frequency characteristics are listed in Table 7.6. In the first set, rather than

Table 7.6 *Description of the word sets used to assess frequency effects after damage*

Word set	(*N*)	$\bar{X}$ Frequency	Frequency range
(1) Very high frequency words	(20)	1859.8	424–5146
(2) Glushko words			
low	(10)	14.6	2–36
low–medium	(8)	69.3	51–88
medium–high	(8)	218.1	108–424
high–very high	(9)	1643.9	630–3941
(3) Body-matched pairs			
lower-frequency member	(11)	167.6	5– 938
higher-frequency member	(11)	1046.8	230–3292

comparing different levels of frequency, we simply selected the 20 virtually highest-frequency exception words in the model's vocabulary, words like ARE, HAVE, ONE, WERE, SAID, WHAT. The question is whether such exalted frequency values 'protect' words from reversing. The second set consisted of 35 items from Glushko's (1979) list of exception words; these are 4–5 letter words with reasonably common spelling patterns (i.e. no orthographically weird words are included), all of which, of course, are in the model's vocabulary. For purposes of evaluating frequency effects, the 35 items were divided into four frequency bands, as shown in Table 7.6. Finally, we selected 22 items consisting of 11 'body'-matched pairs with one higher- and one lower-frequency member in each pair; examples of these items with their K–F frequencies in parentheses are GOOD (807)–HOOD (7) and FOUR (359)–POUR (9). As can be seen in Table 7.6, there was considerable overlap in frequency between the two sets as a whole; none the less, within each matched pair, there was always a substantial discrepancy in frequency. The

highest frequency item is the lower set, WHERE (938), was matched with THERE (2724).

Table 7.7 presents mean error scores (and s.d.'s) for the COR and REG pronunciations for each list, plus reversal rates. It appears that frequency is not the major determinant of susceptibility to reversal in the model. It has

Table 7.7 *Mean (s.d.) phonological error scores for COR and REG pronunciations of the various frequency lists with damage to hidden units p=0.6, plus reversal rates*

Word set	$\bar{X}$ COR	(s.d.)	$\bar{X}$ REG	(s.d.)	Reversals (%)
(1) Very high frequency words	29.5	(8.4)	34.7	(9.0)	29
(2) Glushko words					
low	32.0	(8.1)	35.0	(8.1)	37
low–medium	30.2	(7.4)	36.3	(10.7)	28
medium–high	29.6	(8.6)	34.8	(9.9)	31
high–very high	28.4	(8.2)	35.3	(10.0)	26
(3) Body-matched pairs					
lower–frequency member	29.8	(8.0)	30.7	(8.0)	46
higher–frequency member	28.8	(8.1)	31.6	(7.9)	35

some influence: in the set of body-matched words (list 3), the lower-frequency members reversed more often than the higher-frequency members, and within the four frequency bands of the Glushko words (set 2), the low-frequency items showed the highest reversal rate. On the other hand, considering all word sets in Table 7.5, there are several comparisons where lists with large-frequency differences yield essentially identical reversal rates, for example the very high frequency words and the Glushko low–medium words, or the higher-frequency items of set 3 and the Glushko low words. It is clear (1) that being very common does not protect a word from reversing when the model is damaged; and (2) that we shall have to look to some variable(s) other than frequency if we want to discover the basis for susceptibility to reversal. That is precisely what we shall do next, in Experiment 4.

Experiment 4: preferred pronunciations, phonemic features, and regularizations

As indicated in the introduction to lesioning the model, the current form of output from the model (phonological error scores) requires any pronunciation

of interest for a given letter string to be explicitly tested. Under damaged conditions, the model may, as we have already seen, prefer the regularized pronunciation of an exception word; but this only informs us that the regularization is *a* preferred pronunciation, not that it is *the* preferred pronunciation; another pronunciation could, under the same lesioned conditions, yield a still lower error score. Regularizations as alternative pronunciations derive from some pre-suppositions about the principles underlying translation from orthography to phonology. For a more theoretically neutral exploration of the question of preferred pronunciations, we tested a small set of exception words by varying the vowel segments of each word to include every possible vowel pronunciation within the model's phonemic coding scheme. Vowels seemed a sensible choice since they tend to have more variable letter–sound correspondences than do consonants.

The model's phonemic coding scheme, taken from Rumelhart and McClelland's (1986) past-tense verb learning model and illustrated earlier in Table 7.1, codes vowels in terms of three dimensions: place (front, middle, back), length (long, short), and height (high, low). Thus, for any given word, there are $3 \times 2 \times 2 = 12$ possible vowel pronunciations, which can be illustrated with respect to the test word PINT. The correct pronunciation is of course /pInt/, where the vowel is middle, long and low. There are then four pronunciations which differ from the correct one by a single vowel feature: /pAnt/ and /pWnt/ move the place from middle to front and from middle to back, respectively, without changing length or height; /pant/ changes the length without affecting place or height; and /pOnt/ changes height only. Five pronunciations involve changes in two of the three features: for example, /pent/ involves a change in both place and length, /p^nt/ in both height and length, and so on; and two vowel pronunciations involve a change in all three dimensions.

The exception words used for this evaluation were the first 10 words from the set of 16 listed in Table 7.2. As in other experiments, we did an initial test with the model in its normal, undamaged condition and then 10 runs with damage to hidden units, $p = 0.6$. Instead of the usual two error scores to be compared, each test in this experiment yields 12 error scores for each word, corresponding to the 12 possible pronunciations of the vowel. For this experiment, then, we must distinguish between reversal rate (proportion of occasions on which the single alternative corresponding to the regularized pronunciation yielded the lowest error score). Of course it is possible for more than one alternative (indeed, in principle, for all 11 alternatives) to produce error scores lower than the correct pronunciation; but for simplicity's sake, and because we are interested in actual preferences, we shall only discuss data concerning the lowest score for each word on each test.

Table 7.8 displays the phonological error scores (means and s.d.'s), under both normal and damaged conditions, for COR pronunciations and for pronunciations differing from COR by ONE, TWO, or THREE vowel features. It is clear that the model is highly sensitive to phonemic distance, as

Table 7.8 *Phonological error scores (means and standard deviations) for a set of 10 exception words tested against all possible pronunciations of the vowel segment*

	Normal			Damaged			
	(*N*)	$\bar{X}$	(s.d.)	(*N*)	$\bar{X}$	(s.d.)	Lowest score (%)
COR	(10)	4.3	(1.3)	(100)	30.8	(7.3)	44
ONE	(40)	18.5	(2.5)	(400)	36.0	(8.4)	37
TWO	(50)	32.6	(3.3)	(500)	40.6	(8.8)	17
THREE	(20)	46.4	(4.0)	(200)	44.8	(9.0)	2

Each word has one correct pronunciation, four pronunciations differing from correct by ONE phonemic feature in the model's coding scheme for vowels, five pronunciations differing by TWO features, and two pronunciations differing by all THREE features. The table shows the model's normal performance and also with damage to hidden units, $p = 0.6$. The last column indicates the proportion of damaged tests for which the lowest phonological error score corresponded to the correct pronunciation or to pronunciations differing by ONE, TWO or THREE features.

measured by number of features differing between correct and alternative pronunciations. Especially when undamaged, but also when lesioned, the model's error scores are monotonically related to the number of features altered.

Such differences in error scores, and their associated standard deviations, translate themselves into reversal rates in the way that we have come to expect. As shown in the final colume of Table 7.8, with lesioning, the COR pronunciation yielded the lowest score on only 44/100 tests; thus overall reversal rate was 56 per cent. Of these 56/100 tests resulting in a reversal, the preferred pronunciation was substantially more likely to be a ONE-feature change than a TWO-feature change and was very unlikely indeed to be a pronunciation differing from COR by THREE features.

This result has important consequences for the interpretation of our lesioning results. First of all, it essentially solves a puzzle concerning regularization rates for specific exception words. As mentioned at the end of Experiment 1, the probability that the damaged model will prefer a regularized pronunciation of an exception word varies substantially across different exception words. Because monosyllabic exception words in English by no means constitute an unlimited pool, the same words tend to turn up repeatedly; for example, each of the 14 words in Table 7.9 happens to have

Table 7.9 *For a set of 14 exception words: the number of times each has been tested with damage to hidden units, $p=0.6$; the word's overall regularization rate; and the number of vowel features (in the model's phonemic coding scheme) by which the regularized pronunciation differs from the correct pronunciation.*

Word	(*N* tests)	Regularization rate (%)	No. of features differing between COR and REG
HOOD	(70)	57	ONE
BULL	(50)	56	ONE
COME	(70)	54	ONE
GLOVE	(50)	52	ONE
SOME	(50)	50	ONE
GOOD	(50)	46	ONE
FOOT	(50)	40	ONE
DEAF	(60)	23	TWO
SOUL	(60)	18	TWO
HEAD	(60)	17	TWO
FLOOD	(40)	15	TWO
POST	(60)	7	THREE
PINT	(60)	5	THREE
BOTH	(40)	0	THREE

been examined under conditions of damage to hidden units, $p=0.6$, in no less than four and, for some words, in as many as seven different tests. For these particular words, then, there are very stable estimates of their tendency to regularization. We spent a considerable amount of time and effort attempting to determine what factor(s) might account for the marked variation in regularization rate listed in Table 7.9. Our third experiment demonstrated that frequency was not the crucial variable, and a number of other explorations (such as orthographic neighbourhood: what proportion of words with that 'body' have a regular or an irregular spelling-to-sound correspondence) similarly failed to explain these dramatic differences in reversal rate. As the final column of Table 7.9 indicates, the determining factor is almost certainly the number of vowel features, in the model's phonemic coding scheme, by which the regularized pronunciation differs from the correct pronunciation.

It might be worth adding the reassuring note that this discovery of the major factor contributing to reversals in no way compromises our crucial finding of higher reversal rates for exception than for regular words. Recall that in

Experiment 2, regular and exception words were matched for body and tested with the same two pronunciations. This means that the distance in phonemic vowel features between the correct and the alternative pronunciation was identical for the regular and exception words in Experiment 2. None the less, damage resulted in a substantial difference between the word classes in reversal rate.

The second important implication of this discovery concerns the model's success in simulating the reading performance of surface dyslexic patients. Regularization of the word PINT is often used in descriptions of such patients, partly because it is in fact a frequent error by real patients and partly because investigators of reading disorders thought that they understood why it should be such a frequent error. In the terminology of Henderson (1982) and Patterson and Morton (1985), PINT is a heretic word: all of the 12 other monosyllabic words ending in -INT are pronounced regularly, as in MINT. If a neurological injury could selectively disrupt some component of the system for retrieving pronunciations of whole familiar words, forcing a patient to rely on some other procedure involving grapheme–phoneme mapping rules or analogies with other known words, it seemed obvious that PINT should then be pronounced /pint/. As it happens, though, the vowel in /pint/ differs from /pInt/ not by ONE or by TWO but by THREE phonemic features. Therefore, although the patient data suggest that this ought to be a common regularization error, the model virtually never prefers /pint/ to /pInt/. Note that this is not to say that the damaged model always prefers the correct pronunciation for PINT. In fact, in Experiment 4 /pInt/ yielded the lowest score on only 5/10 tests. The preferred pronunciation in these reversals was, however, not the regularization /pint/, differing from /pInt/ by THREE features, but rather the pronunciation /pAnt/, differing from /pInt/ by only ONE feature.

The obvious next step was to return to the reading data from surface dyslexic patients to see whether their reading performance might be influenced by this variation of phonemic feature distance which so strongly constrains the model's preferred pronunciations. An error corpus from each of two patients, H.T.R. (Shallice, Warrington, and McCarthy 1983) and K.T. (McCarthy and Warrington 1986) was subjected to the following analysis. In order to make the data set as similar as possible to the results from the model, we included only monosyllabic words in which the patient's error was restricted to the vowel segment of the word. This produced an error set of $N=61$ for H.T.R. and $N=88$ for K.T. Each error was coded in terms of the distance (ONE, TWO or THREE features) between the correct pronunciation and the patient's reading response to the word. The reversal errors by the model in Experiment 4 ($N=56$) were coded in the same way. The results, scored as a percentage of responses corresponding to ONE, TWO or THREE features changed, are shown in Table 7.10.

Table 7.10 *(1) The proportions of the model's reversals and of the patients' errors (on the vowel segment of monosyllabic words only) that involve a change in ONE, TWO or THREE phonemic vowel features. (2) The percentage of the errors in (1) corresponding to exact regularizations of the exception target word. (3) As in (2), but restricted to words for which the regularization differs from the correct pronunciation by just ONE feature*

	Model	H.T.R.	K.T.
(*N*)	(56)	(61)	(88)
	(%)	(%)	(%)
(1) ONE	66.1	59.0	52.2
TWO	30.4	31.1	31.8
THREE	3.6	9.8	15.9
(2) % regularizations	19.6	78.8	81.8
(3) % regularizations for words where regularized pronunciation is a ONE-feature change	30.8	79.2	90.2

Much to our surprise, the patients' behaviour in this regard is very well simulated by the model. H.T.R. and, in particular, K.T. are a little more likely than the model to produce responses differing from the correct pronunciation by THREE features (for example, they both read PINT as /pint/!); but the similarities in these values are much more striking than the differences. In fact, this outcome goes beyond mere simulation. It is a prediction from the model to the data, and constitutes an analysis of the patient data that, we claim, no one would have thought of doing without the model's prediction. What this outcome means is that while PINT—/pint/ may be a frequent surface dyslexic reading error, it is not a typical one. Just as in the damaged model, typical reading errors by the patients (at least these two patients) involve a change in just a single phonemic feature.

Although the behaviour of the model and the patients concur closely in this regard, there is in fact one major difference between them. Remember that the 56 observations for the model in this analysis include not just regularizations but all reversals. Likewise, the 61 and 88 errors for H.T.R. and K.T., respectively, are the patient equivalent of reversals: they include not just regularizations but any reading error where the patient's pronunciation was (only) a misreading of the vowel. We can therefore now ask: for the model and for the patients, what proportion of these reversal errors correspond to the single alternative that happens to be the regularized pronunciation of that exception word? As shown in the line of Table 7.10 labelled '% regularizations', this proportion is high for the two patients but low for the model. Although the patients and the model make roughly the same proportions of

errors involving ONE- or TWO-feature changes, the patients appear to differentiate among the various options at each level, selectively favouring the pronunciation corresponding to the exact regularization. For the model, on the other hand, all alternatives within each level seem to be more or less equivalent: the regularization has no special status.

Since the word sets used in this comparison between H.T.R., K.T., and the model were not the same, they do not contain the same proportion of words for which the regularized pronunciation is ONE, TWO or THREE features different from the correct pronunciation. Given the dependence of errors on closeness of phonemic features, such unmatched lists will mean unequal 'opportunities' for regularization. As the last line of Table 7.10 demonstrates, however, the picture is only slightly altered if we restrict the comparison to the subset of words within each of these sets where the regularization involves a ONE-feature change. Maximizing the likelihood of regularization in this way increases the model's proportion of reversals that are regularizations from roughly 20 per cent to 30 per cent; but the comparable values for H.T.R. and K.T., respectively, are 80 per cent and 90 per cent.[2]

This difference in tendency to regularization may in fact be a reflection of a more general difference: the model is much more likely that the patients to produce errors which are 'implausible' realizations of the vowel, in the sense that no existing word in English embodies that pronunciation of the vowel grapheme. The patients do make such errors; for example, H.T.R. read SOUL as /sYl/ and BALD as /bOld/, and K.T. read ROOK as /rok/. There are no English words in which OU is pronounced /Y/, A is pronounced /O/ or OO is pronounced /o/. For want of any better description or account, such errors by surface dyslexic patients have typically been described as 'visual' or 'orthographic' (see, for example, Coltheart *et al.* 1983), and indeed these three examples from the two patients demonstrate why: /sYl/, /bOld/ and /rok/ actually correspond to the phonology of the real words SOIL, BOLD, and ROCK, each of which is orthographically similar to the target word engendering the error response. But, of course, one cannot be sure that these responses represent visual confusions by the patient: they could arise in the process of translation from orthography to phonology just as we assume the patients' regularization errors do.

The point germane to this discussion is that such errors with implausible grapheme–phoneme vowel correspondences are relatively rare in the patients' error corpora: in the subsets of errors being considered here, only $4/61 = 6.6$ per cent of H.T.R.'s errors and $3/88 = 3.4$ per cent of K.T.'s errors were of this type. By contrast, looking at the reversal errors by the model in Experiment 4, $34/56 = 60.7$ per cent of these have implausible correspondences. (Note: this is 34 tokens, i.e. actual instances of reversal, but only 20 types, i.e. different pronunciations.) For example, as already mentioned, the model's most common reversal error for PINT (preferred on 4/10 tests) was /pAnt/; in no

real English word is the single vowel I pronounced /A/. Since regularizations never represent implausible correspondences (on the contrary, they represent the most typical correspondence, which is why they are called regular), we suggest that the apparent difference between the model and the patients in regularization rate may be wholly or partly attributable to a difference in the likelihood that the vowel correspondence will be a legitimate one.

This characteristic of the model's lesioned performance clearly does not provide a good match for the behaviour of real patients. The next step in our investigations, but going beyond the scope of this chapter, will be to explore the basis of this difference between model and patient performance. Below we consider some directions that this investigation could take.

Summary of the lesioning experiments

Like other connectionist models of cognitive processing where the effects of damage have been investigated (e.g. Sejnowski and Rosenberg 1986; Hinton and Shallice, personal communication), the model described here performs in a reasonable manner when lesioned. Phonological error scores, the model's way of indicating its response to a stimulus item, increase monotonically with amount of damage. These augmented scores could be taken to reflect an increase in the proportion of incorrect naming responses, or an increase in the latency of responses, or both. Future work on the model will attempt to differentiate between these two aspects of any oral reading response. The precise location of damage (connections from input units to hidden units; hidden units *per se*; connections from hidden units to output units) has relatively little effect on the size of the error scores; but these locations have differential effects on the variability of error scores, and accordingly on the main measure of interest here: the likelihood that the model will 'prefer' a pronunciation other than the correct one for a given word. Damage to hidden units yields the maximum discrepancy in error rate between regular and exception words.

As Morton and Patterson (1980) insisted for another variety of acquired reading disorder, we must emphasize that there is no precise, fixed characterization of reading performance which qualifies as surface dyslexia. Certain striking features of a patient's overall pattern of reading skill prompt us to use the label 'surface dyslexia'; but each patient is unique. Attempts to simulate the abstract entity called surface dyslexia must be tempered by reminders that real patients are not abstractions. As already noted, some surface dyslexic patients show marked frequency modulation of their success in reading irregular words, while others do not; some patients have an essentially normal ability to read nonsense words, while others do not. Such specific features are only meaningful in relation to the particular patient's

precise processing profile. The same approach must be taken in evaluating the model's performance. It does not greatly matter (though it is of course interesting to know) whether the model shows significant frequency effects in exception word performance after lesioning. What matters is an account of why and under what circumstances one expects to find frequency effects, and whether the presence or absence of an effect fits with other things that we know about the model's or the patient's performance.

With this caveat in mind, plus a reminder of our initial warning about the early stage of these explorations of damage to the model, we suggest that this approach to the study of reading disorders—'lesioning' a working computational model—shows considerable promise. Moreover, several aspects of the initial damage experiments leave us optimistic that the particular model we have been using, or something very much like it, has considerable potential to provide a detailed account of acquired reading disorders. The first result from these experiments is the demonstration that the model can in fact produce the types of errors characteristic of surface dyslexic readers. This is important because the model lacks the non-lexical spelling–sound rules previously thought to be responsible for these errors. The second finding is that both patient and model errors are related to the distance between the correct pronunciation and the error in terms of number of phonemic features. We consider this finding to be important because it shows that the attempt to simulate impaired performance can deepen our understanding of the phenomena. In this case, the relevance of phonemic features to patient errors was not recognized until we attempted to simulate their performance.

A third result of the simulations is that they may offer a different interpretation for the 'visual' errors sometimes noted in surface dyslexic patients. The model produced errors such as PINT→/pAnt/, which do occur (though not commonly) in the error corpora of the surface dyslexic patients H.T.R. and K.T. discussed above, and which occur more frequently in other reported cases. It was thought that such an error '. . . could not arise simply through phonological reading' (Coltheart *et al.* 1983, p. 480) because the single vowel letter I is never pronounced /A/ in an English word. Moreover, if the naming response is treated as the real word PAINT, then its orthographic overlap with the stimulus word PINT is considerable. Therefore, such errors have been called 'visual' or 'orthographic' or, even more literally in the case of PINT→/pAnt/, a letter addition error (Coltheart *et al.* 1983). Our comment on this topic is merely speculative, especially as the model produces these errors with greater frequency than is observed in most patients; but the fact that the model yielded such errors with a completely intact orthographic encoding system suggests that 'visual' errors need not be 'visual' in origin.

We can summarize the relationship between the model's damaged performance and that of patients in the literature as follows. The performance of the patients who have been categorized as surface dyslexic varies in

systematic ways. As a broad generalization, following the description in Shallice and McCarthy (1985), the patients can be divided into two types. Type I patients, including H.T.R. (Shallice, Warrington, and McCarthy 1983), M.P. (Bub *et al.* 1985) and K.T. (McCarthy and Warrington 1986) exhibit the following characteristics:

1. Accuracy of regular word naming is at or near normal levels.
2. Naming latencies are within normal limits.
3. Accuracy in non-word naming is normal.
4. Most errors are regularizations of exception words.
5. Language comprehension and semantic knowledge are severely impaired.

Type II patients, a more heterogeneous lot than Type I, include J.C. and S.T. (Marshall and Newcombe 1973), P.T. (Kay and Lesser 1985) and E.S.T. (Kay and Patterson 1985). These cases exhibit the following characteristics:

1. Naming is poorer for exception than for regular words, but performance on regular words is also impaired.
2. Naming latencies are abnormally slow, and the patient may make a series of attempts to name a single word.
3. Non-word naming is impaired (where tested).
4. Regularization errors do not necessarily account for the majority of errors.
5. There is no marked impairment of semantic knowledge.

Shallice and McCarthy (1985) argue for a qualitative distinction between these two patterns, and they term the first pattern 'semantic dyslexia', reserving the label 'surface dyslexia' for Type II.

Although we began these explorations with the goal of simulating Type I cases, because both they and the model read without semantics, our damage experiments in fact yielded a profile more reminiscent of Type II patients. Our account of these patients cannot be considered complete, because it is likely that they do use partial semantic information derived from the orthographic input to assist the generation of a pronunciation. A more comprehensive account would explain this compensatory strategy and the extent to which it contributes to Type II performance.

The damaged performance of the model clearly does not provide a good fit to the Type I patients. However, it would be inappropriate to conclude that these patients' performance is inconsistent with the model or cannot be simulated by it. Although the types of damage that we have explored do not produce error scores in the normal range for regular words and non-words alongside impaired performance on exception words, this is not to say that such a pattern is an impossible one for the model. First, there are questions about the implementation of the model that need to be explored. Second, the model suggests several other potentially interesting bases for impaired

performance that have not been investigated as yet. It is worth considering briefly these directions for future research.

As Seidenberg and McClelland (1988*a*,*b*) note, several properties of the model seem to be theoretically important. These include the notion that orthographic and phonological representations are distributed, the intermediate level of hidden units, the way the learning rule determines the connection weights, and the idea that naming involves a direct mapping from orthography to phonology. Seidenberg and McClelland also discuss several details of the implemented model that are less theoretically relevant, such as the specifics of the orthographic and phonological encoding schemes, or the particular stimulus set used in training. They argue that the model's ability to capture detailed aspects of normal performance is unlikely to be contingent on these aspects of the implementation. However, we cannot as yet determine exactly how these specifics relate to the effects that we have (and have not!) obtained in regard to surface dyslexia.

For example, there are known limitations to the phonological encoding scheme used in this model and in Rumelhart and McClelland (1986) (see Pinker and Prince 1988; Lachter and Bever 1988). Similarly, it is not clear whether the model's treatment of lexical frequency is adequate; words were sampled during the training phase on the basis of a logarithmic transformation of their Kucera and Francis frequencies. Frequencies in the Kucera and Francis analysis range from about 67 000 to 1; in our scheme the range is only about 16 to 1. The results to this point suggest that these and other aspects of the implementation have little impact on the model's ability to simulate normal performance; however, these limitations may be more important when we turn to making detailed predictions about the exact errors produced by patients. To take one example, the compression of word frequencies may be related to the absence of marked frequency effects in the model's impaired performance. Before any firm conclusions can be drawn, it will be necessary to evaluate versions of the model using different phonological encoding schemes, indices of frequency, amounts of training, etc.

A more severe limitation of the model is that we have not yet implemented procedures for converting the output that it computes into real pronunciations. But this is a limitation on what has been done, not on what can be done. Lacouture (1988), for example, has developed a naming model that computes phonological codes much like the present model. It also exhibits the main types of phenomena concerning, for example, frequency and regularity effects. In Lacouture's model, however, the computed phonological code acts as the input to an autoassociative mechanism, which serves to complete the partially specified phonological code. This pattern completion process could be seen as similar to the process of assembling an articulatory motor program.

The observations from these initial explorations with lesioning suggest to us that it will be important to examine other ways in which the model's

performance can degrade. In particular, we need to consider the possibility that the impaired performance characteristic of Type I patients, such as H.T.R. and M.P., does not derive from damage to knowledge representations at all. The computations performed by the model could be impaired in a variety of ways that do not involve damage to representations. For example, the model computes output by passing activation through the network. We have damaged the system by eliminating connections or units. Imagine, instead, that the model is fully intact, but the net activations of units are incorrectly computed. One characteristic of the implemented model, for example, is that the activation coming into a unit (which is a weighted sum of the activations along the lines coming into it) is passed through a logistic function to yield a net activation between 0 and 1. We could then ask what would happen to performance if activations of hidden units were pathologically limited to a level such as 0.8, preventing output units from being fully activated. What kind of articulatory code would be assembled on the basis of this damped output?

We suggest that this line of inquiry is worth pursuing because there is already some evidence that the kinds of errors characteristic of surface dyslexia can be produced by a system that is wholly undamaged. Consider the following experiment, which we have recently completed. Normal university-student subjects are asked to name words such as the ones presented to the model or to a patient like H.T.R. However, we impose a response deadline, such that subjects must initiate pronunciation earlier than normal. Under these conditions, subjects produce naming latencies that are roughly normal but they make substantially more errors. Moreover, these errors include the following (taken from the actual corpus of responses):

regularizations:	PINT→/pint/; PLAID→/plAd/; STEALTH→/stElth/; DONE→/dOn/
'visual' errors:	TROUGH→tough; BREAD→beard; WALL→well
other errors:	BUSH→/bish/; BURY→/bErE/; DROUGHT→/drOt/; BATH→/bEth/

We assume that these errors arise simply because the deadline forces subjects to begin assembling a pronunciation before the computation of the phonological code is completed. In our implemented model, the activations of output units are computed on a single sweep; in a more realistic model, the activations would build up over time (cf. McClelland 1979; Cohen, Dunbar, and McClelland 1988; Seidenberg and McClelland 1988*b*). The effect of the deadline would be realized by initiating the assembly process before the phonological nodes had reached asymptotic levels of activation. A similar outcome would obtain if nodes were pathologically prevented from reaching these levels.

We are not suggesting that performance under deadline conditions fully mimics the performance of any surface dyslexic. For one thing, the maximum error rate obtained for a normal subject in the deadline condition was 20 per cent, much lower than would be seen in a patient. This might be expected because imposing a deadline that encourages early assembly is not equivalent to a pathological condition that prevents units from reaching asymptotic levels of activation. However, the experiment does show that the types of errors that have been observed in neuropsychological case studies can be produced by subjects whose knowledge representations are intact; the relevance of this observation for accounts of surface dyslexia is a matter worth considering further.

Along the same lines, it is also worth considering whether Type I patients might begin to assemble pronunciations prematurely because their access to semantic information is grossly impaired. The naming task requires that the subject produce the correct pronunciation of a word. The demands of the task change somewhat when the stimuli include non-words, which lack a certifiably 'correct' pronunciation. There may be some trials on which normal subjects check the phonological code computed on the basis of orthography (as in our model) against a phonological code computed on the basis of the orthography→meaning→phonology 'route' implied by Fig. 7.1. Since the meaning-based routine is not available to Type I surface dyslexics (i.e. semantic dyslexics), it would never be checked. The absence of any feedback from other parts of the lexical system might result in relatively rapid use of the pathway from orthography to phonology; the subject has 'nothing to lose' by initiating pronunciation, so to speak.

Finally, one other possibility should be mentioned. Perhaps the simulation results are telling us that something very like the model we have proposed is relevant to normal performance but not to all cases of surface dyslexia. Perhaps the knowledge representations of patients such as H.T.R. and M.P. are damaged to the point where they no longer support pronunciation at all. The patients are none the less asked to pronounce words and non-words. Under these conditions they may utilize other types of knowledge relevant to pronunciation. It is possible that readers have formed some explicit generalizations about the correlations between spelling and pronunciation, perhaps stored in the form of 'rules'. These generalizations could arise in several ways. For example, they could be the detritus of the learning process; children are often taught to read by introducing explicit pronunciation rules. The 'rules' could also reflect generalizations about the properties of a complex computational mechanism like the one in our model. We ourselves often resort to such generalizations in summarizing the behaviour of the model. These generalizations are not accurate in detail and they do not reflect the actual underlying computational mechanisms. It is quite possible that our self-knowledge of complex perceptual and cognitive processes consists of

generalizations of this type. When the normal naming mechanism is impaired, then it is possible that patients rely upon this knowledge which, though of limited applicability, is sufficient to yield correct pronunciations of common spelling patterns.

If this conjecture is correct, we are back to a modified 'dual-route' model as the account of naming *disorders*. There is a normal naming mechanism, like the one in the implemented model; there is a second type of knowledge, definitely 'non-lexical', which supports the naming behaviour of at least some surface dyslexics. It remains to be seen how this account, offered here as speculation, will fare in the light of future evidence. Note, however, that while this account involves two naming mechanisms, it differs from the 'dual-route' model in critical respects. The main assumption of the dual-route model is that separate mechanisms are necessary in order to pronounce exception words on the one hand and non-words on the other (Coltheart 1987). Hence, both routines play a role in normal performance. Our model, in which a single mechanism supports the pronunciation of all types of letter strings, challenges this 'central dogma' of dual-route theories (Seidenberg 1988*b*). This second type of knowledge merely comes into play when the normal system is non-functional. Thus, even this version of the model cannot be taken as an implementation of the dual-route account.

Conclusions

As indicated in the introduction to lesioning the model, the possibility of an account of acquired dyslexia within the model of oral reading that we have discussed is of some considerable theoretical significance. Surface dyslexia, especially in conjunction with its contrasting pattern of impaired reading, phonological dyslexia, has suggested to many that there must be at least two separable routines for the translation of orthography to phonology. Dual-routine theories have already been challenged by the demonstration that the undamaged model can learn to read regular words, exception words and non-words with a single procedure. Such theories will be in further contention if patterns of acquired dyslexia are reproducible by means of damage to the model. Note that the preceding sentence and the first sentence of this paragraph use the modest words 'if' and 'possibility': we are not claiming that the model can now offer such an account, only that it looks promising and well worth further exploration.

By way of summary, the model in its current state does a good job of accounting for what we might term the first-order phenomena in naming, the performance of normal subjects in reading different types of words, and non-words. The model provides the only quantitative account of normal performance; moreover the fit between simulation and behavioural data is

quite close. The model also accounts for the second-order phenomena, such as the types of errors observed in cases of surface dyslexia. The unresolved questions concern third-order predictions, regarding the exact proportions of errors of different types, and the different patterns of performance associated with surface dyslexia. It is not surprising that it is at this level that questions arise concerning limitations of the implemented model. Although substantive questions remain to be addressed, we think that these initial efforts have opened an interesting line of inquiry that is likely to contribute to a deeper understanding of reading and its disorders.

Acknowledgement

We are grateful to Rosaleen McCarthy for providing us with a corpus of reading errors for the patient K.T., studied by McCarthy and Warrington (1986).

Notes

1. Most recent versions of dual-routine models actually posit three pronunciation processes, the 'non-lexical' or 'subword-level' procedure mentioned above, and two 'lexical' procedures that involve accessing stored representations of word pronunciations. These representations, it is argued, can be accessed in two ways: either directly (by a procedure that transcodes from orthography to lexical phonology) or indirectly (from orthography to semantic representations and then to lexical phonology). The hypothesis of a direct lexical but non-semantic procedure has been based partly on patterns of acquired reading disorders (see for example Schwartz, Saffran, and Marin 1980; Funnell 1983) but also on results from normal subjects concerning the interrelationships between naming of words and naming of pictures (see Durso and Johnson 1979, for relevant data and Warren and Morton 1983, for discussion). The main point germane to the present discussion is that in all these accounts, at least two procedures are considered necessary to accomplish successful naming of regular words, exception words, and non-words. The model described here does not reject the notion that written words might be pronounced with reference to their meanings; as Fig. 7.1 suggests, a word could be named by a two-stage process in which meaning is computed from the orthographic input, and the pronunciation from meaning. In contrast to dual- (or triple-) routine models, however, this 'lexical' pathway is not necessary for the pronunciation of any type of letter string. In sum, the model squeezes three routines into two, with the added caveat that the primary procedure for translating from orthography to phonology is sufficient for all types of letter strings.
2. Since this chapter was written, the model has been augmented with a new procedure for assessing its output. The procedure compares the model's output not just to the correct (specified) pronounciation but also to all other pronunciations that can be created by replacing a single phoneme in that word with some other phoneme; it then reports the best match. Since this search only covers a subset of the possible

phonological patterns, we cannot guarantee that the best match among this set of comparisons is the best possible match; but the procedure does provide more comprehensive information regarding the model's 'preferred pronunciations'. Using this procedure, Seidenberg and McClelland (1988*b*) demonstrated that the trained and undamaged model makes errors (i.e. cases where the best fit to the computed pattern is a pronunciation other than the correct one) on only 2.7 percent of the 2897 words in its training vocabulary. The new procedure is of particular value in assessing output from the lesioned model; for example, one can readily determine whether the incorrect best match for an exception word is an exact regularization. Recent simulations using the new procedure suggest that the high levels of damage (60 percent of hidden units) used in most of our initial experiments may not in fact provide the best approximation to surface dyslexia. Although the overall error rate is certainly higher when more hidden units are silenced, the proportion of errors corresponding to exact regularizations actually decreases. In a test using the exception words from Taraban and McClelland's (1987) experiment with 20 percent of hidden units zeroed, nearly half (47 percent) of the model's errors were exact regularizations. Although this is still a somewhat lower regularization rate than that shown by the surface dyslexic patients in Table 7.10, it is a step in the right direction; future explorations may provide still closer approximations.

References

Andrews, S. (1982). Phonological recoding: is the regularity effect consistent? *Memory & Cognition*, **10**, 565–75.

Beauvois, M. F. and Derouesné, J. (1979). Phonological alexia: three dissociations. *Journal of Neurology, Neurosurgery and Psychiatry*, **42**, 1115–24.

Brown, G. D. A. (1987). Resolving inconsistency: a computational model of word naming. *Journal of Memory and Language*, **26**, 1–23.

Caramazza, A. (1986). On drawing inferences about the structure of normal cognitive systems from the analysis of patterns of impaired performance: the case of single-patient studies. *Brain and Cognition*, **5**, 41–66.

Bub, D., Cancelliere, A., and Kertesz, A. (1985). Whole-word and analytic translation of spelling to sound in a non-semantic reader. In *Surface dyslexia: neuropsychological and cognitive studies of phonological reading* (eds. K. E. Patterson, J. C. Marshall, and M. Coltheart). Erlbaum, London.

Cohen, J., Dunbar, K., and McClelland, J. L. (1988). On the control of automatic processes: a parallel distributed processing model of the Stroop effect. AIP technical report 40, Department of Psychology, Carnegie-Mellon University, Pittsburgh, PA.

Coltheart, M. (1985). Cognitive neuropsychology and the study of reading. In *Attention and performance XI* (eds. M. I. Posner and O. S. M. Marin). Erlbaum, Hillsdale, NJ.

Coltheart, M. (1987). Functional architecture of the language-processing system. In *The cognitive neuropsychology of language* (eds. M. Coltheart, G. Sartori, and R. Job). Erlbaum, London.

Coltheart, M., Masterson, J., Byng, S., Prior, M., and Riddoch, J. (1983). Surface dyslexia. *Quarterly Journal of Experimental Psychology*, **35A**, 469–95.

Durso, F. T. and Johnson, M. K. (1979). Facilitation in naming and categorizing repeated pictures and words. *Journal of Experimental Psychology: Human Learning and Memory*, **5**, 449–59.

Ellis, A. W. and Young, A. W. (1988). *Human cognitive neuropsychology*. Erlbaum, London.

Funnell, E. (1983). Phonological processes in reading: new evidence from acquired dyslexia. *British Journal of Psychology*, **74**, 159–80.

Gernsbacher, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness and polysemy. *Journal of Experimental Psychology: General*, **113**, 256–81.

Glushko, R. J. (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance*, **5**, 674–91.

Hebb, D. O. (1949). *Organization of behavior*. Wiley, New York.

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

Hillinger, M. L. (1980). Priming effects with phonemically similar words: the encoding-bias hypothesis reconsidered. *Memory & Cognition*, **8**, 115–23.

Hinton, G. E., McClelland, J. L., and Rumelhart, D. E. (1986). Distributed representations. In *Parallel distributed processing*, Volume 1 (eds. D. E. Rumelhart and J. L. McClelland). MIT Press, Cambridge, Mass.

Humphreys, G. W. and Evett, L. J. (1985). Are there independent lexical and nonlexical routes in word processing? An evaluation of the dual-route theory of reading. *The Behavioral and Brain Sciences*, **8**, 689–740.

Kay, J. (1987). Phonological codes in reading: assignment of sub-word phonology. In *Language perception and production* (eds. A. Allport, D. MacKay, W. Prinz, and E. Scheerer). Academic Press, London.

Kay, J. and Lesser, R. (1985). The nature of phonological processing in oral reading: evidence from surface dyslexia. *Quarterly Journal of Experimental Psychology*, **37A**, 39–81.

Kay, J. and Patterson, K. (1985). Routes to meaning in surface dyslexia. In *Surface dyslexia: neuropsychological and cognitive studies of phonological reading* (eds. K. E. Patterson, J. C. Marshall, and M. Coltheart). Erlbaum, London.

Kucera, H. and Francis, W. N. (1967). *Computational analysis of present-day American English*. Brown University Press, Providence, Rhode Island.

Lachter, J. and Bever, T. G. (1988). The relation between linguistic structure and associative theories of language learning. *Cognition*, **28**, 195–247.

Lacouture, Y. (1988). A connectionist model of the lexicon. Unpublished PhD thesis, McGill University.

McCarthy, R. and Warrington, E. K. (1986). Phonological reading: phenomena and paradoxes. *Cortex*, **22**, 359–80.

McClelland, J. L. and Rumelhart, D. E. (1986). *Parallel distributed processing*, Volume 2. MIT Press, Cambridge, Mass.

Marcel, T. (1980). Surface dyslexia and beginning reading: a revised hypothesis of the pronunciation of print and its impairments. In *Deep dyslexia* (eds. M. Coltheart, K. Patterson, and J. C. Marshall). Routledge and Kegan Paul, London.

Marshall, J. C. and Newcombe, F. (1973). Patterns of paralexia: a psycholinguistic approach. *Journal of Psycholinguistic Research*, **2**, 175–99.

Masterson, J. (1985). On how we read non-words: data from different populations. In *Surface dyslexia: neuropsychological and cognitive studies of phonological reading* (eds. K. E. Patterson, J. C. Marshall, and M. Coltheart). Erlbaum, London.

Meyer, D. E., Schvaneveldt, R. W., and Ruddy, M. G. (1974). Functions of graphemic and phonemic codes in visual word recognition. *Memory & Cognition*, **2**, 309–21.

Morton, J. and Patterson, K. (1980). A new attempt at an interpretation, or, an attempt at a new interpretation. In *Deep dyslexia* (eds. M. Coltheart, K. Patterson, and J. C. Marshall). Routledge and Kegan Paul, London.

Patterson, K. E. (1982). The relation between reading and phonological coding: further neuropsychological observations. In *Normality and pathology in cognitive functions* (ed. A. W. Ellis). Academic Press, London.

Patterson, K. and Coltheart, V. (1987). Phonological processes in reading: a tutorial review. In *Attention and performance XII: the psychology of reading* (ed. M. Coltheart). Erlbaum, London.

Patterson, K., Marshall, J. C., and Coltheart, M. (1985). *Surface dyslexia: neuropsychological and cognitive studies of phonological reading*. Erlbaum, London.

Patterson, K. and Morton, J. (1985). From orthography to phonology: an attempt at an old interpretation. In *Surface dyslexia: neuropsychological and cognitive studies of phonological reading* (eds. K. Patterson, J. C. Marshall, and M. Coltheart). Erlbaum, London.

Pinker, S. and Prince, A. (1988). On language and connectionism: analysis of a parallel distributed processing model of language acquisition. *Cognition*, **28**, 73–194.

Rumelhart, D. E., Hinton, G. E., and McClelland, J. L. (1986). A general framework for parallel distributed processing. In *Parallel distributed processing*, volume 1 (eds. D. E. Rumelhart and J. L. McClelland). MIT Press, Cambridge, Mass.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning internal representations by error propagation. In *Parallel distributed processing*, volume 1 (eds. D. E. Rumelhart and J. L. McClelland). MIT Press, Cambridge, Mass.

Rumelhart, D. E. and McClelland, J. L. (1986*a*). *Parallel distributed processing*, Volume 1. MIT Press, Cambridge, Mass.

Rumelhart, D. E. and McClelland, J. L. (1986*b*) On learning the past tenses of English verbs. In *Parallel distributed processing*, Volume 2 (eds. D. E. Rumelhart and J. L. McClelland). MIT Press, Cambridge, Mass.

Schwartz, M. F., Saffran, E. M., and Marin, O. S. M. (1980). Fractionating the reading process in dementia: evidence for word-specific print-to-sound associations. In *Deep dyslexia* (eds. M. Coltheart, K. Patterson, and J. C. Marshall). Routledge and Kegan Paul, London.

Seidenberg, M. S. (1985). The time-course of phonological code activation in two writing systems. *Cognition*, **19**, 1–30.

Seidenberg, M. S. (1988*a*). Visual word recognition and pronunciation: A computational model and its implications. In *Lexical representation and process* (ed. W. D. Marslen-Wilson). MIT Press, Cambridge, Mass., in press.

Seidenberg, M. S. (1988*b*). Cognitive neuropsychology and language: the state of the art. *Cognitive Neuropsychology*, **5**, 403–26.

Seidenberg, M. S. and McClelland, J. L. (1988*a*). A distributed, developmental model of visual word recognition and pronunciation: acquisition, skilled performance, and dyslexia. In *From reading to neurons: toward theory and methods for research on developmental dyslexia*. MIT Press/Bradford Books, Cambridge, Mass.

Seidenberg, M. S. and McClelland, J. L. (1988*b*). A distributed, developmental model of visual word recognition and naming. *Psychological Review*, in press.

Seidenberg, M. S., Waters, G. S., Barnes, M. A., and Tanenhaus, M. K. (1984). When does irregular spelling or pronunciation influence word recognition? *Journal of Verbal Learning and Verbal Behavior*, **23**, 383–404.

Sejnowski, T. J. and Rosenberg, C. R. (1986). *NETtalk: a parallel network that learns to read aloud*. Baltimore: Johns Hopkins University EE and CS Technical Report JHU/EECS-86/01.

Shallice, T. and McCarthy, R. (1985). Phonological reading: from patterns of impairment to possible procedures. In *Surface dyslexia: neuropsychological and cognitive studies of phonological reading* (eds. K. Patterson, J. C. Marshall, and M. Coltheart). Erlbaum, London.

Shallice, T. and Warrington, E. K. (1980). Single and multiple component central dyslexic syndromes. In *Deep dyslexia* (eds. M. Coltheart, K. Patterson, and J. C. Marshall). Routledge and Kegan Paul, London.

Shallice, T., Warrington, E. K., and McCarthy, R. (1983). Reading without semantics. *Quarterly Journal of Experimental Psychology*, **35A**, 111–38.

Tanenhaus, M. K., Flanigan, H., and Seidenberg, M. S. (1980). Orthographic and phonological code activation in auditory and visual word recognition. *Memory & Cognition*, **8**, 513–20.

Taraban, R. and McClelland, J. L. (1987). Conspiracy effects in word recognition. *Journal of Memory and Language*, **26**, 608–31.

Treiman, R. and Chafetz, J. (1987). Are there onset- and rime-like units in printed words? In *Attention and performance XII: the psychology of reading* (ed. M. Coltheart). Erlbaum, London.

Warren, C. and Morton, J. (1982). The effects of priming on picture recognition. *British Journal of Psychology*, **73**, 117–29.

Waters, G. S. and Seidenberg, M. S. (1985). Spelling-sound effects in reading: time course and decision criteria. *Memory & Cognition*, **13**, 557–72.

Wickelgren, W. A. (1969). Context-sensitive coding, associative memory, and serial order in (speech) behaviour. *Psychological Review*, **76**, 1–15.