In J. Reggia, R. Berndt, & E. Ruppin (Eds.), *Neural modeling of cognitive and brain disorders* (pp. 177–199). New York: World Scientific, 1996.

CONNECTIONS AND DISCONNECTIONS:
A CONNECTIONIST ACCOUNT OF SURFACE DYSLEXIA

KARALYN PATTERSON
Medical Research Council Applied Psychology Unit
15 Chaucer Road, Cambridge CB2 2EF, UK, karalyn@mrc-apu.cam.ac.uk

DAVID C. PLAUT
JAMES L. MCCLELLAND
Department of Psychology, Carnegie Mellon University,
and the Center for the Neural Basis of Cognition, Pittsburgh, USA

MARK S. SEIDENBERG
Neuroscience Program, University of Southern California, Los Angeles, USA

MARLENE BEHRMANN
Department of Psychology, Carnegie Mellon University, Pittsburgh, USA

JOHN R. HODGES
Department of Neurology, University of Cambridge Clinical School, Cambridge, UK

The acquired reading disorder of surface dyslexia, in which lower-frequency words with atypical spelling-sound correspondences (e.g., PINT) become highly vulnerable to error, is presented in a framework based on interaction between distributed representations in a triangle of orthographic, phonological, and semantic domains. The framework suggests that low-frequency exception words are rather inefficiently processed in terms of orthographic-phonological constraints, because these words are neither sufficiently common to have much impact on learning in the network nor sufficiently consistent with the pronunciations of their orthographic neighbors to benefit from shared structure. For these words, then, the interaction between phonological and semantic representations may be especially important for settling on the correct pronunciation. It is therefore viewed as no coincidental association that all reported patients with marked surface dyslexia have also been profoundly anomic, suggesting reduced semantic-phonological activation. The chapter summarizes the simulation of surface dyslexia in the computational model of reading developed by Plaut, McClelland, Seidenberg, and Patterson (1996), and presents new data from three surface alexic patients. The graded consistency effects in the patients' reading performance are more compatible with the distributed connectionist framework than with dual-route models maintaining a strict dichotomy between regular and exception words.

1 Background

One of the main issues in reading research concerns the procedures that readers use to compute the pronunciation of written words. According to the dominant view, two different types of letter string in English demand two qualitatively different procedures: the ability of readers to pronounce letter strings that they have never seen before (for example, a nonsense word such as MAVE) requires a procedure based on rules for translating graphemes to phonemes; and the ability to pronounce familiar "irregular" words that break standard grapheme-phoneme correspondence rules (such as HAVE) requires a procedure in which the word's orthographic lexical entry activates its whole-word pronunciation. Considerable evidence—from normal adult readers, from adults with acquired disorders of reading, and from children with developmental dyslexia—has been interpreted as support for this dual-route model, mainly by Coltheart and his colleagues (Castles & Coltheart, 1993; Coltheart, 1985; Coltheart, Curtis, Atkins & Haller, 1993; Coltheart & Rastle, 1994) but by many other reading researchers as well (e.g., Baluch & Besner, 1991; Funnell, 1983; Paap & Noel, 1991). Dual-route theory, instantiated in computational models both by Coltheart et al. (1993) and by Reggia, Berndt and D'Autrechy (1994), is sufficiently well known and accessible in the literature to obviate the need for a full description here. Simply put, its basic premise is that no model of reading will succeed in explaining the known data unless it incorporates separate lexical and sub-lexical mechanisms for translating an orthographic string into a pronunciation (hereinafter O→P).

Despite the prominence and success of dual-route theory, it has had its critics and alternatives. For example, a number of reading researchers have argued that regular and exception words correspond to points on a consistency continuum rather than a dichotomy (Glushko, 1979; Seidenberg, Waters, Barnes & Tanenhaus, 1984; Shallice, Warrington & McCarthy, 1983). Others have claimed that nonword reading, instead of requiring a separate rule-based "non-lexical" system, could be accomplished by extracting and pooling knowledge from lexical representations for structurally similar words (Henderson, 1982; Humphreys & Evett, 1985; Kay & Marcel, 1981). In the last decade, the proposal that a single O→P mechanism is in fact capable of capturing both the generalizations and the exceptions in spelling- sound relationships has been developed in computational models of reading aloud, first by Sejnowski and Rosenberg (1987) and subsequently by Seidenberg and McClelland (1989). The Seidenberg and McClelland model, as a major theoretical statement about the acquisition and skilled performance of single-word reading, attracted considerable attention, much positive but some critical. For example, dual-route theorists (Besner, Twilley, McCann & Seergobin, 1990; Coltheart et al., 1993) contested the claim that Seidenberg and McClelland had demonstrated the adequacy of a single mechanism for reading both exception words and nonwords, because the original simulation achieved notably less success than most human readers do in generalizing its knowledge to the pronunciation of nonwords.

In the most recent phase of this debate, Plaut, McClelland, Seidenberg and Patterson (1996) presented four new simulations of the O→P computation in English. As one principal development on their predecessor (Seidenberg & McClelland, 1989), the networks in the Plaut et al. model employed orthographic and phonological representations designed to capture more successfully the similarities in orthographic and phonological space. This new design of representations enabled the model to attain accuracy in nonword reading well within the range of real adult readers. Three of the networks had a feedforward architecture; the fourth was an attractor network involving interactivity among the phonological output units and between the phonological units and the hidden-unit layer. Two of the simulations were trained using actual (Kucera & Francis, 1967) frequencies of the 3000 words in the corpus, while the others had some degree of frequency compression, either more or less severe (logarithmic and square-root, respectively). Detailed analyses of the results from these various simulations can be found in the original article; for present purposes, the important summary is the following. Given (a) orthographic and phonological representations that effectively capture spelling-sound consistencies, making the network appropriately sensitive to the range of consistencies in the training vocabulary, and (b) a training regime based on real or approximate word frequencies, making the performance of the network appropriately sensitive to the impact of word frequency, a network with a single O→P procedure can reproduce the pattern of accuracy and response times in naming regular words, exception words, and nonwords that is characteristic of real adult readers.

These results from the fully trained networks of Plaut et al. (1996) establish that a single mechanism, in addition to learning to pronounce both familiar regular words and exceptions, can generalize to novel words. Criticisms of Seidenberg and McClelland's (1989) model of reading, however, did not focus exclusively on its normal reading performance. Early attempts to simulate one prominent form of acquired reading disorder by damaging the Seidenberg and McClelland network had been acknowledged even by its authors as provocative but insufficient (Patterson, Seidenberg & McClelland, 1989); and Coltheart et al. (1993) argued that separate lexical and non-lexical routes are essential to account not only for the correct pronunciation of both nonwords and exception words by normal readers but also for the patterns of performance observed in neurologically acquired disorders of word naming. Another major component of the work by Plaut et al. (1996), therefore, was addressed to the issue of whether and how such disorders, in particular acquired surface dyslexia, might find an explanation in a model that dispenses with separate lexical and non-lexical procedures.

2 Surface Dyslexia

Surface dyslexia is one of the main forms of reading disorder observed when the previously competent reading ability of an adult is disrupted by brain injury or disease (Shallice & Warrington, 1980). This disorder was given its name by Marshall and Newcombe (1973) to convey the idea that, when surface dyslexic patients read a word aloud incorrectly, their errors typically reflect the "surface" structure of the word; the syndrome was contrasted with deep dyslexia, in which errors were construed as reflecting the word's "deep" structure. For the written word PINT, the typical surface dyslexic's error would be /pInt/ (i.e., pronounced like MINT and indeed every word with the spelling pattern -INT in English, except for PINT), whereas a deep dyslexic patient misreading pint would be likely to respond "quart" or possibly "beer" (at least in Britain, where beer in pubs is still served in pints). In describing the typical reading errors of surface dyslexia—often called *regularization* errors, because the irregular word pint is pronounced like its regular neighbors MINT, LINT, PRINT, etc.—Marshall and Newcombe (1973) had identified one of the most salient characteristics of the disorder. Further research (Behrmann & Bub, 1992; Bub, Cancelliere & Kertesz, 1985; McCarthy & Warrington, 1986; and Shallice, Warrington & McCarthy, 1983) established that, in its purest form, acquired surface dyslexia is characterized by reading performance on regular words and nonwords that is within normal limits of both accuracy and speed, and a deficit on irregular words that is strongly modulated by word frequency.

The account of surface dyslexia in Coltheart's DRC (Dual-Route Cascaded) model is as follows: the non-lexical grapheme-phoneme route, which can correctly compute O→P for regular words and nonwords, is intact; the lexical route, which is necessary for correct pronunciations of exception words, is damaged in a manner that still enables success on a high-frequency exception word like HAVE but fails on a less common word like PINT, forcing the patient to respond with the non-lexical route's output for this word, i.e., the regularization error /pInt/. As with the mainstream dual-route account of normal reading, this interpretation has been the leading, but not quite the only, bid on the surface dyslexic table. According to Marshall and Newcombe (1973, 1980)—and, with minor variations, to Hillis and Caramazza (1991) and Howard and Franklin (1988)—an adequate theory of O→P requires two routes, but not precisely the same two as proposed by Coltheart (1985; Coltheart et al., 1993). In these conceptions, the lexical route is a lexical semantic procedure; thus a written word can be translated to a phonological code either by sub-lexical correspondences or by activation of the word's meaning followed by the processes normally used, in object naming and spontaneous speech, to activate phonology from meaning. Surface dyslexic reading is thought to arise from a combination of intact sub-lexical procedure and damaged lexical-semantic route. The account of surface dyslexia offered by Plaut et al. (1996) differs somewhat from all of the above; as will be seen in a moment, however, in one crucial respect it is more akin to these alternative dual-route proposals than to Coltheart's view.

Before we explain our position, we should say that it is still evolving. We present our somewhat preliminary account here in the following spirit: McClelland (in his final discussion at the meeting that engendered this book) emphasized that, although

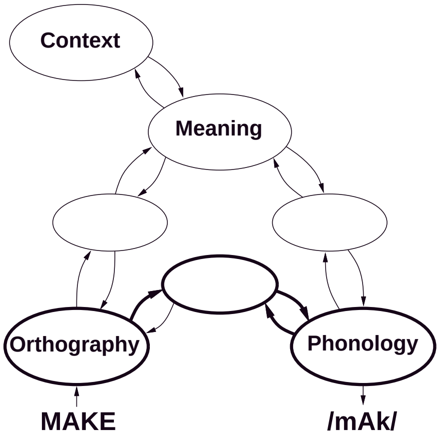

Figure 1: The “triangle” framework for single-word processing from Seidenberg and McClelland (1989) and Plaut, McClelland, Seidenberg and Patterson (1996).

all current models are bound to be wrong, they may nevertheless be groping their way towards some important, and even correct, underlying principles.

3 An Account of Surface Dyslexia in the Plaut et al. (1996) Framework

In the final empirical section of their paper, Plaut et al. turned their attention to surface dyslexia. While stopping short of agreeing with Coltheart’s claim that damage to a single procedure never could, in principle, account for surface dyslexia, they acknowledged that the dramatic pattern of pure surface dyslexia (i.e., normal reading aloud of regular words and nonwords coupled with a severe, frequency-sensitive deficit on exception words) seems unlikely to arise from damage to the kind of single, direct O→P computation developed thus far. “Lesions” to the network sufficiently severe to reproduce the appropriate degree of impairment on exception words also disrupt the model’s performance on regular and nonwords. The Plaut et al. model, however, like the Seidenberg and McClelland model before it, and also like the views of single-word processing proposed by Bullinaria (in press), Kawamoto and Zemblige (1992), and Van Orden and Goldinger (1994), has the broader “triangle” framework sketched in Figure 1. Therefore, despite the demonstration that a fully trained model of direct O→P translation is capable of learning to produce correct pronunciations for all sorts of letter strings, it may be that the typical human reader does not solve the O→P problem in precisely this way. In particular, the possibility remains that some aspects of skilled written word pronunciation might rely on access to semantic representations of words (S). This influence could occur on the basis of either O→S→P activation or O→P↔S activation and interaction; in either case, the important component is the S→P link. The account developed in Plaut et al. is that communication from meaning to phonology is particularly critical for processing of words that are only weakly learned by direct O→P, namely low-frequency words with atypical spelling-to-sound correspondences, and that pure surface dyslexia is thus attributable to reduced activation from S→P. This is not a new idea: although the nature of the direct O→P translation in Plaut et al. differs from that in other theories of the reading process, recall that Hillis and Caramazza (1991), Howard and Franklin (1988), and Marshall and Newcombe (1973, 1980) all implicated word meaning in their accounts of surface dyslexia.

Before presenting the simulation work from Plaut et al. (1996) designed to explore this hypothesis, we shall briefly summarize existing evidence about human reading performance, both normal and abnormal, which supports the idea that semantic representations of words should be considered germane to the process of translating print to pronunciation.

(1) In the usual context for which reading skills are mobilized—text reading —few people would doubt that the pronunciation of written words must be open to semantic influences. For example, readers correctly pronounce heterophonic homographs (such as WIND, LEAD and BASS) when they are reading text aloud. Assuming that the human O→P direct computation has learned to activate both legitimate pronunciations of such words (like the networks in Plaut et al., which were trained on both), the pronunciation appropriate to a particular context (e.g., of the noun WIND, that blows, or the verb WIND, that one used to do to watches) can be selected via S→P activation.

(2) In a study of accuracy and response times (RTs) for single-word naming by normal adult readers, Strain, Patterson and Seidenberg (1995) manipulated a semantic variable, imageability, in their selection of stimulus words. Hypothesizing that a significant impact of this variable on word naming should be observable mainly for words rather weakly supported by the direct O→P procedure, they also included the variables of word frequency and regularity. As predicted, a disadvantage in both accuracy and RT for words with low-imageability ratings was obtained primarily for lower-frequency words with atypical spelling-sound correspondences. This does not of course mean that, for words that are either commonly encountered or that fit the most common spelling-sound patterns, there is no automatic activation of word meaning; for such words, however, O→P activation on its own may be sufficient to achieve rapid and stable phonological representations. This computation is less effective for low-frequency inconsistent words; and then communication from S→P (which is stronger for imageable words with richer semantic representations) may detectably

assist in settling on a pattern of phonological activation.

(3) If surface dyslexia is attributable to some disruption in communication between semantic and phonological representations, then patients with acquired surface dyslexia should be anomic, since naming of objects or concepts relies on the S→P link. As far as we know, there are no exceptions to this association; that is, all published cases of patients with acquired "pure" surface reading have also had a prominent anomia.

(4) It appears that the association between surface dyslexia and anomia may sometimes even respect category specificity: DRB (Franklin, Howard & Patterson, 1995), who was only measurably anomic for abstract words and concepts, was also significantly impaired in reading exception words only if they had abstract meanings.

(5) The entailment or prediction in this account is only for a disruption in communication from S→P. It should not matter, in principle, whether this difficulty arises from degraded semantic representations or from a reduced capacity for activating phonology from meaning. Although the majority of reported cases of pure surface dyslexia have had a profound impairment to semantic memory per se (e.g., Breedin, Saffran & Coslett, 1994; Bub et al., 1985; Funnell, in press; McCarthy & Warrington, 1986; Parkin, 1993; Patterson & Hodges, 1992), there are also surface dyslexic patients whose deficits in naming and reading have been assigned to the S→P link (Graham, Patterson & Hodges, 1995; Watt, Jokel & Behrmann, in press).

(6) In studies of patients whose surface dyslexia is attributed to degraded semantic memory, three separate groups of investigators have reported a significant concordance between reading aloud and comprehension: that is, with irregular words, the items that are named correctly (as opposed to regularized) tend significantly to be the same ones on which the patient succeeds in a comprehension test such matching spoken words to pictures (Funnell, in press; Graham, Hodges & Patterson, 1994; Hillis & Caramazza, 1991).

Finally, it is important to note that surface alexia may sometimes be masked by additional deficits, especially at the level of speech production. One perspicacious question about our predicted association between impaired word comprehension and surface dyslexia has been: should one not then find a common association between surface dyslexia and Wernicke's aphasia, a language disorder standardly interpreted as a comprehension deficit? Our first response to this question is that speech production in Wernicke's aphasia is often so disturbed and distorted by phonological problems that it might be hard to observe an advantage in oral reading for regular over exception words. Our second response is that at least one Wernicke's aphasic patient without a profound phonological deficit has shown the frequency-by-regularity interaction in reading accuracy that is characteristic of surface alexia (Behrmann, unpublished data).

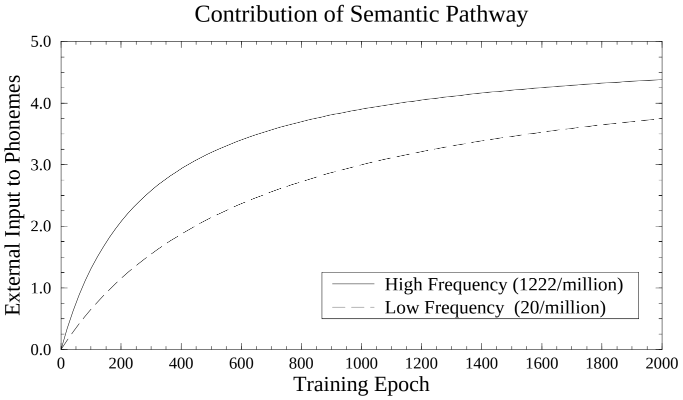

Figure 2: An illustration of the gradual addition of notional semantic input to the phoneme units during O→P training in the Plaut et al (1996) simulation of surface dyslexia; shown here for the Taraban and McClelland (1987) high- and low-frequency words.

4 Simulation of Surface Dyslexia by Plaut et al. (1996)

The instantiation by Plaut et al. (1996) of the way in which word meaning might influence O→P translation did not attempt to provide a genuine representation of meaning; rather, the contribution of meaning was approximated by providing an extra source of input to the phoneme units to push them toward their correct activations. The simulation used a feedforward network with square-root compression of word frequency, plus a small weight decay factor; this biases the network to keep weights small and has the effect of preventing overlearning. The major change from the previous simulations performed without any semantic component was that, over the course of O→P training, the additional source of input to phoneme units (notional semantics) was gradually introduced. The basis for this graded procedure was the assumption that, because O→P mappings in an alphabetic orthography (even a quasi-regular one like English) are much more systematic than O→S mappings, beginning readers learn the former more rapidly than the latter. Furthermore, a larger additional input was provided for high- than for low-frequency words on the assumption that real O→S learning is stronger for frequently encountered words. The gradual and frequency-modulated additional input to phoneme units is illustrated in Figure 2, for a subset of the 2998 monosyllabic words in the training corpus, namely the high- and low-frequency words used in experiments by Taraban and McClelland (1987).

Performance during training is illustrated in Figure 3 for the sets of Taraban and McClelland (1987) words and Glushko (1979) nonwords. At early stages of training (epochs 50–100), adequacy of word pronunciations (percent correct, i.e., whether the

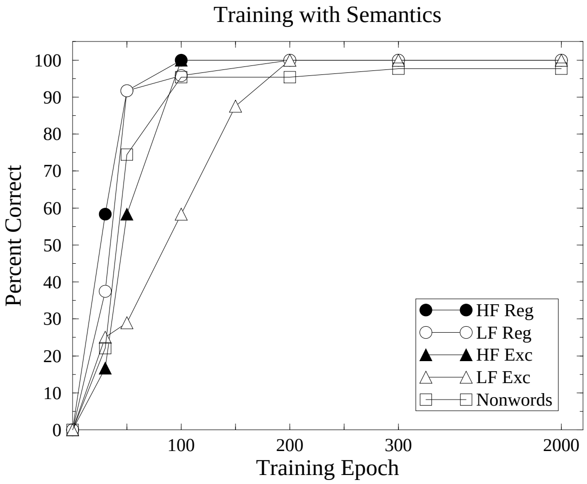

Figure 3: Performance of the Plaut et al (1996) O→P network trained with additional input (notional semantics) to the phoneme units, at various stages of training, on the Taraban and McClelland (1987) high-and low-frequency regular and exception words and the Glushko (1979) nonwords.

correct phoneme unit at each segment of the monosyllable—onset, vowel, coda—is the most active unit) clearly varies as a function of both frequency and consistency; and performance on items which benefit from neither frequency nor consistency (LF Exc words in Figure 3) is still not perfect at epoch 150. By epoch 200, all word pronunciations are correct, and nonwords (which receive no additional "semantic" input) have achieved a level of 95% correct pronunciations. Although this network yields frequency and regularity effects characteristic of networks trained without notional semantics, the net with additional input naturally learns faster than simulations with only O→P input. A net trained with the identical learning parameters and initial random weights, but with no semantic contribution, reached asymptote on percent correct at around epoch 500 rather than 200.

From epoch 200 to 2000, when responses to the different word sets can no longer be distinguished by the percent correct measure (see Figure 3), a significant frequency-by-consistency interaction in the model's performance is still observable in the cross-entropy error score, a measure of the discrepancy between the pattern of activation over the phoneme units generated by the network and the precise target pattern of activation. On the assumption that output patterns approximating more closely to "perfect" will support more rapid responses, this continuing sensitivity to frequency and consistency in the model's fully trained performance can be seen as an analogue of the human skilled reader's response times to name high- and low-frequency regular

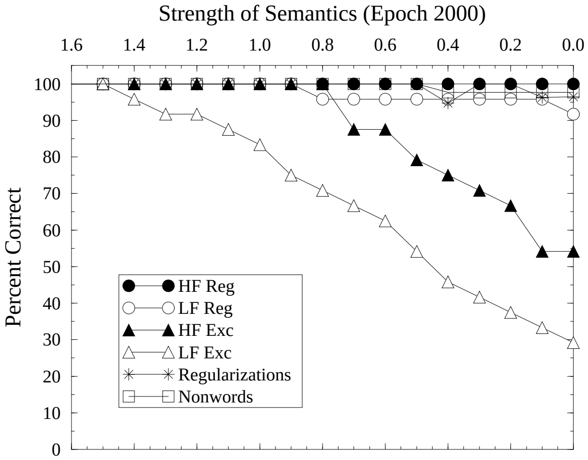

Figure 4: Performance of the Plaut et al (1996) O→P network trained with additional input (notional semantics) to the phoneme units for 2000 epochs, when the strength of the semantic input is gradually reduced; results are shown for percent correct on the Taraban and McClelland (1987) words and the Glushko (1979) words, and for percentage of errors to Exc words that are regularizations.

and exception words.

The results particularly germane to modeling surface dyslexia are presented in Figure 4, which shows the performance of the network that had been trained with notional semantics for 2000 epochs, when the strength of this additional source of activation to the output units is gradually reduced. Although surface dyslexia can result from an abrupt brain insult like a cerebrovascular accident or a head injury, the great majority of reported cases of pure surface dyslexia have suffered from progressive brain disease: either semantic dementia (e.g., Hodges, Graham & Patterson, 1995), in which the primary impairment is apparently in conceptual knowledge itself, or a progressive aphasia which seems to affect mainly the link between semantic and phonological representations (e.g., Graham et al., 1995; Watt et al., in press). In both cases, we assume that there is a gradual reduction in the activation of phonology by semantics, and the simulation was intended to mimic this phenomenon by post-training withdrawal of the additional input to the phoneme units, leaving only orthographic input to phonology.

Although Figure 4 represents post-training performance, what it reveals is the effect of the notional semantic input during training. Because this second source of input pushes the activation of each output unit in the appropriate direction—up if it should be on for the target word, down if it should be off—the gradual increase in semantic input in the training phase reduces the amount of error in the network's responses, which in turn reduces pressure for further learning in the O→P computation. As error decreases, the weights become smaller (recall that this net was also trained with a tendency for weight decay). Larger weights are especially important for exception words, which must compete against the conspiracy of many smaller weights supporting typical spelling-sound correspondences. The impact of the additional semantic input during training is therefore an actual decline in O→P competence, first for LF Exc words and eventually, to some degree, even for HF Exc words. This effect is not observable in the model's accuracy during training because the incrementing semantic input keeps improving the net's performance. When, in the simulation of progressive brain disease, this additional input is gradually withdrawn, the underlying competence of O→P knowledge is revealed. As a function of decreasing strength of semantic input, accuracy of pronunciation (a) remains high for Reg words of any frequency and for nonwords, (b) declines steadily on LF Exc words, and (c) also declines on HF Exc words, though the vulnerability of these commonly encountered words is slower to emerge and always less dramatic.

There is one further curve in Figure 4, somewhat difficult to resolve visually from those indicating essentially ceiling accuracy on regular words and nonwords: the points represented by asterisks indicate the nature of the network's error responses to exception words. Essentially all of the errors were regularizations of the PINT → /pInt/ variety. In the purest cases of reported human surface dyslexia (e.g., Behrmann & Bub, 1992; McCarthy & Warrington, 1986), one also observes virtually perfect reading aloud of regular words and nonwords and almost exclusively regularization errors to exception words.

Although this is a preliminary and certainly incomplete account of surface dyslexia, the success of the simulation is considerable. The values of accuracy for the different word classes at various points along the abscissa in Figure 4 correspond reasonably well to those observed for genuine cases of surface alexia. For example, on similar (though not identical) sets of monosyllabic words, PB and KT (two of the progressive patients reported by Patterson & Hodges, 1992, Table 2, p. 1030) both scored 90–100% correct on high- and low-frequency regular words. PB's scores for the high- and low-frequency exception words were 86% and 48%, respectively, which is very close to the network's performance with "strength of semantics" reduced to 0.5; and KT's high- and low-frequency exception scores were 50% and 8%, respectively: not too far off the model's performance at the point where all putative semantic support has been withdrawn, though the patient's LF Exc value is clearly poorer than the net's. Critically, KT's comprehension loss was much more severe than PB's as measured by other tests such word-picture matching. We acknowledge that our framework (and perhaps those of others as well) is a long way from an understanding of individual differences in both normal and impaired reading skill. In particular, the ideas developed here have recently been challenged by reports of a few patients with deficits on semantic tests who are within the normal range of performance in reading exception words (Cipolotti & Warrington, 1995; Lambon-Ralph, Ellis & Franklin, 1995; Raymer & Berndt, 1994). Some possible interpretations of this dissociation, which is apparently incompatible with our prediction, are discussed in Plaut et al. (1996).

We conclude from this simulation work that, although the story is perhaps somewhat different from and more complicated than that originally envisioned, patterns of both normal and disordered reading can be understood in the framework proposed by Seidenberg and McClelland (depicted in Figure 1). Some, but not all, of the critique of this approach by Coltheart et al. (1993) was apposite. Contrary to their critique, the initial simulations of Plaut et al. (1996) demonstrate that a single mechanism for activating phonology from orthography is capable of acquiring knowledge that supports both pronouncing exceptions and generalizing to novel forms. In line with their critique, however, lesions to such networks do not reproduce the pattern of pure surface alexia. We therefore acknowledge that another "pathway" or, as we have characterized it, another source of input to phonology appears to be necessary to model surface alexia. Perhaps an even better term is the one employed by Kawamoto and Zemblige (1992): source of constraint. They argue that their experimental results (which concern the precise time-course of normal skilled pronunciation of heterophonic homographs like bass and lead) can be explained by a distributed perspective, but only if it provides two constraints on pronunciation, orthographic and semantic. On the basis of a quite different set of human and network results, we have reached the same conclusion.

Despite this new apparent similarity between dual-route and triangle perspectives, some important differences remain. For one thing, and this is the reason that we prefer the phrase "source of input or constraint," we do not think in terms of two wholly separate pathways or routes. Unlike the DRC model of Coltheart et al. (1993), we assume that the same orthographic and phonological representations support both word and nonword reading. Although it has not been fully implemented, the triangle model of Figure 1 is meant to represent a genuinely interactive, recurrent system rather than the combination of two separate pathways, O→P and O→S→P. Secondly, it must be emphasized that the two sources of constraint in this approach do not divide the English language vocabulary neatly into rule-obeying and rule-infringing words as in the two routes of the DRC model. The orthographic source of constraint on pronunciation, even with additional semantic input, embodies considerable knowledge about the inconsistencies of O→P relationships, especially derived from its experience with the more common exception patterns. This is why both the model and the patients still pronounce a fair number of exception words correctly even when all semantic support has been eliminated. Furthermore, as we shall argue below, performance on words with intermediate degrees of consistency—irregular by strict "rule," but with support from other similarly structured words—finds a more natural explanation in the triangle framework than in dual-route theories.

5 Further Observations on Surface Dyslexic Reading

In the final section of this chapter, we introduce some further data from our recent studies of surface dyslexic patients that seem especially compatible with the kind of framework presented here.

5.1 Sub-Regularities

Table 1 presents the performance of three surface alexic patients[a] on 24 monosyllabic exception words which have atypical pronunciations of the body/ rime—words like PINT. The words in Set A were selected on a particular basis, to be explained in a moment, which led us to expect that surface alexic patients might attain an unusual degree of success in naming these particular exception words; the exception words in Set B were then chosen, from a larger list of words named by all of these patients, to match those in set a as closely as possible on two other measures known to affect performance:

(i) Kucera and Francis (1967) written word frequency. Where it was not possible to find a Set B word with a precise frequency match to an a item, the bias was towards selecting a more frequent word for B, to work against our prediction of greater success in Set A. This bias is reflected in the mean frequencies for the two sets, which are close but favor Set B.

(ii) The ratio of regular:exception pronunciations within the set of monosyllabic words sharing that particular orthographic body. Word 8a, for example, is PINT; of the monosyllabic English words ending in -INT, 12 have regular pronunciations (HINT, LINT, PRINT, etc); PINT is the only exception. For each (unique) body represented in Table 1, the regular exemplars outnumber the exception exemplar(s), in most cases by a substantial number, as reflected in the means on this measure shown in the Table.

Although the numbers of items for this contrast are small, all three patients showed a reliable or nearly reliable advantage for Set A over B: MP, 9/12 vs. 4/12, $\chi^2(1)$=4.2, p=.04; PB, 10/12 vs. 4/12, $\chi^2(1)$=6.2, p=.01; AM, 11/12 vs. 7/12, $\chi^2(1)$=3.6, p=.059. Combining the three patients' data yields a highly reliable contrast, 30/36 vs. 15/36, $\chi^2(1)$=13.3, $p<.001$. No word in Set B was named correctly by all patients, whereas 7/12 Set A words were given correct pronunciations by all three. Virtually all of the errors by all three patients on both sets of words were regularizations: all three named PINT to rhyme with "hint" and GROSS to rhyme with "moss"; two of the

---

[a] Descriptions of the general characteristics of the three patients in Table 1 can be found as follows: for MP: Behrmann and Bub (1992), Bub, Black, Hampson and Kertesz (1988) and Bub, Cancelliere and Kertesz (1985); for PB: Patterson and Hodges (1992); for AM: Hodges and Patterson (in press).

Table 1: Two sets of 12 exception words, showing—for each word—the Kucera-Francis frequency, the number of exemplars in the Body Neighborhood with Reg:Exc pronunciations, and the number of correct pronunciations (out of 3) for three surface alexic patients (MP, PB, AM); the lower part of the Table shows performance on the same sets of words by the Plaut et al. (1996) network with various degrees of reduced "semantic" input (as in Figure 4).

Set A				Set B			
Word	Freq	Reg:Exc	+/3	Word	Freq	Reg:Exc	+/3
1a	464	9:1	3	1b	437	9:1	1
2a	37	15:3	3	2b	88	14:2	2
3a	3	10:3	3	3b	14	9:3	2
4a	4	4:1	3	4b	11	5:1	2
5a	760	4:1	3	5b	391	10:2	2
6a	11	3:1	3	6b	16	2:1	2
7a	67	3:2	3	7b	58	8:3	1
8a	5	12:1	2	8b	13	12:1	0
9a	81	9:1	2	9b	66	9:1	0
10a	2	5:1	2	10b	84	5:1	2
11a	94	2:1	2	11b	730	4:2	0
12a	4	4:1	1	12b	9	6:4	1
Mean	127.7	6.7:1.4	0.83		159.8	7.8:1.8	0.42
Network's proportion correct							
semantic strength = 0.6			1.00				0.67
= 0.4			0.92				0.50
= 0.2			0.83				0.42
= 0.0			0.50				0.25

three pronounced PUT to rhyme with “hut” and LOSE to rhyme with “nose”; and so on.

When the Plaut et al. (1996) network (where additional “semantic” input to the phoneme units during training is subsequently reduced in strength, as in Figure 4 described above) is tested on the same 24 words, it yields a similar advantage for Set A over B, as shown in the lower part of Table 1. As it happens, when the strength of additional input is at 0.2, the network’s performance exactly matches the patients’: 83% correct on Set A, 42% correct on Set B. All of the network’s errors were regularizations.

What is it about Set A words that makes them relatively (though clearly not altogether) immune to regularization, by both patients and network? The answer, in our view, is that these words enjoy a kind of sub-regularity, based not on the body/rime but on the combination of the initial consonant and vowel. Almost all (10/12) of the words in set a begin either WA- or WO-, and the remaining two (SWAMP and QUART) have the same character.[b] Leaving aside words such as WAKE or WOKE where the pronunciation of the vowel is signalled by the final -E, a substantial majority of WA- words are pronounced not in accordance with the usual pronunciation of the vowel and body (as in CASH or CART) but rather like WASH or WART. Likewise, a great majority of WO- words are pronounced not with the vowel in NORTH but rather like WORTH. As discussed by Seidenberg (1992), in quasi-regular systems like spelling-sound correspondence and past-tense verb formation in English, a number of patterns that do not follow the most general rule are nonetheless characterized by this sort of shared irregularity, thus forming a sub-regularity. Spelling-sound knowledge apparently reflects this sub-regularity; as a result, compared to other exception words with similar familiarity levels and body neighborhoods, “W-words” depend less on the additional source of constraint on pronunciation and so are less vulnerable to its removal.

Dual-route models like that of Coltheart et al. (1993) can, of course, account for the relative invulnerability of W-words, but only by complicating the rule system.

---

[b] In view of demonstrations that at least normal readers’ pronunciations of words and nonwords may be subject to priming or biasing effects from other similar items in the list context (Kay & Marcel, 1981; Seidenberg et al., 1984), it should be noted that the patients were not asked to read the 12 W-words as a block; these items were embedded in, and well distributed throughout, a much larger list of words and nonwords (total $N = 198$) . Furthermore, as this list contained both regular words and nonwords with the same bodies as the W-words (e.g., FORK and LORK as well as WORK, FARM and DARM as well as WARM; words sharing bodies as well as onsets were well separated throughout the set), any biasing effect from the pronunciation of other W-words should have been offset by effects of these items with the same body but discrepant pronunciations.

5.2 The Fate of Regular Words in Progressive Surface Dyslexia

Over the past five years or so, various authors of this chapter have carried out detailed investigations of around a dozen patients with acquired surface dyslexia. All of these cases have had either a moderate-to-severe impairment of semantic memory (e.g., Behrmann & Bub, 1992; Patterson & Hodges, 1992) or at least a profound impairment in activation of phonology by semantics, as revealed for example by severe anomia (e.g., Graham, Patterson & Hodges, 1995; Watt, Jokel & Behrmann, 1996). All but one of the cases has suffered from a neuro-degenerative disease[c] characterized behaviorally as semantic dementia or progressive aphasia (see Hodges, Patterson, Oxbury & Funnell, 1992, for further description); the one exception is MP, who sustained a major head injury resulting in unusually focal damage to the left temporal lobe (Behrmann & Bub, 1992; Patterson & Behrmann, submitted). With a number of the progressive patients, we have been able to perform longitudinal assessments of reading performance, some of which are still in progress. On initial assessment, virtually all of the patients could be described as having a pattern of pure surface dyslexia, in the sense that their accuracy of word naming was notably outside normal limits for exception words but within normal range for regular words. This is illustrated in Figure 5 for PB and FM, two of the cases from the Patterson and Hodges (1992) study. These data are from the list of words employed in that study, which consists of 126 pairs of monosyllabic regular and exception words (e.g., PINE and PINT, BLACK and BLOOD, etc) matched for frequency, length and initial phoneme, in three frequency bands. Figure 5 also shows the two patients' performance on the same list approximately three (FM) or four (PB) years later. Although there is still a highly reliable advantage for regular over exception words for both patients at this stage, regular words now yield both errors and a degree of frequency sensitivity (with a slight reversal for FM between medium- and low-frequency regular words).

What is the nature of these emerging errors on regular words, and how are they to be explained? Over the course of the 3–4 year period, PB and FM were asked to name this list of words a total of five and eight times, respectively. Table 2 provides a classification of the entire set of errors made by each patient, on both regular and exception words, into two broad categories. Taking the less interesting category of "Other" errors first, both PB and FM—like virtually all patients with surface alexia (see for example Coltheart et al., 1983), and indeed like virtually all patients with any kind of reading disorder—make a certain number of errors where the response bears a relationship to the target word that is neither "surface" nor "deep"

---

[c] For readers interested in the underlying pathology of these conditions: three of the patients reported in Patterson and Hodges (1992) have come to post-mortem analysis. Two had Pick's disease. The third had Alzheimer pathology but in a highly atypical distribution: the profound focal left temporal atrophy and severe neuronal loss in this region was more characteristic of semantic dementia due to Pick's disease than of AD.

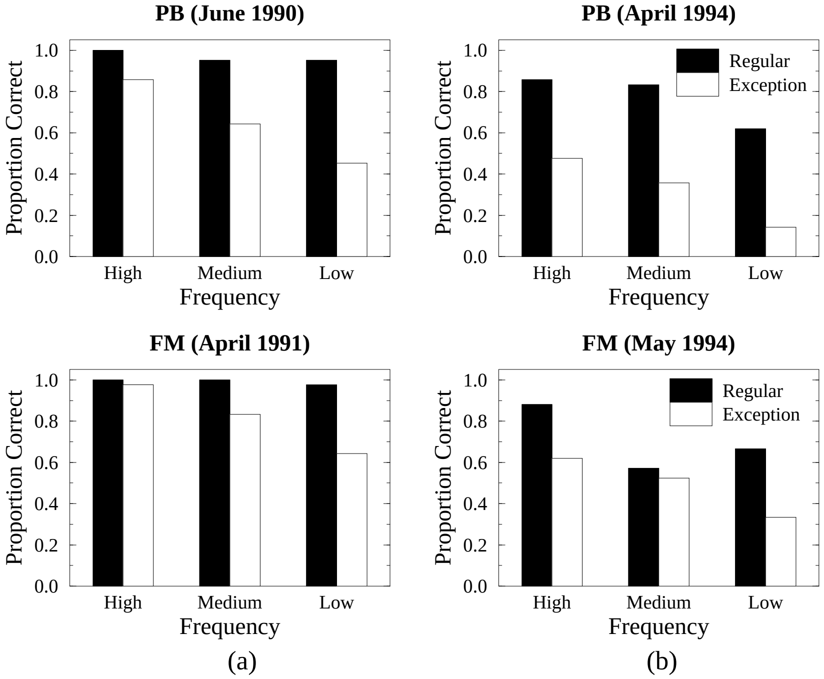

Figure 5: Reading performance by two progressive surface dyslexic patients, PB and FM, (a) at their initial assessment, on the Patterson and Hodges (1992) high-, medium- and low-frequency regular and exception words; and (b) about 4 (PB) and 3 (FM) years later.

Table 2: Error data from multiple administrations of the surface list to PB and FM, showing each patient's total numbers of errors and the classification of these into two broad classes: (1) LARC errors (short for Legitimate Alternative Reading of Components), e.g., PINT (an exception word) pronounced to rhyme with "mint" and HOOT (a regular word) pronounced to rhyme with "foot"; (2) Other errors (variously described by other authors as visual, orthographic, phonological), e.g., ONCE→"ounce."

Patient	Times Tested	Error Type	Word Type Regular	Exception
PB	5			
		Total	71/630	291/630
		LARC	34	246
		Other	37	45
FM	8			
		Total	129/1008	337/1008
		LARC	41	243
		Other	88	94

but simply resembles it orthographically and/or phonologically. FM in particular has become more prone to "Other" errors as her reading disorder has worsened; for example, she now makes a number of the letter confusion errors, mainly between visually similar letters (e.g., PRAY→"bray"), that are more characteristic of pure than of surface alexia. Most, though not all, of the "Other" errors are substitutions of an orthographically/phonologically similar real word for the target, as in ONCE→"ounce," THROAT→"trout," etc. Both PB and FM produced roughly equal numbers of "Other" errors to regular and exception words. This is as we would expect, provided that the regular and exception words are reasonably well matched in orthographic and phonological characteristics (such as similarity to other words in their neighborhoods), and differ only in the predictability of the relationship between spelling and sound.

The error category germane to the present discussion is what we have dubbed *LARC* errors (a term first used by Patterson, Suzuki, Wydell & Sasanuma, 1995), short for Legitimate Alternative Reading of Components. LARC means that the incorrect response reflects a legitimate pronunciation of each component of the word, in the sense that the orthographic component takes that pronunciation in other English words. The "Other" response THROAT→"trout" cannot be classed as a LARC error because there are no English words (at least in the dialect of English spoken by PB and FM), in which TH is pronounced /t/, and there are also no words in which OA is pronounced /au/ as in "trout." The quintessential LARC error is a regularization like PINT→/pInt/; but there are other types of such errors as well. First of all, exception words can and do yield LARC errors which are not pure regularizations. For example, the regularization of BLOOD would rhyme with "food," but PB pronounced it like "good"; the regularization of SWEAT would rhyme with "treat" but he pronounced it like "great"; and so on. Secondly, of even greater interest, regular words also yield LARC errors: PB named HOOT to rhyme with "foot," YEAST like "breast," HEAR like "bear"; FM pronounced BROWN with the body/rime of "blown," HEAT like "threat," COST like "post"; and so on. Not surprisingly, since the very definition of an exception word is that at least one of its components has a different, legitimate, and indeed more common, pronunciation, both PB and FM made far more LARC errors to exception words than to regular words. The important observation is that while errors to regular and exception words may differ in quantity, they do not differ in nature and therefore do not require a different kind of account.

Many of the regular words yielding LARC errors are, of course, the type of word known (since Glushko, 1979) as *regular inconsistent*. A regular inconsistent word (like HOOT) takes the pronunciation that is most typical of its body neighborhood; but one or more words with the same body have a conflicting pronunciation (e.g., FOOT). According to the framework and simulations in Plaut et al. (1996), such regular inconsistent words—rather akin to the W-words— represent an intermediate case between words with completely predictable components and true exceptions. The relatively infrequent but illuminating reading errors of surface dyslexic patients to these words strike us as another significant match between the triangle model and real data. As with the decreased vulnerability of W-words in surface dyslexia, the slightly increased vulnerability of regular inconsistent words is a direct prediction of the triangle framework. No doubt these data can also be given an explanation in a dual-route framework. In the DRC model, this would presumably involve interaction between the phoneme system (which is activated by the GPC rule system, and should support correct reading of a regular word like hoot) and the phonological output lexicon. If presentation of HOOT partially activates the lexical representations for orthographically similar words like FOOT and SOOT, the rule-based pronunciation of HOOT might occasionally succumb to this influence and be ir-regularized. As suggested by Sasanuma, Itoh, Patterson and Itoh (in press), to the extent that such interactive influences between lexical and non-lexical systems provide a major explanatory principle in the DRC model, the differences between the two approaches become less critical.

We conclude with an observation which, in the context of this book, may constitute preaching to the converted: that both components of the book's topic are proving important in the effort to understand the human brain and its capabilities. Not only must we build computational models whose predictions can be tested; the data against which the models are tested must include disordered as well as normal functioning.

Acknowledgments

The work reported here was supported in part by grants from the NIMH (USA) to J. L. McClelland and colleagues, and from the MRC (UK) to J. R. Hodges and colleagues. We are grateful to Naida Graham and Eamon Strain for research assistance.

References

Baluch, B., & Besner, D. (1991). Visual word recognition: Evidence for strategic control of lexical and nonlexical routines in oral reading. *Journal of Experimental Psychology: Learning, Memory and Cognition, 17*, 644–652.

Behrmann, M., & Bub, D. (1992). Surface dyslexia and dysgraphia: Dual routes, single lexicon. *Cognitive Neuropsychology, 9*, 209–251.

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

Breedin, S. D., Saffran, E. M., & Coslett, H. B. (1994). Reversal of the concreteness effect in a patient with semantic dementia. *Cognitive Neuropsychology, 11*, 617–660.

Bub, D., Black, S., Hampson, E and Kertesz, A. (1988). Semantic encoding of pictures and words: Some neuropsychological observations. *Cognitive Neuropsychology, 5*, 27–66.

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

Bullinaria, J. A. (in press). Connectionist models of reading: Incorporating semantics. In *Proceedings of the 18th Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.

Castles, A., & Coltheart, M. (1993). Varieties of developmental dyslexia. *Cognition, 47*, 149–180.

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

Coltheart, M. (1985). Cognitive neuropsychology and the study of reading. In M. I. Posner and O. S. M. Marin (Eds.), *Attention and performance XI* (pp. 3- 37). Hillsdale, NJ: Erlbaum.

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

Coltheart, M., Masterson, J., Byng, S., Prior, M., & Riddoch, J. (1983). Surface dyslexia. *Quarterly Journal of Experimental Psychology, 35A*, 469–495.

Coltheart, M., & Rastle, K. (1994). Serial processing in reading aloud: Evidence for dual-route models of reading. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 1197–1211.

Franklin, S., Howard, D., & Patterson, K. (1995). Abstract word anomia. *Cognitive Neuropsychology, 12*, 549–566.

Funnell, E. (1983). Phonological processing in reading: New evidence from acquired dyslexia. *British Journal of Psychology, 74*, 159–180.

Funnell, E. (in press). Response biases in oral reading: An account of the co- occurrence of surface dyslexia and semantic dementia. *Quarterly Journal of Experimental Psychology*.

Glushko, R. J. (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance, 5*, 674–691.

Graham, K. S., Hodges, J. R., & Patterson, K. (1994). The relationship between comprehension and oral reading in progressive fluent aphasia. *Neuropsychologia, 32*, 299–316.

Graham, K. S., Patterson, K., & Hodges, J. R. (1995). Progressive pure anomia: Insufficient activation of phonology by meaning. *Neurocase, 1*, 25–38.

Henderson, L. (1982). Orthography and Word Recognition in Reading. London: Academic Press.

Hillis, A. E., & Caramazza, A. (1991). Mechanisms for accessing lexical representations for output: Evidence from a category specific semantic deficit. Brain and Language, 40, 106–144.

Hodges, J. R., Graham, N., & Patterson, K. (1995). Charting the progression in semantic dementia: Implications for the organisation of semantic memory. *Memory, 3*, 463–495.

Hodges, J. R., & Patterson, K. (in press). Non-fluent progressive aphasia and semantic dementia: A comparative neuropsychological study. *Journal of the International Neuropsychological Society*.

Hodges, J. R., Patterson, K., Oxbury, S., & Funnell, E. (1992). Semantic dementia: Progressive fluent aphasia with temporal lobe atrophy. *Brain, 115*, 1783–1806.

Howard, D., & Frankln, S. (1988). *Missing the meaning?* Cambridge, Mass: MIT Press.

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

Kawamoto, A. H., & Zemblige, J. H. (1992). Pronunciation of homographs. *Journal of Memory and Language, 31*, 349–374.

Kay, J., & Marcel, A. J. (1981). One process, not two, in reading aloud: Lexical analogies do the work of nonlexical rules. *Quarterly Journal of Experimental Psy-*

chology, 33A, 397–414.

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

Lambon-Ralph, M., Ellis, A. W., & Franklin, S. (1996). Semantic loss without surface dyslexia. *Neurocase, 1*, 363–369.

Marshall, J. C., & Newcombe, F. (1973). Patterns of paralexia: A psycholinguistic approach. *Journal of Psycholinguistic Research, 2*, 175–199.

Newcombe, F., & Marshall, J. C. (1980). Transcoding and lexical stabilization in deep dyslexia. In M. Coltheart, K. Patterson & J. C. Marshall (Eds.), *Deep dyslexia* (pp. 176–188). London: Routledge.

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

Paap, K. R., & Noel, R. W. (1991). Dual route models of print to sound: still a good horse race. *Psycholinguistic Research, 53*, 13–24.

Parkin, A. J. (1993). Progressive aphasia without dementia due to focal left temporo-frontal hypometabolism—A clinical and cognitive analysis. *Brain and Language, 44*, 201–220.

Patterson, K., & Behrmann, M. (1996). Frequency and consistency effects in a pure surface dyslexic patient. Submitted to *Journal of Experimental Psychology: Human Perception and Performance*.

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

Patterson, K., Seidenberg, M. S., & McClelland, J. L. (1989). Connections and disconnections: Acquired dyslexia in a computational model of reading. In R. G. M. Morris (Ed.), *Parallel distributed processing: Implications for psychology and neuroscience* (pp. 131–181). Oxford: OUP.

Patterson, K., Suzuki, T., Wydell, T., & Sasanuma, S. (1995). Progressive aphasia and surface alexia in Japanese. Neurocase, 1, 155–165.

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

Raymer, A. M., & Berndt, R. S. (1994). Models of word reading: Evidence from Alzheimer's disease. *Brain and Language, 47*, 479–482.

Reggia, J. A., Berndt, R. S., & D'Autrechy, C. L. (1994). Connectionist models in neuropsychology. In F. Boller & J. Grafman (Eds.), *Handbook of Neuropsychology, Vol. 9* (pp 297–33). North Holland, Elsevier.

Sasanuma, S., Itoh, H., Patterson, K., & Itoh, T. (in press). Phonological alexia in Japanese: A case study. *Cognitive Neuropsychology*.

Seidenberg, M. S. (1992). Connectionism without tears. In S. Davis (Ed.), *Connectionism: Advances in theory and practice* (pp 84–137). Oxford: OUP.

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

Seidenberg, M. S., Waters, G. S., Barnes, M. A., & 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., & Rosenberg, C. R. (1987). Parallel networks that learn to pronounce English text. *Complex Systems, 1*, 145–168.

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

Shallice, T., Warrington, E. K., & McCarthy, R.(1983). Reading without semantics. *Quarterly Journal of Experimental Psychology, 35A*, 111–138.

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

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

Van Orden, G. C., & Goldinger, S. D. (1994). Interdependence of form and function in cognitive systems explains perception of printed words. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 1269–1291.

Watt, S., Jokel, R., & Behrmann, M. (in press). Surface dyslexia in non-fluent progressive aphasia. *Brain and Language*.