Connectionist modelling of surface dyslexia based on foveal splitting: Impaired pronunciation after only two half *pint*s

Richard Shillcock*†
Padraic Monaghan*
`rcs/pmon@cogsci.ed.ac.uk`
* Institute for Adaptive and Neural Computation, Division of Informatics
† Department of Psychology
University of Edinburgh, 2 Buccleuch Place, Edinburgh, EH8 9LW, UK

Abstract

In cases of surface dyslexia and phonological dyslexia there is a dissociation between the reading of irregular words and nonwords. This dissociation has been captured in connectionist models of dyslexia in terms of impairments to the models' phonological representations. We report a series of connectionist simulations based on an alternative neuro-anatomically motivated theory that dyslexia is at least partly caused by hemispheric desynchronisation. Problems of interhemispheric transfer affect the mapping between orthography and phonology because the human fovea is precisely vertically split: a fixated word is initially split and the two parts contralaterally projected to the two hemispheres of the brain. Much of lexical processing can be reconstrued as the integration of this information (Shillcock, Ellison & Monaghan, 2000). We demonstrate that the dissociation between the reading of irregular words and nonwords can be understood in terms of a failure to integrate the initially split input.

Introduction

The word *pint* is pronounced differently from its orthographic neighbours *dint, hint, lint, mint, tint.* In this sense its pronunciation is irregular. Because of the presence of such words in its lexicon, English is said to have a deep orthography. Irregular pronunciations have been observed to cause processing problems both for normal readers and for certain impaired readers: in the former a low frequency irregular word such as *pint* may take measurably longer to pronounce than a matched regular word, and in the latter such a word might be erroneously regularised (*pint* pronounced to rhyme with *mint*). For these reasons irregular words have been used in a large number of studies of how readers translate the orthographic representation of a word into its phonological specification. In this paper we approach the pronunciation of irregular words from the perspective of a new model of visual word recognition based on the neuro-anatomical observation of foveal splitting (Shillcock, Ellison & Monaghan, 2000). We will argue that this model provides a principled motivation for some of the problems associated with irregular pronunciations in both normal and impaired readers.

The measurable processing problems that normal adult readers have with low frequency irregular words (*pint*) do not extend to high frequency irregular words such as *have* (whose pronunciation cannot be predicted from *cave, rave, pave, knave*) (Paap & Noel, 1991; Seidenberg, Waters, Barnes & Tanenhaus, 1984; Taraban & McClelland, 1987). Perhaps the most productive approach to modelling this interaction between regularity and frequency has been that developed by Seidenberg and McClelland in their (1989) connectionist model of reading and by Plaut and others subsequently (e.g. Plaut & McClelland, 1993; Plaut, McClelland, Seidenberg & Patterson, 1996; Harm & Seidenberg, 1999). Seidenberg and McClelland's original model was a simple feedforward model; later models have contained recurrent connections and more structured input and output layers. In general, in these models the apparently rule-based behaviour that generates regular pronunciations is recast in statistical terms, so that a low frequency irregular pronunciation (*pint*) is militated against by the more frequently occurring regular pronunciation of the relevant vowel in similar contexts. Thus, when the irregular pronunciation is itself high frequency (*have*), that particular mapping between orthography and phonology is sufficiently emphasised to co-exist with the regular mapping. The superpositional storage that characterises these models means that a minority mapping such as that required by *pint* is disproportionately demanding in terms of computational resources compared with the efficient generalisation represented by a regular pronunciation. This fact is also demonstrated in Plaut and McClelland's (1993) model in which attractors behave componentially when the input and output representations of words are split into onset, nucleus and coda. In their model, pronunciation of the vowel in *pint* involves the connections from all three input slots (onset, nucleus and coda), whereas the regular, default pronunciation of the *i* involves only the nucleus.

Irregular words pose particular problems for surface dyslexics, who are liable to produce regularisation errors (Patterson, Coltheart & Marshall, 1985; Manis et al., 1996). In connectionist models, the learning and retention of irregular pronunciations are generally vulnerable. For instance, Seidenberg and McClelland showed that restricting the number of hidden units impairs learning low frequency irregular words, and Harm and Seidenberg (1999) produced a similar effect by lowering the learning rate overall. Plaut et al. (1996) and others explore the idea that a division of labour occurs between the direct orthography-phonology mapping and the same mapping mediated by semantics: the irregular pronunciations are seen as relying more on the route that proceeds via semantics, leaving the direct route to concentrate on the regular pronunciations. This behaviour of the models seems to capture the observation that developmental dyslexics frequently have impaired phonological processing (see Snowling, 2000, for a comprehensive review of the phonological processing problems of dyslexics).

Some of the most critical data in this area concerns the dissociation between the ability to pronounce irregular words and the ability to pronounce nonwords. This dissociation is found between the surface and phonological subtypes of dyslexia (Beauvois & Derouesné, 1979). Surface dyslexics cope moderately well with novel words and nonwords, but are liable to make regularisation errors on known irregular words, whereas phonological dyslexics cope moderately well with irregular words but are disproportionately impaired when reading novel words and nonwords. This dissociation motivated Marshall and Newcombe's (1973) original dual-route model and its later development and computational implementation by Coltheart and others (e.g. Coltheart, Curtis, Atkins & Haller, 1993), in which a lexical route and a non-lexical route (containing grapheme-to-phoneme correspondence rules) can be differentially impaired to produce the desired impairments.

In this paper we claim that developmental surface dyslexia, characterised by problems with irregular words, arises naturally from impaired hemispheric interaction in a model based on the observation that the human fovea is precisely vertically split. There are long-standing observations that dyslexia is frequently associated with problems of callosal transfer (e.g. Davidson & Saron, 1992). By modelling word recognition within a split network, we ground these observations of impaired reading in an implemented model of normal reading. We claim that impaired hemispheric interaction is a fundamental, qualitative explanation of problems in pronouncing irregular words, and is a more parsimonious account than resource-based explanations.

The split-fovea model of reading

Shillcock, Ellison and Monaghan (2000) present a model of lexical access based on the precise vertical splitting of the human fovea. Information presented in the left visual field (LVF) projects, initially, to the right hemisphere (RH), whereas information in the RVF projects to the LH. This long-recognised initial contralateral projection of the visual field to the two hemispheres of the brain is also true of the human fovea – a fact that has not been extensively explored in research in visual word recognition. When a word is directly fixated, the two parts of the word on either side of the fixation point are projected to different hemispheres. In order for a word to be recognised and pronounced correctly, the information in the two hemispheres has to be integrated. Shillcock et al. (2000) investigate some of the implications of foveal splitting for a full-sized lexicon and show that the initial splitting of the word is an informationally attractive starting point for word recognition, rather than being merely an inconvenience.

Consider the word *pint*, centrally fixated. The two sides of a split model will receive the two letters *pi* and *nt*, respectively. In order to pronounce the vowel correctly, the model must process information from both sides of the model: *pi* on its own may be pronounced as in *pine*, or as in *pill*. If this integration is not complete then a regularisation error is likely to occur, so that *pi* will be pronounced in its most frequent form: /pI/ (see Table 1). As Harm and Seidenberg (1999) observe, the task of reading irregular words is akin to solving the XOR problem. In the case of a split model without recurrence, this task is impossible, as the structure is akin to a perceptron.

Table 1: Pronunciation of *pi*.

Pronunciation of vowel	Example	Count	Frequency (per million)
/I/	pith	14	424
/ii/	piece	2	175
/&I/	pint	10	164
/I@/	pier	2	16

When a single word is read it may be fixated to the left of the first letter, to the right of the last letter, or at all points in between. Elsewhere (e.g. Shillcock & Monaghan, *2001*) we have implemented an idealised version of the initial splitting of these different visual inputs in reading single words in a series of neural network models, so that there are five different fixation positions across the input for a four letter word (only "fixations" between letters are considered). In the simulations we report here, we employ a simplified version of this model, in which each word is only fixated at one fixation point within the word. This simplification allowed us to stay closer to Harm and Seidenberg's (1999) simulations, which provide an important point of comparison. We show that impairments to the integration of information in the two parts of the word results in the behaviour associated with surface dyslexia.

One version of the split model of reading is shown in Figure 1. The model comprises two orthographic input layers, corresponding to the left and right visual fields. Each input layer has 4 letter slots. If a letter is present in a slot then one of 26 units representing the letters of the alphabet will be active. The output layer is a representation of phonology, with two slots each for onset, nucleus and coda. Phonemes are represented in terms of 11 phonological features. We have used the features described by Harm and Seidenberg (1999), although we have augmented their phonology to accommodate the transcriptions found in the CELEX database (Baayen, Pipenbrock & Gulikers, 1996). These changes to the phonological transcription principally involved the representation of diphthongs and the role of schwa; the changes consid-

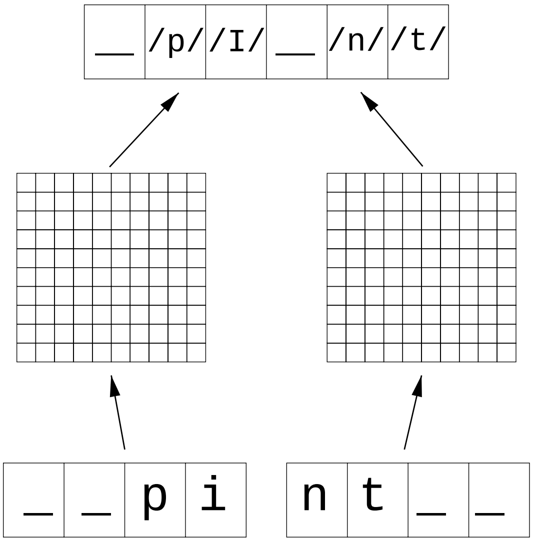

Figure 1: The split model of reading.

erably increased the problems of learning the mapping from orthography to phonology. Each input layer in the model is fully connected to one of two hidden layers each with 100 units. These hidden layers are fully connected to the output. The model has to learn to map orthography onto phonology given the constraints of a split input.

The model was trained on 3835 single-syllable word lemmas from the CELEX English database with up to two consonants in the onset and coda. Four words were omitted because they contained orthographic consonant clusters that would not fit within the input[1]. Words were presented in the input so that the first vowel was justified immediately to the left of centre. In total, 57.5% of words were fixated at or to the left of their physical centre, which is the usual site of fixation during reading (Rayner, 1998). Words were presented randomly to the model according to the log-frequency of the word, and backpropagation was used to train the connections between the layers.

Several versions of the model were trained: (a) a feedforward version with no phonological attractors (“Split-No Attractors”), (b) a recurrent version with phonological attractor units (“Split-Attractor”), and (c) a recurrent and interconnected version with phonological attractors, with the hidden layers fully connected to one another (“Split-Interconnected”). The attractor models had a set of 35 “clean-up” units connected to the output layer, which were pretrained in learning a phonology to phonology mapping. Orthography to phonology trials were then interdispersed with phonology to phonology trials during training of the attractor models. Two nonsplit models were also trained as controls, equivalent to those employed by Harm and Seidenberg (1999). These models were required to make the same mapping between orthography and phonology, except that all the input units were connected to a single hidden layer of 100 units. Feedforward (“Nonsplit-No Attractor”) and recurrent attractor (“Nonsplit-Attractor”) versions were trained.

The 361 nonwords used by Harm and Seidenberg (1999) were employed[2]. The model was tested with 44 irregular words taken from Taraban and McClelland’s (1987) materials[3].

We predicted that the Nonsplit-Attractor model would perform well on both irregular words and nonwords, following Harm and Seidenberg’s demonstration of the capabilities of that model. The Nonsplit-No Attractor version should perform relatively poorly on irregular words and nonwords; this model corresponded to Harm and Seidenberg’s (1999) unimpaired model, and their phonologically impaired model, respectively). We predict that the split models will exhibit surface dyslexia to varying degrees according to the level of interaction between the two hemispheres. Furthermore, this deficit will be robust in the face of further training on the model – additional training will not reverse the pattern of difficulties in reading irregular words and nonwords, as happens in Harm and Seidenberg’s (1999) delayed model of reading.

Results

Figures 2, 3 and 4 show the performance of the different models, in terms of percentage of words correctly pronounced, at different stages of training. Figure 2 shows how well the models learned the whole training set, and we see that the NonSplit-No-Attractors model performs comparably to the same architecture presented by Harm and Seidenberg, climbing steadily to levels in excess of 90% correct. Even though the current training set contained more elaborate phonological representations than those used by Harm and Seidenberg for the same type of model, we see that the curve has not asymptoted even after the presentation of 5M words in training; further training promises to improve performance even more. We see a similar level of success for the NonSplit-Attractors model, though this model is slower to learn than the NonSplit-No-Attractors model due to the interleaving of phonology to phonology trials during training. These models replicate Harm and Seidenberg’s (1999) success with the same architectures.

In contrast, Figure 2 shows that the Split-No-Attractors model asymptotes early, after about 3M words of training, and never exceeds 70% of words correct. The simple split model is fundamentally incapable of more than this modest performance. The remaining learning curves in Figure 2 demonstrate the value of connectivity between the two halves of the split model: the Split-Attractors model behaves on a par with the different

---

[1] The omitted words were *eighth*, *borscht*, *touched*, and *schnapps*.

[2] Three nonwords were omitted because they appeared as real words in our training corpus: *plop*, *mo*, and *peep*.

[3] 4 of the original 48 items were omitted because they were wordforms that did not occur in our word lemma training set: *does*, *said*, *says*, and *were*.

Non-Split models, and the most rapid learning of all occurs in the Split-Corpus-Callosum model, although the latter model is not directly comparable with the others as it contains more weighted connections and hence more computational resources. The principal result from the training of the different models is that sharing information between the two halves of the input is critical to successful learning.

Figure 3 shows relatively successful generalisation by all of the models to the set of nonwords. The Split-No-Attractors model performs least well as we might predict from Figure 1, but all of the models asymptote within the 55%-70% region, in generalising to pronounce nonwords that were not encountered in training.

The central result of all of these simulations can be seen by comparing Figures 3 and 4. In Figure 4 we see differences between the models in their performance on pronouncing irregular words. The Split-No-Attractors model performs extremely poorly on these words, pronouncing only about 50% of the irregular words correctly. The Split-Attractors and NonSplit-Attractors models perform comparably well, and the NonSplit-No-Attractors and Split-Corpus-Callosum models perform extremely well on these irregular words. In relation to the performance of the other architectures, the simple split architecture shows a dramatic dissociation between

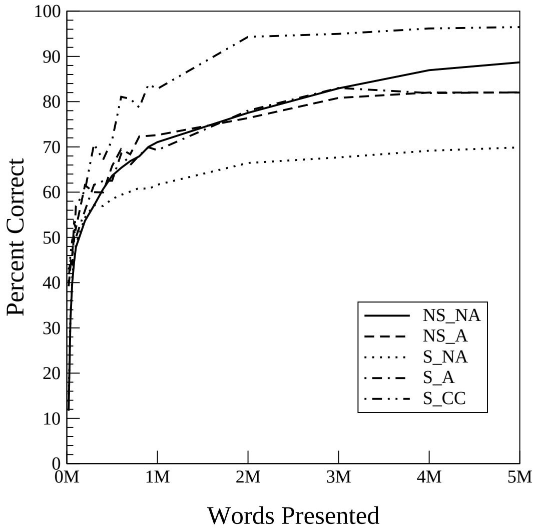

Figure 2: Performance of the split and nonsplit models on the training set.

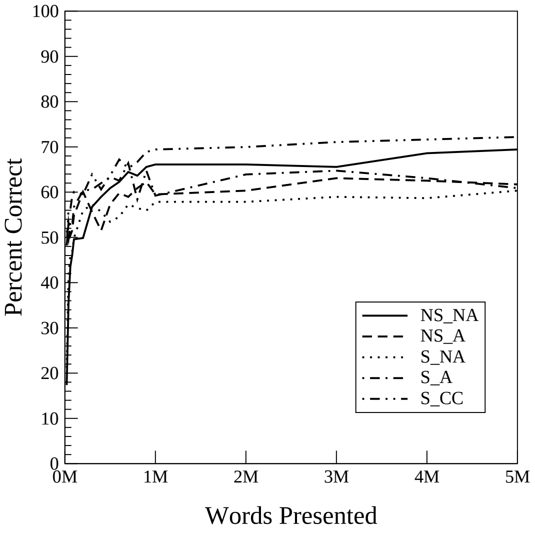

Figure 3: Performance of the split and nonsplit models on nonwords.

The errors produced by the simple split model resembles those found in surface dyslexia. Table 2 compares the performance of the Split-No Attractors and NonSplit-No-Attractors models on an example set of irregular words after 5 million words had been presented to the models. The NonSplit model converges to correct pronunciations for all these words, whereas the Split model produces regularisation errors even after such lengthy training. The differences between the Split and NonSplit models are not only quantitative but qualitative, with the Split model producing plausible over-regularisations for irregular words.

Table 2: Productions of the Split and NonSplit No-Attractor models for irregular word examples.

	Pronunciation		
Word	Correct	NonSplit model	Split model
bind	b&Ind	b&Ind	b&nd
broad	brOOd	brOOd	brOOd
come	kVm	kVm	kVm
hood	hUd	hUd	huud
mild	m&Ild	m&Ild	mEIld
pear	pE@r	pE@r	pIEr
pint	p&Int	p&Int	pInt
quay	kii	kii	kEI
tomb	tuum	tuum	tOm

Discussion

We have successfully reproduced the dissociation between the reading of irregular words and nonwords observed in surface dyslexia. We started by observing that the human fovea is precisely vertically split, and that a fixated word is initially divided between the two hemispheres. We reconstrued the task of lexical processing as one of integrating the information contained in the two hemispheres. We explored the performance of neural network models of reading which had been similarly vertically split, and demonstrated that simple split archi-

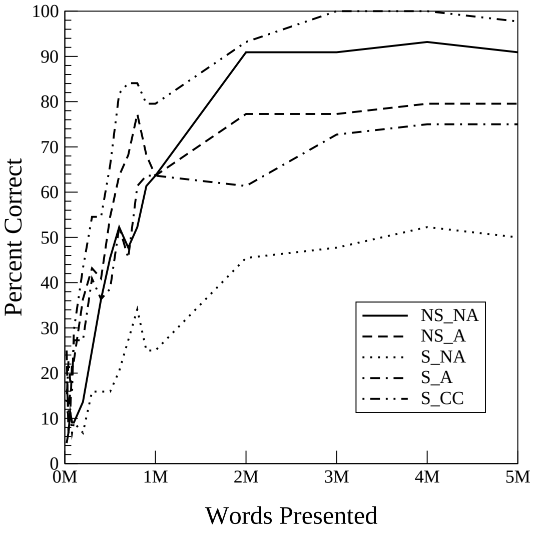

Figure 4: Performance of the split and nonsplit models on irregular words.

tectures are crucially limited in their success in mapping from orthography to phonology.

The problem of pronouncing a divided irregular word like *pint* resembles the XOR problem. The two halves cannot map directly and independently onto the output to solve the problem. There is no purely componential solution, as there is with the pronunciation of a regular word like *mint*. Instead, there have to be intermediate representations that combine the two halves and then map onto the eventual solution. The problem is not restricted to the pronunciation of vowels, although they are the main part of the problem. The consonant cluster at the beginning of *chef* is liable to be regularised to resemble that in *chad* and *cheat* if the whole of the word is not present.

A paradox of the connectionist modelling of cognitive neuropsychological data is the emergence of complex dissociations from relatively undifferentiated architectures. Such demonstrations are parsimonious accounts of the data because they seem to emerge from the structure of the problem, ”out there” in the world rather than from the details of some putative functional processing architecture. Researchers have produced successful connectionist accounts of the pattern of dissociations found in the different dyslexias. These dissociations have been achieved in a variety of ways, and we can distinguish three broad strategies. The first is the *Representational Strategy*. Thus, in Plaut and Shallice’s (1993) model of deep dyslexia the authors show that the proportion of the different types of error produced by the lesioned network can be changed by lesioning in different parts of the model; lesioning around the model’s semantic clean-up units causes more semantic errors, for instance. Similarly, Harm and Seidenberg (1999) show that impairing the phonological attractors in their model increased the errors on irregular words. The psychological reality of this strategy is relatively easy to assess. For instance, Harm and Seidenberg’s impairing of the phonological attractors is motivated by the data showing that surface dyslexia is often accompanied by phonological impairment (see, e.g., Snowling, 2000). The second strategy we identify is the *Parametric Strategy*. Examples are the manipulation of the computational resources available to the model, or changes in the details of training. Seidenberg and McClelland (1989) reduced the numbers of hidden units available to the orthography-to-phonology mapping, showing that it militated against learning the irregular words. Similarly Harm and Seidenberg (1999) produced the same outcome by reducing the learning rate. Such manipulations are quantitative, compared to the qualitative effects of the Representational Strategy, and their psychological reality is more difficult to assess. There are clear demonstrations that the orthography-to-phonology mapping for irregular words is the hardest aspect of the pronunciation problem, but equally there are demonstrations that the problem can be solved by a network model with only a few hundred nodes, given the correct representations. Are irregular words still hard for a processor with the resources of the human brain, and with several years to spend on the problem? Additional training of Harm and Seidenberg’s model with a low learning rate producing surface dyslexia behaviour leads to the convergence of performance on regular and irregular words, for example. Parametrically based models of dissociations in dyslexia carry with them costly assumptions concerning the capacities of the human brain.

Our own approach has been to introduce a discrete, qualitative neuro-anatomical distinction into the modelling: the effects of foveal splitting. We characterise surface dyslexia as being caused, at least in part, by hemispheric desynchronisation. No amount of extra training and no amount of extra computational resources devoted to either half of a split processor can improve performance on irregular words in a direct mapping between orthography and phonology. The problem is a qualitative one.

Thus, we have provided an account of the dissociation between irregular words and nonwords in surface dyslexics. We do not see this hemispheric desynchronisation account as a necessarily exclusive one. There may also be a contribution from the impairment of phonological representations, as is generally assumed. The relative contribution of each account, and any possible interaction between them, is an empirical question. However, it may be possible to ground phonological impairment itself in hemispheric desynchronisation, despite the fact that expressive phonology is conventionally viewed as the sole preserve of the LH.

A further aspect of the dissociation between regular words and nonwords concerns the performance of phonological dyslexics, who can be highly proficient readers of known words, but can also be dramatically poor at reading nonwords and unknown words, even to the extent of not being able to generate the sound of an isolated letter. We interpret phonological dyslexia as also resulting from hemispheric desynchronisation. We propose that the desynchronisation is more severe in phonological dyslexia, compared with surface dyslexia, and that a different relationship emerges between the orthographic, phonological and semantic forms of words to compensate for the inability to integrate the orthographic information in the two hemispheres and to map it directly onto a phonological representation. We propose that orthographic information is mapped directly onto semantic information independently in the two hemispheres; each hemisphere partially activates the semantic representations of all the words corresponding to its own orthographic input. Identification of the word is achieved by the interhemispheric transfer of semantic information. The routes by which semantic information might be transferred interhemispherically are more extensive, compared with those concerned with vision, and the problem of finding the intersection of two sets of partially activated semantic representations is, we claim, an easier problem than integrating the corresponding visual information. In this account, a novel word can only be pronounced ad hoc by analogy with known words. This account is therefore a qualitative explanation of the often dramatic problems with novel words observed in phonological dyslexia.

Acknowledgments

This work was supported by Wellcome Trust grant 059080.

References

Baayen, R.H., Pipenbrock, R. & Gulikers, L. (1995). *The CELEX Lexical Database (CD-ROM)*. Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA.

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

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.

Davidson, R.J & Saron, C.D. (1992). Evoked potential measures of interhemispheric transfer time in reading disabled and normal boys. *Developmental Neuropsychology 8,* 261-277.

Harm, M. W. & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: insights from connectionist modelling. *Psychological Review*, *106*, 491–528.

Manis, F., Seidenberg, M., Doi, L., McBride-Chang, C. & Peterson, A. (1996). On the basis of two subtypes of developmental dyslexia. *Cognition*, *58*, 157–195.

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

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

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

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

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

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

Rayner, K. (1998) Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*, 373–422.

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.

Shillcock, R. Ellison, M. T. & Monaghan, P. (2000). Eye-fixation behaviour, lexical storage and visual word recognition in a split processing model. *Psychological Review*, *107*, 824–851.

Shillcock, R. & Monaghan, P. (1998). Using physiological information to enrich the connectionist modelling of normal and impaired visual word recognition. *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society* (pp. 945–950). Hillsdale, NJ: Lawrence Erlbaum Associates.

Shillcock, R. & Monaghan, P. (2001). The computational exploration of visual word recognition in a split model. *Neural Computation*, *13*.

Snowling, M. (2000). *Dyslexia: A Cognitive Developmental Perspective* (Second Edition). Oxford: Blackwell.

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