In: *Proceedings of the GALA '97 Conference on Language Acquisition*, pages 316-321.

Modelling the Acquisition of Reading Skills

John A. Bullinaria

Centre for Speech and Language, Department of Psychology
Birkbeck College, Malet Street, London WC1E 7HX, U.K.
j.bullinaria@psyc.bbk.ac.uk

Abstract

Existing direct route connectionist models of reading aloud (i.e. text to phoneme conversion) can now learn to perform on their training data and unseen words with accuracy comparable to that of humans. They also exhibit a number of developmental, reaction time and brain damage effects that are observed experimentally. However, various deficiencies (such as their failure to perform reliable lexical decision and to show the pseudohomophone effect) indicate the importance of incorporating some form of lexical/semantic system into these models. In this paper we present a preliminary investigation of this problem. A general framework is outlined that models network activation flow between orthography, phonology and semantics. Explicit simulations of the learning process in small scale networks of this type show how this approach can account for many aspects of reading (and related tasks) that are not possible without the influence of the semantic system.

1. Introduction

Recent improved versions (Bullinaria, 1994, 1997; Plaut et al., 1996) of earlier connectionist models of reading (Sejnowski & Rosenberg, 1987; Seidenberg & McClelland, 1989) now perform on their training data and generalize to new words nearly as well as human subjects. These models also exhibit various realistic developmental and brain damage effects, and simulated reaction times (i.e. naming latencies) show many of the regularity, frequency and consistency effects found experimentally. Of particular relevance here is that they show realistic normal developmental effects (such as the learning of regular pronunciations more quickly than that of frequency matched irregular items) and under appropriate circumstances can exhibit developmental surface dyslexia (in which irregular word performance is poor compared with that of regular words). Moreover, these properties are largely independent of the modelling details. These models also constitute an existence proof against earlier claims that separate processing routes are needed to learn the regular and exception words. However these single route models, that map directly from orthography to phonology, are unable to account for several important pieces of experimental data. In particular, they are unable to perform reliable lexical decision and they do not show the pseudohomophone effect whereby non-words that sound like real words are pronounced faster than matched non-words that do not (McCann & Besner, 1987). It has always been clear that some form of lexical/semantic system would eventually have to be incorporated into any complete model of reading (e.g. Seidenberg & McClelland, 1989). However, it has not been so clear what form this additional sub-system should take and how it should interact with the direct route which already accounts for so many aspects of human reading abilities.

Here I shall present a preliminary investigation of the properties of class of these more complete fully connectionist models and their implications for understanding the learning to read by children. Not surprisingly, these models have many features in common with existing non-connectionist models of the same data (e.g. Coltheart et al., 1993; Coltheart & Rastle 1994). If we believe that the human brain is a neural network that allows activation to flow between separate orthographic (O), phonological (P) and semantic (S) representations, we naturally have a framework of the form shown in Figure 1. We initially assume, for simplicity, that the input representations are the same as the output representations and that there is one hidden layer of processing units on each link to allow non-linearly separable associations. These assumptions can always be relaxed later if necessary. The sub-systems consisting of the O to P mapping and the P to O mapping correspond to existing direct route reading and spelling models. It is our aim here to investigate how the links to and via S affect the networks' performance. The first problem we face is that (ignoring morphological effects) the mapping between O/P and S is essentially random, and learning random mappings is extremely costly in terms of computational resources. However, if we choose carefully simplified representations and learning algorithms, it is now possible for some preliminary small scale simulations to be carried out. This paper presents a number of such simulations within this framework, which indicate that it is (or soon will be) able to provide successful models of many aspects of human reading abilities.

2. A Simplified Model

To allow reasonable network training times (i.e. months rather than years) it is necessary to restrict ourselves to a fairly small subset (of about 500) of all monosyllabic words (compared with the 3000 or more words typically used in single route models). Then, to ensure that each possible grapheme and phoneme is sufficiently well represented in this training set that realistic generalization can occur, and to ensure that the density of words within the space

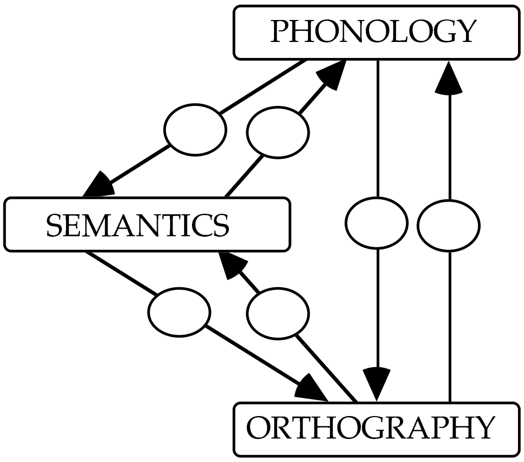

Figure 1: The basic connectionist framework with activation flow between each pair of input/ output representations via hidden units.

of orthography and phonology is not unrealistically sparse, it is also necessary to reduce the set of allowable graphemes and phonemes. To this end, our phonology representation consisted of an input/output unit for each of the most common onset phoneme clusters (12), vowel clusters (10) and offset phoneme clusters (12), and the orthography consisted of units for the corresponding three sets of letter clusters (12, 9 and 13) plus two units to code for the presence or absence of a final 'e'. This constitutes a simplified version of the Plaut et al. (1996) representation, that (after removing homographs and homophones) allows the representation of 513 real monosyllabic words from the standard Seidenberg & McClelland (1989) corpus. Rather than attempting to generate realistic semantic representations for these words, we simply assigned each a random sparse binary vector. Each activated unit (3 out of 27) is taken to represent the presence of a particular semantic micro-feature (e.g. as in Plaut & Shallice, 1993). The 'lexical entries' are implicit in the distributed patterns of activations of these units. Choosing to have 500 hidden units in each fully connected link completes the specification of the network in Figure 1.

We can regard this network as simultaneously providing simplified models of six basic language processing tasks: orthography to phonology (reading aloud), orthography to semantics (reading quietly), phonology to orthography (spelling), phonology to semantics (listening), semantics to orthography (writing) and semantics to phonology (speaking). We assume that taking our orthographic units to represent graphemes (rather than, say, individual letters or their constituent line segments) and the phonological units to represent phoneme clusters (rather than, say, individual phonemes or features such as front, back, nasal, fricative, etc.) will not make much difference to the general performance, though this will clearly have to be explored more carefully in the future. We have (again for simplicity and training time considerations) not included any explicit recurrent connections between or within layers, though the architecture does allow for recurrence such as activation flow from O to P to S and back to P. It is such interaction between the various sub-systems that we can expect to account for data such as the pseudohomophone effect.

We next come to the problem of training our network. Unfortunately it is not simply a matter of training each of the six sub-networks separately and then connecting them together. It is the interaction between the various possible pathways and the 'division of labour' between them during learning that gives rise to many of the interesting effects. Thus, we need to train the whole network at once. We also want to model various time course effects in a reasonably realistic manner (to give reaction times, etc.), so we think in terms of activation cascading through the network (e.g. McClelland, 1979) as in recurrent networks rather than the typical one pass approach of standard feed-forward networks. At each discrete time slice t we have:

$$Out_i(t) = Sigmoid(Sum_i(t))$$

$$Sum_i(t) = Sum_i(t-1) + \lambda \sum_j w_{ij} Prev_j(t) - \lambda Sum_i(t-1)$$

so the output $Out_i(t)$ of each unit i is the usual sigmoid of the sum of the inputs into that unit at that time. The sum of inputs $Sum_i(t)$ is given by the existing sum at time t–1 plus the additional weight w_{ij} dependent contribution fed through from the activation $Prev_j(t)$ of the previous layer and a natural exponential decay of activation depending on some time constant λ. The network can be trained using the traditional gradient descent procedure to modify the weights w_{ij} iteratively to produce the correct outputs. We clamp one input set (e.g. orthography), then use the above equations to update the activations to one of the output sets (e.g. semantics) and modify the weights to reduce the output errors, then do the same for the other output set (i.e. phonology). For each training word we repeat this process for a number of time slices and the network eventually settles into a stable output pattern. If we choose the training words in random order, and choose which input set to clamp at random for each word, and keep the time scale parameter λ and learning rate η sufficiently small that large fluctuations in the weights and activations do not occur, then the network eventually learns to produce the correct outputs for each input word form. Here we shall discuss one particular network trained with $\lambda = 0.1$, $\eta = 0.0001$, 150 time slices per word, logarithmically compressed word frequencies, cross-entropy error measure and no momentum.

3. General Performance

We noted above that it was appropriate to train the whole network at once, yet it is clear that children generally have some command of the mappings between phonology and semantics before they begin to learn to read and write. Consequently, we initially trained the network for 9200 epochs on just the phonology and semantics, by which time it was able

PROCESS	ENGLISH		SCRAMBLED	
	ACTIVATION ERROR	CORRECT OUTPUTS	ACTIVATION ERROR	CORRECT OUTPUTS
O to P (both routes)	5.1	98.1%	11.4	96.7%
O to P (direct)	10.8	97.5%	18.6	88.9%
O to P (via S)	37.4	0.0%	31.0	37.8%
O to S (both routes)	13.3	89.1%	10.5	93.6%
O to S (direct)	37.4	0.2%	21.8	73.1%
O to S (via P)	18.2	87.3%	27.4	45.0%

Table 1: The divisions of labour for the reading processes 'O to P' and 'O to S'.

to produce the correct phonology for 97.3% of the semantic inputs and the correct semantics for 94.3% of the phonology inputs. Not surprisingly it produced no correct outputs for orthographic inputs and no correct orthographic outputs. It was then trained for a further 4600 epochs on the full set of mappings. The mappings between orthography (O) and phonology (P) are very regular and hence much easier to learn than the mappings involving semantics (S). The network thus learns quickly that it can use the existing P-S pathways together with the easy O-P pathways to do everything, at the expense of not developing the O-S pathways. This is clearly seen in Table 1 which shows the error scores corresponding to the separate pathways for reading. Although the O-S pathways perform poorly on their own, they do contribute to the good total performance. Additional, more indirect, routes (e.g. O to P to S to P) bring the errors down even further (e.g. the final O to P performance is 99.6% with errors of regularization on only two low frequency irregular words).

We next have to consider the extent to which this division of labour is inevitable, so let us examine the extreme cases. First, suppose that we trained the full network from scratch rather than achieving good P-S performance before attempting to learn O. In this case the O-S and P-S pathways will be more equally developed, but the regularity of the O-P mappings is likely to still leave the direct O-P pathway more influential than the indirect O-S-P route. However, we now expect the direct O-S pathway to have more influence on the full O-S mapping.

Second, if the O-P mapping were much less regular (e.g. as in Chinese), we would expect the indirect O-S-P route to be much more influential on the full O-P mapping, and the direct O-S route to have much more influence on the full O-S mapping. This is also what we might expect for the more irregular words in English, particularly if they have 'strong' semantic representations (e.g. Strain, Patterson & Seidenberg, 1995). This was confirmed by repeating the above network simulation, but training with the orthography to phonology correspondences scrambled. The lack of regularities made the second stage of learning much slower but the network eventually (after 16000 epochs) reached a total performance similar to before in the predicted manner as summarised in Table 1.

Since there *is* considerable experimental evidence of phonological mediation between O and S in English (e.g. Van Orden et al., 1990) of the form found in our original model, we shall restrict our attention to that. We noted above that this network learnt all but two low frequency exception words and our experience with larger single route models (e.g. Bullinaria, 1997; Plaut et al., 1996) leaves us confident that these will also be learnt if we allow sufficient further training. Generalization was tested on two distinct sets of 200 non-words. Our restricted phonology and orthography made it rather difficult to construct pseudohomophones (i.e. non-words that sound like real words) and so we used all 200 that were possible and a matched set of 200 control non-words (e.g. word 'came', pseudohomophone 'caim', control non-word 'com'). Table 2 summarises the generalization performance, which is surprisingly good (thanks to our reduced phonology and orthography) and exhibits the pseudohomophone advantage found in experiments (McCann & Besner, 1987). As with human subjects (Glushko, 1979) and the larger models, many of the acceptable non-word pronunciations are not strictly regular but analogous to irregular words.

4. Naming Latencies

Since our whole model is based on the cascaded activation approach, it is straightforward to extract reaction times (RTs) from it (such as naming latencies). We simply clamp the chosen input set (e.g. orthography) and count how many time slices it takes for the required output set (e.g. phonology) to become appropriately activated. How we define 'appropriately activated' is not totally obvious since the outputs of our network will need to be further processed for most realistic tasks (e.g. for naming our phonological output needs to be converted into real speech). It seems reasonable to assume that the time taken for the integrated phonological activation to reach some fixed threshold will give a reasonable estimation of real naming latencies (Bullinaria, 1995a). Doing this does give a realistic (noisy skewed Gaussian) distribution of RTs for our training words. With such a small set of training words it is difficult to achieve statistically significant regularity, frequency and consistency effects, though the network does show the right pattern of mean RTs (high freq. exceptions, 69.1; low freq. exceptions, 94.1; high freq. exception controls, 60.6; low freq. exception controls, 63.1; high freq. regular

TYPE	FULL NETWORK			DIRECT ROUTE		
	Correct (Reg)	Correct (Any)	RT	Correct (Reg)	Correct (Any)	RT
Words	100.0%	100.0%	62.3	99.0%	100.0%	96.8
Pseudo-homos	95.0%	99.0%	68.1	94.5%	99.0%	132.5
Non-words	88.0%	95.0%	71.9	94.5%	99.5%	130.1

Table 2: Performance on matched words, pseudohomophones and control non-words.

inconsistents, 62.9; low freq. regular inconsistents, 67.2; high freq. r.i. controls, 61.8; low freq. r.i. controls, 61.3). By reducing the thresholds we also have a natural procedure for simulating the familiar speed-accuracy trade-off effects.

A well known problem for single route reading models is the pseudohomophone effect whereby pseudohomophones are pronounced faster and more accurately than matched control non-words (McCann & Besner, 1987). From our original matched sets of 200, any triples containing errors or RTs longer than 110 were removed leaving matched sets of 163 items of each type. The mean RTs for these words and non-words are given in Table 2. For the full network, we can clearly see both the non-word and pseudo-homophone effects, which are highly significant (all RT differences $p < 0.001$). The direct O-P route on its own also exhibits the required non-word effect ($p < 0.001$), but fails to show the pseudohomophone effect ($p = 0.77$), which is the pattern usually found in single route models. It is easy to understand what is happening here. If a word or non-word orthography causes activation at the phonological level similar to that of a real word by virtue of the O-P route, then that will in turn activate semantics which will then feed back to reinforce the phonology. The semantic activation will be reduced by interference from the O-S route for non-words, and the O-P route will be less efficient for non-words based on irregular words, so the non-word effect will be reduced but not counteracted totally for the pseudohomophones. Similarly, we can see how priming by semantically related words can occur.

5. Lexical Decision

Another difficulty with single route reading models is that they do not embody any criteria for performing lexical decision. The addition of a semantic system will provide such criteria, and it will also allow the possibility of modelling semantic and associative priming. The performance of reliable lexical decision in networks like our phonology and semantics sub-network has been discussed in some detail elsewhere (Bullinaria, 1995b), so we shall restrict ourselves to providing a brief over-view here.

We know experimentally that lexical decision is speeded by prior presentation of semantically related words (semantic priming: e.g. ‘leap’ primes ‘jump’) or associated words (associative priming: e.g. ‘pillar’ primes ‘society’). It is therefore reasonable to assume that the time taken to activate semantics provides at least one factor in the lexical decision process. In Bullinaria (1995b) it was shown how the consistency of the input and output phonology in a P-S-P network could be used to perform reliable lexical decision. A network was trained on 200 words in the same way as described above, but with the added realism that the inputs were allowed to build up and decay linearly over 15 time slices with no network re-setting between words. Pairs of words were classified semantically related/unrelated (S+/S–) if they had two/no semantic units in common, and associated/non-associated (A+/A–) if they occurred one after the other 25.0%/0.5% of the time during the training process. The 200 words were split into 40 sets, each with one target and four primes A+S+, A+S–, A–S+, A–S– involving no phonological overlap. The network was activated for the prime word, then the target word was presented and the time taken for the output and input phonology to ‘match’ was measured. Such RTs were found to be significantly different ($p < 0.001$) depending on the preceding prime word: mean 39.0 for A+S+, 43.4 for A+S–, 48.5 for A–S+, and 52.8 for A–S–. This is the same pattern of semantic and associative priming that is found experimentally with humans (Moss et al., 1994). Visual lexical decision effects follow via the efficient O-P pathway. Clearly real lexical decision is much more complex than this, but it is encouraging to see that such a simple ‘activate and check’ model can account for so much of the data.

This leaves us with the second pseudohomophone effect whereby visual lexical decision for pseudo-homophones is slower than that for matched control non-words. Since our system does not have an explicit model of the complete lexical decision process, we cannot provide reliable RTs as we did for the naming latencies, but it is not hard to imagine why the conflicting ‘yes’ from phonology and semantics and ‘no’ from orthography for the pseudo-homophones should take longer to process than the unanimous ‘no’ for the control non-words.

6. Acquired Dyslexias

Various forms of dyslexia place strong restrictions on all models of reading. The loss of exception words in surface dyslexia and the loss of non-words in phonological dyslexia constitute a double dissociation and this is traditionally taken to be indicative of some degree of modularity. Consequently, it is usually explained by the loss of separate direct and semantic routes in a some form of dual route model (e.g. Coltheart et al., 1993).

Since the direct O-P route in connectionist models such as ours can deal effectively with both regular and exception words, we cannot account for surface

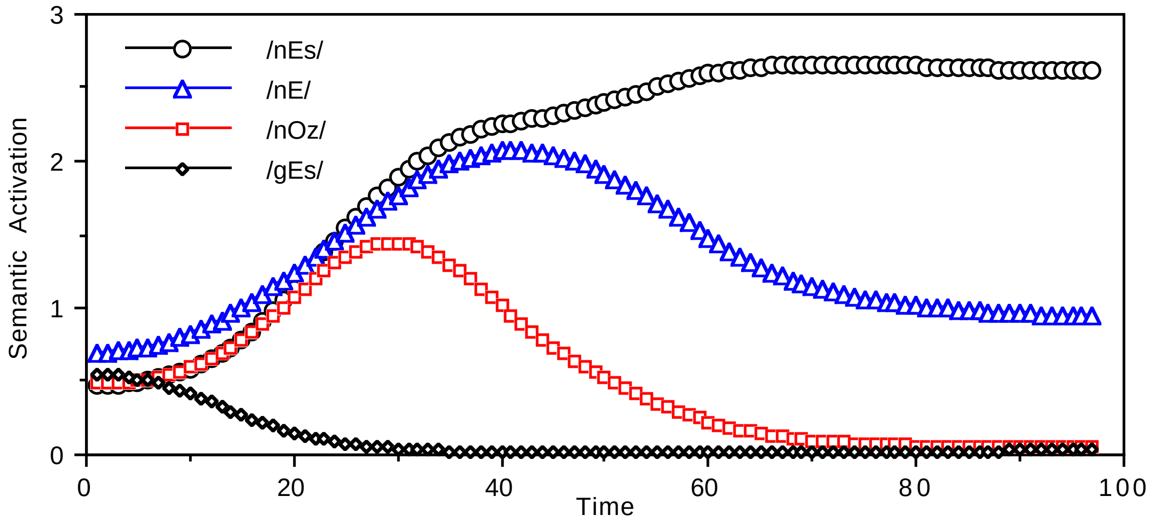

Figure 2: The 'cohort effect' arising automatically from sequential speech input.

dyslexia simply by losing the semantic route. However, simulating various types of damage to the direct O-P route does give surface dyslexic effects and these can be understood as a natural regularity effect (e.g. Bullinaria, 1994). What is potentially more problematic is the fact that the O-S-P route in our model cannot read effectively on its own and hence we cannot simply explain phonological dyslexia in terms of a near complete loss of the direct route. Fortunately, there is an alternative explanation of phonological dyslexia, with further experimental evidence, in terms of problems with phonological output assembly (Patterson & Marcel, 1992). We still have modularity, but now the modules operate in series rather than in parallel.

Many aspects of another acquired reading disorder, known as deep dyslexia, have already been modelled in a connectionist system (Plaut & Shallice, 1993). If we damage our model, visual errors (e.g. 'cat' for 'cot') can occur, and we do find concrete words are more robust than abstract words defined as having fewer activated semantic units (e.g. 90% correct versus 45%). However, semantic errors (e.g. 'cat' for 'dog') and combination errors (e.g. 'trees' for 'orchid' via 'orchard') tend not to occur. Whether, this is because we need basins of attraction in our model (that are able to perform 'lexical capture'), or simply because of our un-natural (binary) semantic vectors, remains to be seen.

7. Sequential Inputs

We have ignored many important questions such as the alignment of the orthography and phonology and the mechanisms by which the activation proceeds into and out of the input/output 'buffers' of our network. Many of these problems have already been addressed elsewhere (e.g. Bullinaria, 1997; Houghton et al., 1994) and we can hope that such features will not prove too difficult to introduce into our general framework. One feature, however, that does appear to seriously conflict with this framework is that some inputs (e.g. real speech) are naturally sequential in nature and there are well known 'cohort effects' in semantic activation that cannot possibly be captured by our static inputs (e.g. Marslen-Wilson, 1987). One way to accommodate this would be to present the phonological input sequentially at the input of the network as it occurs and then allow a series of recurrent connections to store the necessary information over time. Not only does this not fit in well with our simplified framework, but it is well known to be extremely difficult (if not impossible) for networks to learn to remember information across many time slices. An alternative that fits in better with the above is to assume that some earlier stage of processing is able to convert the speech sequence into a static auditory image (e.g. Patterson et al., 1995) that is essentially our networks phonological input buffer. If we assume that the phonemes arrive in this buffer in order sequentially, then we can expect this sequentiality to propagate through to the semantic system as well.

To test this idea we retrained a phonology to semantics sub-network with the input phonology building up over time rather than appearing complete instantaneously. For each word the activation of the onset unit increases linearly from zero to one over the first 20 time slices and then stays at one, the vowel activation increases from zero between slices 20 and 40 and then stays at one, and finally the offset activation increases from zero between slices 40 and 60 leaving the whole phonology activated from slice 60. Figure 2 shows how the semantic activation varies for one representative set of four words. Initially all words consistent with the first phoneme /n/ are activated. Then, as the second phoneme /E/ appears in the input, the word /nOz/ loses activation whilst /nE/ and /nEs/ continue rising. Finally, as the whole phonology is available, the unique consistent word remains active whilst the others fall to the levels appropriate to their semantic relatedness to the actual word. It is easy to see how a similar procedure can be used to account for the left-to-right serial nature of reading (cf. Coltheart & Rastle, 1994).

8. Conclusions

We have presented a general framework (shown in Figure 1) for the connectionist modelling of reading, spelling and related tasks, together with the results of various explicit small scale simulations that show how this framework allows one to remedy some of the major deficiencies of earlier single route models of reading. In particular, these models exhibit the pseudohomophone effect for naming and are able to perform reliable lexical decision with realistic semantic and associative priming. It is also interesting to note that the various sub-systems do not naturally interact in the manner assumed by more traditional dual route models with separate direct and lexical/semantic routes (e.g. Coltheart, 1993).

We still need to experiment with larger versions of the networks with more training words, more realistic input/output representations and training regimes, uncompressed frequency distributions and context information to deal with homographs and homophones. However, the successes of the models presented here suggest that we do have a promising framework for the modelling of reading acquisition and the related tasks of spelling, speech recognition, and so on. Moreover, we can see that many of the properties of these models are actually independent of their precise details and hence we can expect to make the various aspects of them increasingly realistic without undoing our early successes. If we are to understand, and be in a position to remedy, the various problems that children face when learning to read (such as phonological dyslexia) we need to be sure that we are looking in the right place for those problems. The next stage of this work (in progress) is hence to consider more explicitly the semantic system and the speech input and output mechanisms and to investigate their properties when operating in conjunction with the models outlined here. We will then also be in a good position to model the emergence of morphological effects, time course effects, cross modal priming, and so on.

9. References

Bullinaria, J.A. (1994). Internal Representations of a Connectionist Model of Reading Aloud. *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society*, 84-89. Hillsdale, NJ: Erlbaum.

Bullinaria, J.A. (1995a). Modelling Reaction Times. In L.S. Smith & P.J.B. Hancock (Eds), *Neural Computation and Psychology*, 34-48. London: Springer.

Bullinaria, J.A. (1995b). Modelling Lexical Decision: Who needs a lexicon? In J.G. Keating (Ed.), *Neural Computing Research and Applications III*, 62-69. Maynooth, Ireland: St. Patrick's College.

Bullinaria, J.A. (1997). Modelling Reading, Spelling and Past Tense Learning with Artificial Neural Networks. *Brain and Language*, in press.

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. & 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.

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.

Houghton, G., Glasspool, D.W. & Shallice, T. (1994). Spelling and Serial Recall: Insights from a Competitive Queuing Model. In G.D.A. Brown & N.C. Elllis (Eds), *Handbook of Spelling*, 365-404. Chichester: John Wiley.

Marslen-Wilson, W. (1987). Functional parallelism in spoken word recognition. *Cognition*, **25**, 71-102.

McCann, R.S. & Besner, D. (1987). Reading Pseudohomophones: Implications for Models of Pronunciation Assembly and the Locus of Word-Frequency Effects in Naming. *Journal of Experimental Psychology: Human Perception and Performance*, **13**, 14-24.

McClelland, J.L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review*, **86**, 287-330.

Moss, H.E., Hare, M.L., Day, P. & Tyler, L.K. (1994). A Distributed Memory Model of the Associative Boost in Semantic Priming. *Connection Science*, **6**, 413-427.

Patterson, K., & Marcel, A. (1992). Phonological ALEXIA or PHONOLOGICAL Alexia? In J. Alegria, D. Holender, J. Junça de Morais & M. Radeau (Eds), *Analytic Approaches to Human Cognition*, 259-274. Amsterdam: Elsevier.

Patterson, R.D., Allerhand, M.H. & Giguère, C. (1995). Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform. *Journal of the Acoustical Society of America*, **98**, 1890-1894.

Plaut, D.C., McClelland, J.L., Seidenberg, M.S. & Patterson, K.E. (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.

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

Sejnowski, T.J. & Rosenberg, C.R. (1987). Parallel Networks that Learn to Pronounce English Text. *Complex Systems*, **1**, 145-168.

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.

Van Orden, G.C., Pennington, B.F. & Stone, G.O. (1990). Word Identification in Reading and the Promise of Subsymbolic Psycholinguistics. *Psychological Review*, **97**, 488-522.