THE UNIVERSITY *of* EDINBURGH

An anatomical perspective on sublexical units: The influence of the split fovea

An anatomical perspective on sublexical units: The influence of the split fovea

Richard Shillcock & Padraic Monaghan

University of Edinburgh

January 10, 2003

11,522

Abstract

We discuss the problem of how to represent the internal structure of English words, and consider the solutions adopted in a number of implemented models of visual word recognition and naming. We describe two sets of simulations with the split-fovea model, an implemented connectionist cognitive model of single-word reading, whose architecture is based on the precise vertical splitting of the human fovea. We show that the model can capture critical human data from two effects concerned with the parallel activation of lexical competitors: (a) the transposed letters effect, in which pairs of words like salt and slat, or clam and calm, interact during processing, and (b) the neighbourhood effect, in which large lexical neighbourhoods facilitate naming. We discuss the results in terms of the coarse coding generated by the architecture of the split-fovea model and the naming task. Finally, we consider some of the implications for the processing of different languages and different orthographies, and for language impairment.

An anatomical perspective on sublexical units: The influence of the split fovea

We have demonstrated elsewhere that the fact that the human fovea is precisely vertically divided is a crucial point of departure for the computational modelling of visual word recognition (see, e.g., Monaghan, Shillcock & McDonald, submitted; Shillcock, Ellison & Monaghan, 2001; Shillcock & Monaghan, submitted). Researchers have proposed a number of solutions to the question of how to represent the structure of words in reading. In this paper we review some of these solutions and situate our own approach with respect to them. We then present new simulations showing that the split-fovea model can capture human data from experiments on the interaction of words related by letter transposition, such as slat and salt, and on the recognition of words with large or small lexical neighbourhoods.

One of the strengths of the computational modelling approach is that it requires the researcher to be explicit about representational assumptions. In the computational modelling of visual word recognition and naming, the representation of the visually presented word must be specified; the positions of the constituent letters must be coded. How does information about letter location participate in lexical processing? Word recognition is a task for which mammalian visual cognition did not specifically evolve. The human visual system has evolved to remember and to distinguish such things as conspecifics, objects and places in the natural world. Words, however, are unlike any of these things. They are typically asymmetric, and one half of a word does not usually determine what the other half contains. Every letter, at every point in the word, counts towards the identity of the word. Finally, there are tens of thousands of words in an adult lexicon, all different and unique. Given the nature of the problem, we are entitled to return to first principles. We will first consider some of the cognitive requirements of visual word recognition. We will then look at two observable effects, the transposed letters effect and the neighbourhood facilitation effect, that have implications for how items in the lexicon compete to be matched against a visual stimulus. We will discuss a number of models of lexical processing from the perspective of these two effects, and draw general conclusions about how best to specify letter position. Finally, we will present simulations of the two effects, using an implemented version of the split-fovea model. Our example language – and that of the simulations described – will be English, but we will consider the issue of different orthographies when we explore one of the critical dimensions in lexical processing, the coarseness of the coding required.

The requirements of visual word recognition

We will frame our discussion of word recognition in terms of two related questions: What types of information are used, and how much information is necessary? In naturalistic reading tasks, the participant needs to distinguish the word being read from all of the other possibilities in the lexicon. Efficiently achieving this goal may not require the brain to access all of the potential information about the word being read, or about the words stored in the lexicon. (At the same time, it should be appreciated that the many different relationships between words are often highly intercorrelated, and there is considerable redundancy.) If the reading task is relatively demanding – for instance, proofreading, reading degraded stimuli, or rejecting particularly word-like nonwords – we can expect more of the potential information available to be accessed, and for strategic effects to become more evident. In general, in this discussion, we will be concerned with the more naturalistic reading tasks, particularly naming.

In order to appreciate the role of information about letter position and sublexical units, it is useful to distinguish between the global and the local information available about the word being recognized. Perhaps the most salient aspect of global information about a word is its length: in the normal reading of text, it may be available parafoveally, and it plays a pervasive role in determining fixation behaviour (e.g., just & Carpenter, 1980; Rayner, 1979). Indeed, word length might be a good candidate for the early, fast uptake of low spatial frequency information by the magnocellular pathways. Word length is a powerful predictor of lexical processing in certain tasks. In the large-scale word naming datasets reported by Balota and Spieler (1998) and Spieler and Balota (1997), word length in number of letters is a good predictor of reaction time, for both young and old participants (r = .365, N = 1589, p < .001; r = .320, N = 1589, p < .001, respectively). Shorter words are named faster. Indeed, number of letters is a better predictor than number of phonemes, as defined in the MRC Psycholinguistic Database (Coltheart, 1981) (r = .265, N = 1545, p < .001; r = .200, N = 1545, p < .001, respectively). However, in Balota, Cortese and Pilotti’s (1999) large-scale visual lexical decision datasets, word length is not a strong or consistent predictor of reaction time for young or old participants: for number of letters: r = .024, N = 1598, n.s.; r = .046, N = 1599, p = .033, respectively. For number of phonemes: r = .035, N = 1553, p = .085; r = .053, N = 1554, p = .018, respectively. The effect is in the same direction as for naming, but other variables such as age of acquisition or word frequency are better predictors and, indeed, when age of acquisition is controlled in a partial correlation, no significant relationship survives between word length in number of letters and lexical decision time (for the younger participants, r = .0069, N = 448, n.s.). In contrast, word length becomes a stronger predictor of naming time when age of acquisition is controlled for (r = .402, N = 444, p < .001). Taken together, there results seem to indicate that it takes readers more time to generate the pronunciation of longer words than shorter words, and that this takes place prior to naming the word (cf. Rastle, Harrington, Coltheart & Palethorpe, 2000). Word length in letters correlates significantly with a range of important lexical variables specified in the MRC Psycholinguistic Database, such as age of acquisition (r = .202, N = 466, p < .001), concreteness (r = .080, N = 1180, p = .003), familiarity (r = -.156, N = 1215, p < .001), word frequency (Kucera and Francis's "number of samples") (r = .203, N = 1685, p < .001), imagability (r = .106, N = 1209, p < .001), and meaningfulness (r = .146, N = 993, p < .001). However, despite these relationships, there are no grounds for claiming that the most salient global information about a word, its length in letters, itself contributes useful information towards lexical decision, within the constraints of the dataset used, which employed monosyllabic words of 2–8 letters in length. In the modelling approach we have pursued, we have claimed that at least some processing occurs based on the part of the word in each hemifield, initially in the absence of information about the length of the whole word.

Another set of claims concerning global processing in word recognition concerns supra-letter features, such as outline shape. Although there has been a longstanding interest in this issue (e.g., Cattell, 1886), evidence for it has generally been from less naturalistic tasks such as proofreading (e.g. Healy & Cunningham, 1992), and there have been numerous failures to find the predicted effect in those words – high frequency words

– in which it might be expected to be strongest (e.g., Besner, 1989). Nonetheless, several researchers have included in their models of word recognition, particular routes that specialize in fast, global checking of the word as opposed to analysis into abstract letter identities (see, e.g., Allen, Wallace & Weber, 1995; Besner & Johnston, 1989). Allen et al. (1995) specifically implicate the magnocellular pathway in carrying a spatial frequency characterization of the whole word, although see Majaj, Pelli, Kurshan and Palomares (in press) for the claim that encoding orthography in terms of spatial frequency channels may not be a flexible or strategically controlled process.

Finally, Sergent (1987) describes a visual lexical decision study with split-brain participants in which they were shown brief presentations of words straddling the fixation point. They were able to give a verbal report only about the half of the word that appeared in the right visual field, as only this part of the word was available to the expressive phonological processing in the left hemisphere. However, they performed significantly above chance in lexical decision, and they were significantly more likely to attempt to guess the identity of the stimulus when it was a real word as opposed to a nonword. Sergent concluded that they were able to assess the whole word at some level of processing. We might speculate that such stimuli had been the subject of some process of subcortical, pre-attentional novelty detection. These divided brains seemed capable of deciding whether a stimulus had been seen before. This study seems to provide evidence for processing of the whole word in a situation in which neither hemisphere had full access to the full extent of the word. We can make a distinction between global information about a word (e.g., its length or outline shape) and limited processing of all of the word, as in some subcortical representation. The latter seems to exist.

In summary, the processing of global features of words has received far less attention, in theories and experiments, than the analysis of words into their constituent letters or graphemes. The evidence seems to suggest that the role of such global processing is smaller than that of analytic processing, and may only be demonstrable in certain tasks. Nevertheless, its contribution may still be critical. In modelling dyslexia with the split fovea model (see Monaghan & Shillcock, in preparation), we have claimed that hemispheric desynchronization is a powerful and parsimonious explanation for the impairment of pronunciation. The coordination of the lexical processing in the two cerebral hemispheres may be particularly assisted by the early availability of a low spatial frequency, global representation of the fixated word. Indeed, a strong case has been made for the involvement of magnocellular impairment in at least some proportion of dyslexics (see, e.g., Stein & Talcott, 1999; Stein & Walsh, 1997).

Whatever cannot be explained in terms of global representations, falls into the domain of analytic processing. We will restrict this brief review to a consideration of what information is strictly necessary for the identification of a visually presented word. Elsewhere we present an analysis of how much information about letter location is required to identify single words from a large lexicon (Shillcock, Ellison & Monaghan, 2000). Considering only four-letter words, if no location information was available at all, then 34% of such words would be ambiguous; it would be impossible to distinguish between anagrams like tsar and star. The earliest available and largest-grain information about letter location concerns the contents of the two hemifields: the fixated word is divided at the fixation point and, because of foveal splitting, the two parts are initially projected to the contralateral hemispheres. At this stage in the processing of four-letter words, only 4.7% of words remain ambiguous; item and time cannot be distinguished, for instance. Readers may well have access to more detailed location information about words, but such information is not strictly required to identify most centrally fixated words. Since the anatomy of the visual system automatically provides this coarse-grained information about location when a word is centrally fixated, there is no reason to lose this advantage by adopting any more abstract analytic approach to specifying letter position (for instance, by forcing the letters of the word into some left-justified template). This analysis emphasizes letter identity over letter location. However, there is considerable evidence that the first and last letters of an isolated word receive preferential processing (see, e.g., Jordan, 1990, 1995). We will see below that the split-fovea model automatically prioritizes the processing of these “exterior letters” (see, also, Shillcock & Monaghan, 2001). Indeed, if the identity of the letters in the two hemifields of a centrally fixated word is known, then specifying the first and last letter of the word necessarily identifies all the four-letter words, and only leaves a fraction of one percent of the lexicon ambiguous: these final ambiguities are pairs like trail and trial (see Shillcock et al., 2000 for full details).

In summary, to identify a word for lexical decision or naming may not require a full specification of all of the letter position information. This analysis implies that we should find parallel and equal activation of lexical candidates that are only distinguished by such redundant information. The prime examples of such words in English are the (relatively rare) transposed-letters pairs, such as trial and trail, and slat and salt. We turn now to a consideration of these words, together with the more frequently studied example of parallel activation of lexical candidates, the effect of lexical neighbourhood on processing.

Lexical neighbourhoods and transposed-letter confusions

Lexical representations have been seen, in various models of word recognition, as being activated in parallel and competing against each other for recognition. In implemented computational models of word recognition, which we review below, such competition has been explicitly specified in terms of levels of representation, sublexical units, and the nature of the connectivity in the models' architectures. One way in which experimental psycholinguists have sought to clarify issues of lexical representation and competition has been by determining the sublexical units involved. This approach has led to attempts to specify how much lexical candidates need to resemble each other for them to compete for the recognition of a visually presented word. Experiments in visual word recognition have conventionally been directed at the study of very similar words, although modelling approaches have enabled researchers to explore subtler, more distant relationships between words, as in models based on superpositional storage (see, also, Shillcock, Kirby, McDonald & Brew, submitted). Lexical neighbourhoods have attracted experiments at least since Coltheart, Davelaar, Jonasson and Besner (1977) reported that large neighbourhoods slowed nonword classification, where neighbourhood (N) is defined as the number of words that can be created by changing one letter of the target word. In the context of the current paper, we are principally interested in neighbourhood effects on the naming of real words; we will explore the behaviour of a model that maps from the orthographic form of a word to its phonological form. As recent reviews of neighbourhood effects (Andrews, 1997; Mathey, 2001) make clear, a number of studies have been reported that have shown that a large lexical neighbourhood facilitates the naming of a word (e.g., Andrews, 1989; 1992). This effect of neighbourhood size is the core effect in this field, but it is hedged by contradictory findings when manipulations include word frequency, language, type of task, and neighbourhood structure. The neighbourhood size effect is a challenge to models of word recognition and naming, but there is a second challenge in proposing a modelling framework that can resolve the apparent contradictions in this area.

A loosening of the criteria for defining a lexical neighbourhood allows the order of letters to be minimally disturbed. In English, this manipulation places transposed-letters pairs such as <u>salt</u> and <u>slat</u> in the same neighbourhood. Chambers (1979) first showed that such pairs appear to interact during lexical access, slowing lexical decision responses both for words and for nonwords. Specifically, she investigated this transposed-letters confusability effect on the lower frequency member of each pair, and she concluded that information about letter order could not prevent the parallel activation of the higher frequency member by its lower frequency pair. Andrews (1996) explored the effect more fully, and showed that in the naming task there was a significant interaction between transposed-letters status and word frequency: the high frequency member of the pair (e.g. <u>salt</u>) was inhibited compared with a control word, and the low frequency member (e.g. <u>slat</u>) was somewhat facilitated. Figure 3a shows the pattern of results. (The results in the lexical decision task were comparable.) Although this effect was not robust over every analysis of every experiment, it represents a challenge to models of word recognition and naming, in that letter position information seems to have been relaxed in favour of letter identity information.

In summary, the neighbourhood size effect and the transposed-letters effect both speak to the question of how information about letter identity and position are coded. We now review a number of implemented models of word recognition and naming and assess how each has responded, or might respond, to these critical data.

Implemented models and sublexical units

One of the major distinctions between implemented computational models of lexical processing is that between models in the hierarchical, symbolic, localist tradition of the Interactive-Activation Model (IAM), and models in the distributed tradition of the Seidenberg and McClelland's (1989) developmental model. However, as Andrews (1996) notes, there is no necessary difference between their behaviours: the latter models could, in principle, converge on internal representations that were effectively the same as those hand-wired into the IAM. In that case, their behaviours would be functionally the same. Both types of model are essentially architectures that calculate relatively opaque statistics over the lexicon and generate outputs accordingly. Although their capacity for graded effects on the output may be affected by architectural parameters such as the number of layers between input and output, or the degree and type of recurrence in the connections, their input representations will also have an important effect. Below, we briefly review a number of implemented computational models of lexical processing and discuss the ways in which their input representations might condition their behaviour, with particular reference to the transposed letter and neighbourhood effects reviewed above.

The Interactive-Activation Model. The Interactive-Activation (IAM) model is the architecture that inspired the last two decades of cognitive modelling research in visual word recognition (McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1981). It defined the paradigm in many ways, and has been variously developed (e.g., Jacobs & Grainger, 1992; Coltheart, Rastle, Perry, Langdon & Ziegler, 2001). Words were presented as visual features over an input divided into letter positions. The features activated letters, which in turn activated words. The recurrent flow of activation between the word level and the letter level, together with winner-take-all competitions within the higher levels, conspired to allow complex graded behaviours in terms of the rise and fall of patterns of activation over the output units. The initial version of the model illustrated some important principles. First, its authors made minimal assumptions and allowed the behaviours of the model to emerge, largely from the complexity of its lexicon, which was psychologically realistic along at least one dimension – that of the number of English four-letter words it could recognize. Second, Rumelhart and McClelland established the convention of representing letter position in terms of slots, one for each letter position in an abstract wordform. Third, they showed that this simple architecture could be modified to produce more specific behaviour. For instance, they showed that the timing of the input could be manipulated so as to encourage the model to prioritise the representation of the exterior letters, in order to capture the exterior letters effect, in which such letters were apparently processed first, or accorded a higher priority, in the Reicher-Wheeler task (cf. Jordan, 1990, 1995).

On the one hand, the IAM is a critical demonstration of the flexibility and behavioural productivity of a simple architecture. Coltheart and Rastle (1994) have shown that a version of the IAM is capable of simulating the facilitation of lexical processing by a large lexical neighbourhood. On the other hand, quite different behaviours may emerge depending on the nature of the parameter setting (cf. Jacobs & Grainger, 1992). Whatever parameters are used to constrain such a model need to be well motivated, and the same parameters need to elicit a range of different data.

Alternative coding schemes have been developed, and the slot-based ones have been elaborated. For instance, in McClelland's (1986) PABLO programmable blackboard model, bigrams were used, so that letters were represented as following or preceding another letter. This model was designed to allow anagrams to be activated, by encoding medial letters in pairs rather than in trigrams. A further alternative has been to allow neighbouring letter positions in slot-based models to influence each other (cf. Peresotti & Grainger, 1995).

In summary, interactive-activation architectures can capture the neighbourhood size effect, but even the simplest architectures, incorporating a slot-based input schema, show such flexibility that the problem has become one of constraining them in a credible way. Slot-based inputs have always begged the question of how impervious each position is to interference from neighbouring positions. The architecture can be elaborated, using bigrams or spreading activation from adjacent slots, to capture the transposed letters effect, but it is critical to motivate such elaborations by something other than the human data from this particular effect.

BLIRNET. Mozer (1987) describes a way of representing letters and words that employs a variant of the Wickelgraph (cf. Wickelgren, 1969). In his model, BLIRNET, words are input to the model in terms of five different spatial features appearing at any of the points in its 36 x 6 “retina”. In a series of additional layers, increasingly complex and increasingly location invariant representations are developed, until BLIRNET’S output layer codes words in terms of triples of letters across four consecutive slots. Thus, cat is represented by the letter-cluster units **C, **_A, *CA, *_AT, *C_T, CAT, C_T*, CA_*, AT*, A_**, T**, in which * stands for a white space and the _ stands for any letter. BLIRNET’S final weights, the ones closest to the output, were modifiable. BLIRNET was designed to capture psychological data concerning tasks in which different stimuli were presented at the same time. Thus, the set of activated letter clusters corresponding to two simultaneously presented words would be parsed into separate lexical representations by another component of the model, the “pull out net”. The units in this component were connected with excitatory and inhibitory links that captured the orthotactic and lexical regularities in the lexicon. From this summary, it can be seen that this way of representing words allows neighbourhood relationships to participate in processing in so far as lexical neighbours possess similar components and the training of the last set of weights will be affected by the frequency of the words presented. BLIRNET was not designed to model the fine detail of single-word processing. However, Mozer and Behrmann (1992, p. 415) state that among the spurious letter-clusters the model partially activated when resources were limited were ones in which adjacent letters in the stimulus were transposed (e.g., ENY being activated by MONEY), thus raising the possibility that their model would show some version of the transposed-letters effect.

<u>Developments of the IAM</u>. Brown (1987) and Norris (1994) both describe developments of the IAM architecture in which sublexical units are stipulated in the architecture: Brown implemented a multiple levels model of pronunciation using a hierarchy of letters, bigrams, trigrams and words, and Norris implemented an interactive-activation architecture using the sublexical units proposed by Shallice, Warrington and McCarthy (1983), dividing the output into onset, nucleus and coda. Brown was concerned only with the implications for consistency in pronunciation and he reports only small-scale simulations. However, there is every likelihood that this model would capture neighbourhood effects, due to the greater frequency of those constituents that occur across different words. Some researchers have seen the inhibitory connections within the levels of interactive-activation architectures as the locus of neighbourhood effects (see, e.g., Mathey, 2001); the absence of such connections between competing phonological units in Brown's model is therefore of interest in this respect. Norris demonstrated that his model, when parameterized using known datasets, could capture facilitatory neighbourhood effects on word naming, as reported by Andrews (1989). In contrast, both models stipulate letter position within the word when it is input to the model, and therefore offer no obvious prospect that they could capture the TL effect.

<u>The Seidenberg and McClelland Developmental Model</u>. In a move away from the limitations of the IAM's hand-coded architecture, Seidenberg and McClelland's (1989) developmental model was able to learn the mapping between distributed orthographic and phonological representations of words. The original version of the model contained

an orthographic coding scheme that was based on Wickelgraph triples of letters: salt was represented as #sa, sal, alt, lt#. The authors introduced a further, more idiosyncratic level of coding to ensure that the lexical representations were coarse, to promote generalization between similar words. They used the same input unit to represent a number of different Wickelgraphs.

This coding scheme has the advantage of allowing words to be represented without explicitly specifying their formal structure or even their overall length. However, the TL pairs like salt and slat do not share a single Wickelgraph: #sa, sal, alt, lt# against #sl, sla, lat, at#, respectively. Some degree of overlap could be introduced by encoding the beginning and end of the word by a double hash, as in BLIRNET, so that the first and last Wickelgraphs – ##s and t## – are shared, but even coding the end of the word with a single hash means the theoretical expense of specifying additional information beyond the letters themselves.

In summary, as Andrews (1996) observes, the conventional Wickelgraph representation used in Seidenberg and McClelland's original model provide no grounds for the model to be able to capture the TL confusability effect. In contrast, this Wickelgraph representation captures some of the similarity between conventional lexical neighbours: for instance, salt and silt share one out of their four Wickelgraphs, and salt and sale share two. The prediction that this Wickelgraph representation will support the neighbourhood facilitation effect in English is borne out (Seidenberg & McClelland, 1989).

Developments of the triangle model. Seidenberg and McClelland's Developmental Model framework has been elaborated and explored to provide coverage of a wide range of data (e.g., Harm & Seidenberg, 1999, 2001; Plaut, McClelland, Seidenberg & Patterson, 1996; Seidenberg, Plaut, Petersen, McClelland & McRae, 1994). One of the principal representational advances in this "triangle model" framework has been the replacement of the Wickelgraph input with a graphemic input parsed into the syllabic constituents of onset, nucleus and coda. This input is simpler than a raw slot-based input, and it allows the model to capitalise on orthotactic/phonotactic constraints present in the language. Performance in nonword pronunciation is humanlike. However, one of its drawbacks is that it seems to provide only limited grounds for capturing the transposed-letters effect. In our salt/slat example, the l moves between the onset and the coda. The two words will share s in the onset, a in the nucleus, and t in the coda, which motivates some shared processing, but no more than between salt and spat. To our knowledge, there is no report of a version of the triangle model capturing the transposed-letters effect. However, a neighbourhood facilitation effect is captured by these models.

SERIOL. Whitney (2001a, b) describes the SERIOL model of lexical processing, which was designed principally to account for letter-ordering data from experiments with normal and impaired readers. In this framework, the ordering of letters is achieved by means of a serial process corresponding to an underlying oscillation. The architecture consists of levels containing letters, bigrams and words, and the input from the letter level to the bigram level is graded, monotonically decreasing from left to right for English. This activation gradient, corresponding to a locational gradient, is achieved at the lowest level by assuming that there are hemispheric differences in inhibition, and that the fixated word is divided between the two hemifields/hemispheres; effectively, the reader learns to transform the symmetrical acuity gradient into a left to right activation gradient. In the reading of the word cart, the activation gradient is established at the feature level, the constituent letters are activated in order at the letter level, the ordered bigrams CA, AR, RT, CR, AT, and CT become active at the bigram level, and finally cart becomes most active at the word level. Although Whitney does not specifically test the SERIOL model against either the neighbourhood size effect or the transposed-letters effect, the structure of the model suggests that its default behaviour would allow it to respond to the neighbourhood structure of the lexicon in a similar way to other interactive-activation architectures. SERIOL's precise behaviour with transposed-letters pairs is an empirical question, but the fact that cart partially activates bigrams such as CR and AT suggests that interaction between such pairs may not be beyond the model's abilities.

From this brief and partial review of implemented models of visual word recognition, we can see that the issue of input representation is critical: how are the letters of words represented in their respective positions? The facilitation of lexical processing by large neighbourhoods seems to be within the abilities of most models, given that most specify the common processing of sublexical constituents shared between different words. However, the transposed-letters effect seems to require the modeller to stipulate the relevant representational types (such as ngrams containing non-adjacent letters) or additional architecture (such as graded excitation between adjacent letter slots), to avoid a slot-based model specifying letter position too inflexibly. Whatever elaboration of the model is necessary to allow it to capture both effects needs to be motivated, and will be more legitimate if it enables the model to capture a range of human data from word recognition experiments. We now turn to the split-fovea model, which we first describe, and then test against the two effects.

The split-fovea model

We have developed a model of lexical processing, the split-fovea model, based on the observation that the human fovea is precisely vertically divided. The model captures the fact that a fixated word is split into two parts at the point of fixation, and the two parts are initially projected to the contralateral hemispheres. We have grounded the model in this anatomical constraint, allowing the behaviours of the model to emerge from the interaction of this constraint with comprehensive representations of the lexicon at the orthographic, phonological and semantic levels in different versions of the model. This architecture has allowed us to simulate a wide variety of normal and impaired reading behaviours. For a full description of the development of the model from these anatomical assumptions, and for analysis of its behaviours, see Shillcock, Ellison and Monaghan (2000), Shillcock and Monaghan (2000), Monaghan, Shillcock and McDonald (submitted), and Shillcock and Monaghan (submitted).

A version of the model is shown in Figure 1, with an architecture designed to deal with words of up to five letters. Figure 1 shows the model's "fixation point" (i.e. the split point) falling between the second and third letters of the input word. Analyses of corpora of eye-tracking data show that, although there are tendencies towards particular fixation locations within words, any particular word may eventually be fixated at any point along its length, depending on the context in which it occurs, and the shorter a word is, the more likely it is that it may not be directly fixated at all (see Rayner, 1998, for a review). We have so far made minimal assumptions about fixation behaviour in reading, by requiring the model to be able to process each word when it is fixated at all positions between letters, including the two positions in which either one or the other end of the word directly abuts the fixation point. The full training regime for a four-letter word is illustrated in Figure 2: the model is required to be able to process each word in every possible position with respect to the fixation point (with the exception of eccentric locations in which the fixation point does not abut the end of the word).

In Figure 1, the crossed connections to the hidden units reflect the contralateral organization of the brain, and allow us to equate the right-hand side of the model with the RH of the brain. Connectivity between the two halves of the model is necessary for it to be able to deal with non-componential pronunciations like pint. Without such connections the model, faced with generating the pronunciation of pint, resembles a perceptron unable to solve a linearly inseparable mapping problem like Exclusive-or. The model's "callosal connections" between the two sets of hidden units allow it to produce the necessary internal representations. For any one input, activation passes from the input units to the hidden units, and, at successive points in time, activation begins passing between the two sets of hidden units, and from the hidden units to the output units, which converge over time onto the desired output pattern.

The split-fovea model incorporates many of the same insights that have motivated the models reviewed above. Like many of the models, it employs a slot-based representation. Unlike some of them, it contains no explicit specification of the beginning and end of the word, or of any subsyllabic structure within the word. However, as Figure 2 shows, the training regime presents the model with the range of consecutive letter positions, some of them correspond to onsets, rimes, and codas. Furthermore, the model may very well develop representations of ends of the word and of formal constituents in the intermediate structure it learns. Like BLIRNET, the split-fovea model carries out a shift-invariant mapping: it can process an input word irrespective of position. Like SERIOL, it capitalises on the fact that the human fovea is split. Like Seidenberg and McClelland's developmental model, it is able to learn whatever intermediate, graded, sublexical representations it needs, and like the later versions of the triangle model, it maps onto a phonologically structured output. Finally, like many of the other models, the split-fovea model uses recurrence to develop its output over time. Despite the complexities of its motivation, the version of the split-fovea model we present here is functionally simple: complex processing may arise due to its input schema and to its callosal connectivity, but there is a simple, anatomical foundation for such structures in the model.

Does the split-fovea model behave in ways that are comparable to human readers? If it is intuitively difficult to see how the integrity of a lexical representation is maintained under conditions of superpositional storage (Besner, Twilley, McCann & Sergobin, 1990), then it is even more difficult to do so when multiple staggered representations are stored, when words of different length are used, and when the majority of the input words have been cut into two parts. It is critical for us to be able to show that lexical identity does not simply disintegrate under the pressure of the task. In the two experiments reported below, we tested the hypothesis that the split-fovea model would be able to capture the human data from the two effects discussed above, namely transposed-letter confusions and the facilitation of lexical processing by large neighbourhoods.

EXPERIMENTS

Architecture

The architecture of the version of the split-fovea model that we employed is shown in Figure 1. Each half of the input layer contained five letter slots. A letter was represented in a particular position by the activation of one unit from the 26 units in the appropriate letter slot. The output layer consisted of six phoneme slots, each of which was composed of 11 phonological features, taken from Harm and Seidenberg (1999). There were two slots each for the onset, nucleus and coda of each word.

Training and testing

We extended our previous simulations, which involved only four-letter words, by training the model to read monosyllabic words of length one to five. We took the words from the CELEX English database (Baayen, Pipenbrock & Gulikers, 1995), omitting only those wordforms with onsets or codas with more than two phonemes. This procedure left 5165 words. We presented each word equally often at each fixation position. For a word of length n, there were $n + 1$ possible presentation positions. Words were presented according to their log-frequency, divided by the number of presentation positions for that word (e.g., dividing by 6 for 5-letter words, and dividing by 2 for 1-letter words). Log-frequency was used to increase the probability that the model would be exposed to low-frequency words during a training regime of a few million words (for discussion of this issue, see Plaut, McClelland, Seidenberg & Patterson, 1996).

The model was trained to map the orthographic representations onto their respective phonological representations, using a standard recurrent backpropagation learning algorithm with gradient descent. In a single training trial, at time 0, the word was presented at the input layer. At time 1, activity reached the hidden layer. At time 2, activity reached the output layer, and also crossed the callosal connections between the two halves of the hidden layer. Over five further time steps, the model settled onto the required phonological representation of the word, with activity cycling between the two halves of the hidden layer and propagating to the output layer. After seven time steps, connections between units were updated, the activity in the model was reset to a resting state, and the next word was presented at the input. We presented 10 million tokens of words altogether, at which point the model had learned the task well.

Experiment 1: Transposed letters

In the first simulation, we tested the hypothesis that the split-fovea model would capture the naming time data from the transposed letter effect reported by Andrews (1996), reviewed above. There should be an interaction between the transposed letter effect and word frequency.

Materials. We took from Andrews (1996, Appendix) the pairs of words with transposed letters (TL words: e.g. salt and slat) and the pairs of control words, each matched with their TL words for initial letter and frequency (e.g., for salt and slat, the controls were sand and slop). In each pair of TL words, there was a higher-frequency and a lower-frequency item: for instance, salt and slat respectively. We omitted sets that contained a TL word that was longer than five letters, or that was polysyllabic, or that had more than two phonemes in its onset. One pair of TL words, fist and fits, had relative frequencies in the CELEX corpus that were opposite to those reported by Andrews, so we reversed set membership for these items and their control words. There were 16 words in each condition.

Results

We averaged a word's mean squared error (MSE) across all presentation positions to simulate the response time to name that word. We entered MSE into an ANOVA with TL status as a within-items variable and frequency as a between-items variable. There was no main effect of frequency, $F(1, 30) = 0.160$, $p = 0.692$, and no main effect of TL status, $F(1, 30) = 0.396$, $p = 0.534$. However, the interaction between TL status and frequency was marginally significant, $F(1, 30) = 2.965$, $p = 0.095$. The interaction was qualitatively similar to that reported by Andrews (1996) for the human data: higher frequency TL words had increased response time (increased error) over lower frequency TL words, compared with control words. Higher frequency TL words had MSE of 4.402 and higher frequency controls had MSE of 3.665, whereas lower-frequency TL words had MSE of 3.563 and lower frequency controls had MSE of 5.149. Figure 3 compares the data from the simulation with the human data.

There was a high standard deviation across the means, which contributed towards the absence of a main effect of frequency. Part of the variance in the data was due to the model reading some words incorrectly – misreading one phoneme could result in a large increase in error. We excluded data from a word at a particular input position if the resulting pronunciation contained an "incorrect" phoneme, that is, one that was not closest to the target phoneme. This procedure effectively removed the outputs with very high error. We performed an ANOVA on the MSE with these particular outputs omitted. The effect was similar to that of the initial analysis. There were no significant main effects, but the marginally significant interaction between TL status and frequency remained, $F(1, 30) = 4.151$, $p = 0.051$. Higher frequency TL words had higher MSE than control words (4.279 and 3.420, respectively), whereas lower frequency TL words had lower MSE than control words (3.508 and 4.204, respectively). Post hoc comparisons between these values were not significant.

Discussion

In Andrews' (1996) analyses, the main effects and interactions were clearest in the analyses by subjects, and often were not found in the analyses by items. For example, the by items interaction for the naming latency data in Andrews' Experiment 1 reached a significance level of 0.05, precisely as in our simulation. Andrews' Experiment 3 also allows a direct comparison with our simulation; in this experiment, the interactions for naming latency and for accuracy were only significant in the by subjects analyses.

The analysis we report above is a by items analysis as the results are obtained from one simulation, one run of the model. The data from a number of additional runs, entered as subjects, would allow a clearer interaction between TL status and frequency to emerge. Our simulation results were also limited by the fact that we could not employ the full set of words from Andrews (1996), meaning that there is lower power in the statistical analyses of our simulations than in the analyses of the human data. Nevertheless, within the constraints of the current simulations, we have found an interaction between TL status and frequency that approached significance, and qualitatively resembled the human data.

Finally, the transposed letters effect did not emerge in pilot studies employing versions of the split-fovea model that were trained to auto-associate orthographic representations onto identical orthographic representations. This difference between models mapping orthography to orthography and orthography to phonology suggests that the transposed letters effect is at least partly due to the coarser coding inherent in generating the pronunciation of a visually presented English word, compared with the fine coding, letter by letter solution that is possible when the task is simply one of reproducing an orthographic representation of a word. We return to this issue in the General Discussion.

Experiment 2: Neighbourhood effects

In the second simulation, we tested the hypothesis that the split-fovea model would capture the basic data associated with the neighbourhood effect, as reported by Andrews (1992), as reviewed above. A large lexical neighbourhood should facilitate word naming.

Materials

We took the set of words in Sample 1 from Andrews (1992). This list comprised 24 large neighbourhood words and 24 small neighbourhood words, with each set containing 12 high-frequency words and 12 low-frequency words. One low frequency, large neighbourhood word, sire, was not in the training set, and so we omitted one word from each of the other three sets.

Results

As in Experiment 1, we measured MSE on the output patterns to simulate the response time to name a word. An ANOVA conducted on the MSE data, with frequency and neighbourhood size as between-items factors, revealed no significant main effect of frequency, $F(1, 40) = 0.666$, $p = 0.419$, but there was a significant main effect of neighbourhood size, $F(1, 40) = 4.329$, $p = 0.044$. Small neighbourhood words produced a larger MSE. The interaction between frequency and neighbourhood size was not significant, $F(1, 40) = 0.948$, $p = 0.336$. The neighbourhood effect in the simulation appears to be largely due to the difference in MSE between the groups of low-frequency words: MSE for large and small neighbourhood low-frequency words was 3.323 and 8.417, respectively. The difference between the two high- frequency groups was smaller, but still in the same direction: for large neighbourhood words MSE was 3.586, compared with 5.433 for small neighbourhood words. However, post hoc comparisons were not significant for either set (adjusted $p = 0.186$, and $p = 0.562$, respectively).

Three of the high frequency, small neighbourhood words were not read correctly in any position: town, sure, and both were read as /tıUn/, /soor/ and /bəUt/ instead of /taUn/, /ʃoor/ and /bəUθ/, although in each case only one feature was misread. We omitted these items and balanced the groups for neighbourhood size by omitting three high-frequency, large-neighbourhood words, but did not find a significant main effect of neighbourhood size, $F(1, 34) = 2.323$, $p = 0.137$. No other effects were significant.

Discussion

The model simulated the faster naming of words with large neighbourhoods in the human data: the MSE of the phonological output for the large neighbourhood words was smaller than that for with small neighbourhoods. Andrews (1992) reported an interaction between neighbourhood size and frequency, which the model did not capture: in the human data, low frequency words with large neighbourhoods were read more quickly and accurately than low frequency words with small neighbourhoods, and there was a small trend in the opposite direction for high frequency words. The absence of a comparable interaction in our simulation was due to neighbourhood size facilitating naming for both high and low frequency words; however, the effect of neighbourhood size was larger for low frequency words.

The small size of the neighbourhood effect in the simulation is not surprising. Andrews (1992) only found a main effect for neighbourhood size in the by subjects analysis. Multiple runs of the model would provide a significant by subjects interaction between neighbourhood size and frequency.

GENERAL DISCUSSION

We have responded to Andrews (1996) challenge to provide a model that can produce both the neighbourhood size effect and the interaction between transposed-letters status and frequency in English. The pronunciation of higher-frequency TL words was inhibited in the model, whereas the pronunciation of lower-frequency TL words was facilitated, due to the latter receiving a boost from their orthographically similar higher-frequency

TL pair-word. Large lexical neighbourhoods facilitated the accurate pronunciation of their constituent words, and this support appears to be particularly helpful for low frequency words. The simulations we have reported were conducted with subsets of the stimulus materials from the human experiments, with the same parameters for the two simulations and with no delicate manipulation of the parameters to achieve the effects. The reported behaviours are qualitatively close to the human data despite the fact that the model is relatively idealised in many respects.

The model's performance in the Experiments reported above deviated from the human data in one critical respect: We did not find a significant main effect of word frequency in either simulation. However, over all the words in the training set, the correlation between frequency and MSE was significant, $r = -0.13$, $N = 5165$, $p < 0.001$, with higher frequency words resulting in lower MSE. The split-fovea model, in common with other connectionist models of lexical processing, simulates the word frequency effect: repeated training on particular mappings insinuates those mappings into the learned structure of the model more effectively. Over the entire training set, the word frequency effect is significant, but does not account for a very substantial proportion of the variance. However, when we look more closely at the human data, this apparently small role for word frequency in the model is more interpretable. In Balota et al.'s naming time database (Balota & Spieler, 1998; Spieler & Balota, 1997) the correlations between naming time and the various frequency measures available in the MRC Psycholinguistic Database range from $r = -.008$ ($N = 1577$, n.s.) to $r = -.236$ ($N = 1577$, $p < .001$), with median values of -.065 ($p = .018$) and -.067 ($p < .005$). Although the word frequency effect has seemed for a long time to be ubiquitous in processing, it does not always account for a very large part of the variance in processing, and it may well be that there is an important role for other variables with which word frequency is highly intercorrelated. The split-fovea model produces a significant word frequency effect overall, but the model excludes many dimensions of psychologically realistic lexical processing that are correlated with word frequency and, were they to be included in the model, might increase the apparent role for word frequency.

Consider a simple model with four letter slots and a central split, representing the split fovea. In an idealised single presentation to such a model, in which the word salt or slat was centrally fixated and the letters cleanly projected to the contralateral hemispheres, a transposed-letters pair like salt and slat will only share the exterior letters: sa or sl will go to the RH and lt or at will go to the LH. Such a model does not seem to provide a basis for the transposed-letters effect. However, the split-fovea model we have explored differs critically from such a simple model, in that it contains a shift-invariant mapping in which a word of length n is presented at n + 1 different positions and is mapped onto the same output. Furthermore, there are recurrent “callosal” connections between the two halves of the model. The model prioritises the representation of the exterior letters because these letters fall predominantly – and for some presentations, exclusively – in a particular hemifield (see Shillcock & Monaghan, 2001, for further details). The other side of this behaviour is that the model gives relatively less priority to the representation and processing of the interior letters. A corollary of these behaviours is that the shift-invariant mapping allows a relaxation of the binding of a particular letter to a particular slot, and the consequences are apparent in the interaction of transposed-letters words. We have presented data for the mean MSE, calculated over all possible positions for a word. We have not explored any differences between the presentation positions because fixation position was not controlled in the human experiments. Jordan, Patching and Milner (1998) have shown that precise inferences about where participants are fixating in single-word presentations cannot be made on the basis of instructions alone. It is likely that the human data reflect more than one fixation on each word, meaning that the average MSE across all presentation positions is the most reliable measure in simulating the effect. Nevertheless, it does seem to be implicit in the model that there should be less interaction between non-adjacent transposed letters (e.g. ceihf and chief) compared with adjacent transposed letters (e.g. chief and chief), which matches intuitions about anagram solving. It should be noted that a minority (five out of 32) of the stimulus items in Andrews' experiments and in our simulations (three out of 16) involved an exterior letter (e.g., coats and coast, busy and buys). The default prediction of the split-fovea model, given the presence of the exterior letters effect, is that these items should diverge in behaviour from the other stimuli in which the transpositions involve only the internal letters. However, the number of available stimuli is too small for any exploration of that effect.

A further factor contributing to the emergence of the transposed-letters effect in the split-fovea model is the extent to which the central letter slots code the behaviour of consonants when they appear both in onset and in coda position. As Figure 2 shows, for four-letter TL words like calm and clam, the central four slots will each code both of the transposed letters in onset and in nucleus position. The split-fovea model thus partially transcends what Plaut et al. (1996) have called the "dispersion problem" – the fact that simple slot-based models cannot transfer any of what they have learned about the role of, for instance, l in onset position to l when it occurs in coda position. Partly overcoming the dispersion problem is an aid to better generalization, from one perspective, but it may also contribute to the parallel activation of TL confusable words. For further discussion of the dispersion problem and the comparison of local and distributed representations, see Bowers (submitted), Page (2000), and Shillcock and Monaghan (submitted).

The neighbourhood size effect emerges from the model because the pronunciation of English words requires the coding of representations greater than the individual letter. Some pronunciations like mill or tint can be arrived at componentially, by one-to-one mappings from each letter to its dominant pronunciation, whereas other pronunciations require more letter context. The distribution of the contexts required prefigures the formal analysis of syllabic structure, in which, for instance, the rime plays an important role. At the non-componential extreme of this spectrum are the low frequency words like pint, for which the whole of the onset and the coda together constitute the "context" necessary for the correct pronunciation of the vowel. The split-fovea model provides a range of possible pronunciation contexts by splitting each word between the two hemifields in a series of constituencies that are continuous with one or other end of the word. We discuss the pronunciation behaviours of the model more fully elsewhere (Shillcock & Monaghan, submitted). Unlike other modellers, we have not needed to specify the sublexical units into which each input word must be broken, nor have we introduced into the architecture insights gained from the formal analysis of syllable structure. Instead, the solution has emerged from the anatomy of the visual system in interaction with the nature of the task. The end result is that letter position is specified in the model in a rich, opaque manner, determined by the interaction of the anatomical givens with the requirements of the task.

Letter position is coded coarsely, as is required by the pronunciation task, but the development of this coarse coding is steered by the structure provided by the shift-invariant mapping and by the need to transfer information across the model's "callosal" connections. The consequence is that sets of letters that are shared between words are able to mediate lexical neighbourhood behaviour in the model.

Our interpretation of the split-fovea model's simulation of the transposed-letters effect and the neighbourhood size effect has drawn up on the idea of coarse coding – the representation of letter and position information partly in terms of their broader context. We have noted above that pilot simulations with the split-fovea model failed to find a significant transposed-letters effect when the model was only required to map from orthography to orthography, a relatively transparent task that may be best achieved using fine-coded, single-letter representations. This interpretation suggests a principled distinction between different tasks and between different languages to explain some of the inconsistent results regarding lexical neighbourhood effects that have been reported in the literature. Where a task – such as naming – demands relatively coarse coding, there will tend to be significant neighbourhood effects. Where a task can be accomplished using fewer coarse-coded representations, neighbourhood effects may be less strong. Whether the neighbourhood effect is facilitatory or inhibitory may depend on precise contingencies within the lexicon and within the task. Differential neighbourhood effects between languages may reflect the depth of the orthography in those languages. A language with a deep orthography, like English, should produce more reliable neighbourhood effects than a language, like Spanish, with a relatively shallow orthography. Note, however, that pronunciation inherently relies on coarse coding to some degree: even if the pronunciation of vowels is predictable, there will still be consonant clusters that might be resolved into one or more graphemes.

Finally, we turn to the issue of impairment. We have seen that the transposed-letters effect manifests itself in normal readers in terms of rather finegrain timing differences and small increases in errors, under the pressure of a timed task. As a general principle, we might expect to find qualitatively similar behaviour manifesting itself in an amplified form in impaired individuals, even under less stressful reading conditions. Such pure cases of letter migration errors are rarely reported, and the relevant behaviour seems mostly found in patients who present with other types of dyslexic behaviours and attentional problems, and even then the specific problem in coding position-specific letter information is chiefly reported for letter arrays rather than words (see Katz & Sevush, 1989; Price & Humphreys, 1993; Warrington, Cipolotti & McNeil, 1993). Friedmann and Gvion (2001) present the purest cases of what they term letter position dyslexia: two patients, BS and PY, presented with a highly selective reading problem in which word-medial letters changed place in a manner directly comparable to the TL effect observed in normal readers. It is instructive that such a pure form of this dyslexic behaviour should be found in Hebrew, given that Hebrew has an extremely deep orthography. In "unpointed" Hebrew, read by adult readers, the vowels are largely omitted meaning that the reader needs to infer the identity of the vowels by coarsely coded context. Hebrew morphology is based on tri-consonantal roots and consonant transpositions are very likely to generate legal words in Hebrew (compared with the rarity of transposed-letters word pairs in English). The two patients described by Friedmann and Gvion had both suffered strokes, which invariably disrupt the balance of hemispheric processing and/or the hemispheric transfer of information. From the perspective of the split-fovea model, we interpret this letter position dyslexia in terms of disruption of the coarse-coded orthographic representations in one or both hemispheres and in the transfer of such information between the hemispheres.

CONCLUSIONS

It is one of the goals of cognitive neuropsychologists eventually to ground cognitive processing in observable anatomy. We have produced further evidence for our assertion that foveal splitting is a fundamental characteristic of human reading, and one that conditions cognitive processing at the lexical level. The split-fovea model is constructed to take advantage of the foundational insights obtained from earlier models, particularly those in the tradition of the triangle model. We have adhered to the principles of grounding the model in the anatomy of the visual system, of using comprehensive repertoires of representations at input and output and in the training regime, and of proceeding from the simplest architectures so as to allow maximal scope for different behaviours to emerge in interaction with the structure of the problem. In the current simulations, we have successfully addressed human data that are potentially problematic for some classes of architecture and for some types of input representations. We have argued that the sublexical processing that occurs in the split-fovea model is of a coarse-coded, relatively opaque, "microfeatural" type, as opposed to processing that exploits sublexical structure provided by the modeller. Further, we have claimed that the shift-invariant input, in conjunction with the split architecture, provides the basis for the granularity of the coding, exposing particular continuous subsets of letters to further processing. Finally, we have pursued the implications of the simulations into case studies of cognitive impairment and suggested a processing explanation for a relatively rare form of dyslexia.

References

Allen, P.A., Wallace, B. & Weber, T.A. (1995). Influence of case type, word frequency, and exposure duration on visual word recognition. Journal of Experimental Psychology: Human Perception & Performance, 21, 914–934.

Andrews, S. (1992). Frequency an d neighbourhood effects on lexical access: Lexical similarity or orthographic redundancy? Journal of Experimental Psychology: Learning, Memory and Cognition, 18, 234–254.

Andrews, S. (1996). Lexical retrieval and selection processes: Effects of transposed-letter confusability. Journal of Memory and Language, 35, 775–800.

Andrews, S. (1997) The role of orthographic similarity in lexical retrieval: resolving neighborhood conflicts. Psychonomic Bulletin and Review, 4, 439-461.

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

Balota, D.A., Cortese, M.J., & Pilotti, M. (1999). Item-level ana
lyses of lexical decision performance: Results from a mega-study. In Abstracts of the 40th Annual Meeting of the Psychonomics Society (p. 44). Los Angeles, CA: Psychonomic Society.

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

Besner, D. (1989). On the role of outline shape and word-specific visual pattern in the identification of function words: NONE. Quarterly Journal of Experimental Psychology, 41, 91–105.

Besner, D. & Johnston, J.C. (1989). Reading and the mental lexicon: On the uptake of visual information. In W. Marslen-Wilson (Ed.), Lexical Representation and Process (pp. 291–316). Cambridge, Massachusetts: MIT Press.

Besner, D., Twilley, L., McCann, R. & Sergobin, K. (1990). On the association between connectionism and data: are a few words necessary? Psychological Review ,97, 432-446.

Bowers, J.S. (submitted). Challenging the widespread assumption that connectionism and distributed representations go hand-in-hand.

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

Cattell, J.M. (1886). The time taken by cerebral operations. Mind, 11, 220–242.

Chambers, S.M. (1979). Letter and order information in lexical access. Journal of Verbal Learning and Verbal Behavior, 18, 225–241.

Coltheart, M. (1981). The MRC Psycholinguistic Database. Quarterly Journal of Experimental Psychology: Human Perception and Performance, 33, 497–505.

Coltheart, M., Davelaar, E., Jonasson, J.T. & Besner, D. (1977). Access to the mental lexicon. In S. Dornic (Ed.), Attention and performance VI. London: Academic Press.

Coltheart, M. & Rastle, K. (1994). A left-to-right serial process in reading aloud. Journal of Experimental Psychology: Human Perception and Performance, 20, 1197–1211.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001) The DRC model: A model of visual word recognition and reading aloud. Psychological Review, 108, 204 – 258.

Friedmann, N., & Gvion, A. (2001). Letter position dyslexia. Cognitive Neuropsychology, 18, 673-696.

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

Harm, M.W. & Seidemberg, M.S. (2001). Are there orthographic impairments in phonological dyslexia? Cognitive Neuropsychology, 18, 71–92.

Healey, A.F. & Cunningham, T.F. (1992). A developmental evaluation of the role of word shape in word recognition. Memory & Cognition, 20, 141–150.

Jacobs, A.M. & Grainger, J. (1992). Testing a semistochastic variant of the interactive activation model in different word recognition experiments. Journal of Experimental Psychology: Human Perception and Performance, 18, 1174–1188.

Jordan, T.R. (1990). Presenting words without interior letters: Superiority over single letters and influence of postmark boundaries. Journal of Experimental Psychology: Human Perception and Performance, 16, 893–909.

Jordan, T.R. (1995). Perceiving exterior letters of words: Differential influences of letter-fragment and non-letter-fragment masks. Journal of Experimental Psychology: Human Perception and Performance, 21, 512–530.

Jordan, T.R., Patching, G., & Milner, A.D. (1998). Central fixations are inadequately controlled by instructions alone: Implications for studying cerebral asymmetry. Quarterly Journal of Experimental Psychology, 51, 371-391.

Just, M.A. & Carpenter, P.A. (1980). A theory of reading: From eye fixations to comprehension. Psychological Review, 87, 329–354.

Katz, R.B. & Sevush, S. (1989). Positional dyslexia. Brain & Language, 37, 266–289.

McClelland, J.L. (1986). The programmable blackboard model of reading. In J. L. McClelland & D. E. Rumelhart (eds.), Parallel Distributed Processing: Explorations in the microstructure of cognition, Vol. 2. Psychological and biological models. Cambridge, MA: MIT Press.

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

Majaj, N.J., Pelli, D.G., Kurshan, P. & Palomares, M. (in press). The role of spatial-frequency channels in letter identification. Vision Research.

Mathey, S. (2001). Influence of orthographic neighbourhood in word recognition tasks. Canadian Journal of Experimental Psychology, 55, 1–23.

Mozer, M.C. (1987). Early parallel processing in reading: A connectionist approach. In M. Coltheart (Ed.), Attention and Performance XII: The psychology of reading (pp. 83–104). London: Erlbaum.

Monaghan, P. & Shillcock, R.C. (in preparation). The split-fovea model of reading: Part 2. Connectionist modelling of dyslexia.

Monaghan, P., Shillcock, R.C. & McDonald, S. (in press). Hemispheric asymmetries in the split-fovea model of semantic processing. Brain and Language.

Mozer, M.C. & Behrmann, M. (1992). Reading with attentional impairments: A brain-damaged model of neglect and attentional dyslexias. In Connectionist approaches to natural language processing (Eds. R.G. Reilly and N. E. Sharkey), Lawrence Erlbaum, Hove, UK.

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

Page, M. (2000). Connectionist modelling in psychology: A localist manifesto. Behavioural and Brain Sciences, 23, 443-500.

Peressotti, F. & Grainger, J. (1995). Letter position coding in random consonant arrays. Perception and Psychophysics, 37, 875–890.

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.

Price, C.J. & Humphreys, G.W. (1993). Attentional dyslexia: the effect of co-occurring deficits. Cognitive Neuropsychology, 10, 569–592.

Rastle, K., Harrington, J., Coltheart, M. & Palethorpe, S. (2000). Reading aloud begins when the computation of phonology is complete. Journal of Experimental Psychology: Human Perception and Performance, 26, 1178–1191.

Rayner, K. (1979). Eye guidance in reading: Fixation locations within words. Perception, 8, 21–30.

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

Rumelhart, D.E. & McClelland, J.L. (1981). An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effect and some tests and extensions of the model. Psychological Review, 89, 60–94.

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

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

Sergent, J. (1987). A new look at the human split brain. Brain, 110, 1375–1392.

Shallice, T. and Warrington, E. K. (1977). The possible role of selective attention in acquired dyslexia. Neuropsychologia, 15, 31–41.

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

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

Shillcock, R.C., Kirby, S., McDonald, S. & Brew, C. (submitted). Exploring systematicity in the mental lexicon.

Shillcock, R.C. & Monaghan, P. (2001). The computational exploration of visual word recognition in a split model. Neural Computation, 13, 1171–1198.

Shillcock, R.C. & Monaghan, P. (submitted). The split-fovea model of reading: Part 1. A connectionist implementation.

Shillcock, R.C. & Monaghan, P. (submitted). The split-fovea model of reading: Part 2. Connectionist modelling of normal reading.

Spieler D. H., & Balota, D. A. (1997). Bringing computational models of word naming down to the item level. Psychological Science, 6, 411-416.

Stein, J.F. & Talcott, J. (1999). Impaired neuronal timing in developmental dyslexia – the magnocellular hypothesis. Dyslexia, 5, 59–77.

Stein, J.F. & Walsh, V. (1997). To see but not to read; the magnocellular theory of dyslexia. Trends in Neuroscience, 20, 147-152.

Warrington, E.K., Cipolotti, L. & McNeil, J. (1993). Attentional dyslexia: A single case study. Neuropsychologia, 31, 871–885.

C.S. Whitney (2001a). Position-specific effects within the SERIOL framework for letter-position coding. Connection Science, 13, 235–255.

Whitney, C. (2001b). How the brain encodes the order of letters in a printed word: The SERIOL Model and selective literature review. Psychonomic Bulletin & Review, 8, 221–243.

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

Acknowledgements

The research was supported by Wellcome Trust grant 059080. The first author was supported by ESRC fellowship R/000/27/1244. The second author is employed on a Human Frontiers Science Program grant at the University of Warwick.

Author Note

Richard Shillcock, Department of Psychology and Institute for Adaptive and Neural Computation, Division of Informatics, University of Edinburgh.

Padraic Monaghan, Department of Psychology, University of Warwick,

Correspondence concerning this article should be addressed to Dr. Richard Shillcock, Department of Psychology, 7 George Square, Edinburgh, EH8 9JZ, or to the authors at rcs@inf.ed.ac.uk..

List of Figures

.

Figure 1.

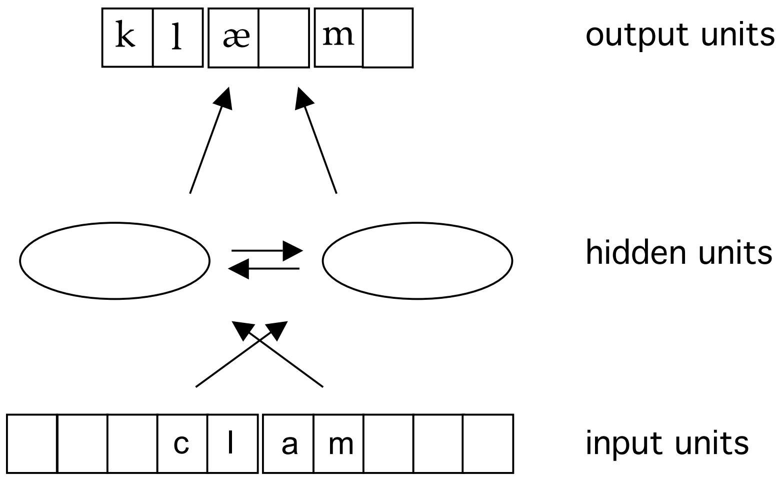


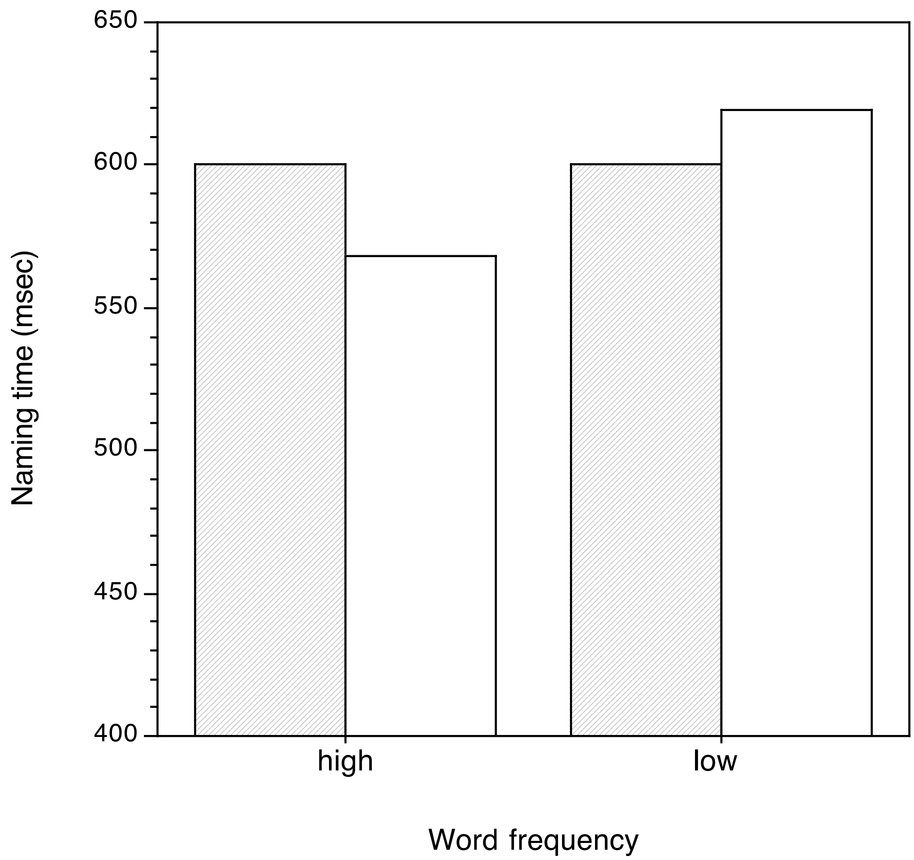
Figure 2.

	c	l	a	m
		c	l	a
			c	l
				c

m				
a	m			
l	a	m		
c	l	a	m	

Figure 3.

(a)


650
600
550
500
450
400
Naming time (msec)
high
low
Word frequency

(b)