

## Orthographic Neighbourhood Effects in Parallel Distributed Processing Models

CHRISTOPHER R. SEARS, *University of Calgary*  
YASUSHI HINO, *Chukyo University*  
STEPHEN J. LUPKER, *University of Western Ontario*

**Abstract** Recent research in visual word recognition suggests that the speed with which a word is identified is influenced by the reader's knowledge of other, orthographically similar words (Andrews, 1997). In serial-search and activation-based models of word recognition, mental representations of these "orthographic neighbours" of a word are explicitly assumed to play a role in the lexical selection process. Thus, it has been possible to determine the specific predictions that these models make about the effects of orthographic neighbours and to test a number of those predictions empirically. In contrast, the role of orthographic neighbours in parallel distributed processing models (e.g., Plaut, McClelland, Seidenberg, & Patterson, 1996; Seidenberg & McClelland, 1989) is less clear. In this paper, several statistical analyses of error scores from these types of models revealed that low frequency words with large neighbourhoods had lower orthographic, phonological, and cross-entropy error scores than low frequency words with small neighbourhoods; and that low frequency words with higher frequency neighbours had lower error scores than low frequency words without higher frequency neighbours. According to these models then, processing should be more rapid for low frequency words with large neighbourhoods and for low frequency words with higher frequency neighbours.

A word's orthographic neighbourhood is classically defined as the set of words that can be created by changing one letter of the word while preserving letter positions (Coltheart, Davelaar, Jonasson, & Besner, 1977). For example, the words PINE, POLE, and TILE are all orthographic neighbours of the word PILE. In recent years, there have been a number of studies examining the effects of a word's orthographic neighbourhood on identification latencies (see Andrews, 1997, for a review), and a considerable, although sometimes contradictory, database on this topic has now emerged.

Many models of the word recognition process do assume that the lexical representations of the orthographic neighbours of a presented word will be activated and will play an

important role in the lexical selection process. In what follows, we first examine the predictions of serial-search models (Forster, 1976; Paap, Newsome, McDonald, & Schvaneveldt, 1982) and activation-based models (Grainger & Jacobs, 1996; McClelland & Rumelhart, 1981) with regard to orthographic neighbourhood effects. We then consider the role of orthographic neighbours in parallel distributed processing models (i.e., Plaut, McClelland, Seidenberg, & Patterson, 1996; Seidenberg & McClelland, 1989), which constitute the main focus of the present investigation.

### ORTHOGRAPHIC NEIGHBOURHOOD EFFECTS IN SERIAL-SEARCH MODELS

In serial-search models which incorporate a frequency-ordered search through a candidate set of lexical entries (e.g., Forster, 1976; Paap, Newsome, McDonald, & Schvaneveldt, 1982), the size of a word's orthographic neighbourhood will influence the speed with which a correct match is found. More specifically, because a target word's orthographic neighbours will typically be members of an activated candidate set (due to their similarity to the target), increases in the number of neighbours will typically lead to increases in the size of the candidate set, which will in turn produce increases in the time required for lexical selection. According to serial search models then, words with large neighbourhoods should typically be processed more slowly than words with small neighbourhoods (such an effect can be referred to as "an inhibitory neighbourhood size effect").

Because the search through the candidate set is frequency-ordered in these models, however, it is actually not the absolute neighbourhood size of a word that is critical, but the number of higher frequency neighbours in the word's orthographic neighbourhood. That is, only higher frequency neighbours would delay lexical selection, because only those candidates would have to be evaluated prior to the word itself during the frequency-ordered search for the target's lexical representation. Consequently, although large neighbourhoods would, typically, delay lexical selection (because words with large neighbourhoods usually possess higher frequency neighbours), it is not the existence of

neighbours, per se, but rather the existence of higher frequency neighbours that produces a processing delay. Thus, a basic prediction that serial search models make is that words with higher frequency neighbours should be processed more slowly than words without higher frequency neighbours (such an effect is often referred to as "an inhibitory neighbourhood frequency effect").

The literature to date has provided only minimal support for these predictions. Andrews (1989, 1992), for example, found that lexical decision and naming latencies for low frequency words with large neighbourhoods were shorter than those for low frequency words with small neighbourhoods, a result which is exactly the opposite of the inhibitory neighbourhood size effect predicted by serial-search models. (For high frequency words, neighbourhood size had little or no effect on response latencies. Thus, there is typically a neighbourhood size by frequency interaction.) Facilitatory neighbourhood size effects have also been reported by Forster and Shen (1996) and Sears, Hino, and Lupker (1995), with the latter investigators also reporting an interaction between word frequency and neighbourhood size. In fact, in a recent review of the existing literature, Andrews (1997) noted that virtually all of the studies which have examined the neighbourhood size effect with the lexical decision task have reported either facilitatory or null neighbourhood size effects, a situation which is clearly problematic for serial-search models.

On the other hand, there is at least some evidence for the existence of an inhibitory neighbourhood frequency effect (Carreiras, Perea, & Grainger, 1997; Grainger, 1990; Grainger & Jacobs, 1996; Grainger, O'Regan, Jacobs, & Segui, 1989; Grainger & Segui, 1990; Huntsman & Lima, 1996; Jacobs & Grainger, 1992; Perea & Pollatsek, 1998). That is, all these studies seem to show that lexical decision latencies to low frequency words with higher frequency neighbours are slower than those to low frequency words without higher frequency neighbours. Further, using a multiple regression analysis, Paap and Johansen (1994) have reached a similar conclusion (although see Sears, Lupker, & Hino, in press, for an alternative explanation for this finding).

The story is not so simple, however, because it is complicated by the fact that the inhibitory neighbourhood frequency effect is typically not observed in studies which use English stimuli (the majority of studies have used either French, Dutch, or Spanish stimuli). Indeed, as Andrews (1997) noted in her review, only two of the eight experiments that have examined the effect of neighbourhood frequency for English words in the lexical decision task have reported an inhibitory effect (Huntsman & Lima, 1996; Perea & Pollatsek, 1998). In the remaining experiments, null or facilitatory neighbourhood frequency effects were reported (Forster & Shen, 1996; Sears et al., 1995). For example, in the Sears et al. study, in which neighbourhood

size and neighbourhood frequency were factorially manipulated, responses to words with higher frequency neighbours were actually faster than responses to words without higher frequency neighbours. The lack of a clear and consistent inhibitory effect of higher frequency neighbours coupled with the consistent finding of a facilitatory neighbourhood size effect would appear to cause severe problems for serial-search models.

#### ORTHOGRAPHIC NEIGHBOURHOOD EFFECTS IN ACTIVATION-BASED MODELS

Although activation-based models of word recognition have fared a bit better, the situation is similarly complicated. The interactive-activation model (McClelland & Rumelhart, 1981) would seem to readily accommodate reports of facilitatory neighbourhood size effects (Andrews, 1989, 1992; Forster & Shen, 1996; Sears et al., 1995) because the orthographic neighbours of a word are assumed to contribute in a positive way to the activation of the word's lexical unit. More specifically, in this model, lexical selection is achieved when a word's lexical unit reaches a critical activation threshold. When a word is presented, activation starts to accumulate in the lexical units of both the presented word and its orthographic neighbours. These partially activated units send excitatory feedback back down to their sublexical units. In turn these units send activation back up to the lexical units, increasing lexical activation and, ultimately, helping to push the activation of one of those units over threshold.

According to Andrews (1989), everything else being equal, low frequency words with large neighbourhoods would benefit more from reciprocal activation than would low frequency words with small neighbourhoods, because a greater number of lexical units would participate in the reciprocal activation process. Thus, low frequency words should show a facilitatory neighbourhood size effect. In contrast, high frequency words, which are assumed to have higher resting activation levels than low frequency words, would be less sensitive to the effects of these lexical-sublexical reverberations, because they could reach an activation threshold quite quickly through direct activation alone. Thus, high frequency words should show no neighbourhood size effects as Andrews and others (e.g., Sears et al., 1995) have reported.

Reports of facilitatory neighbourhood frequency effects for low frequency words (e.g., Sears et al., 1995) could, in theory, also be explained by the same mechanism. That is, higher frequency neighbours, which possess higher resting levels of activation, could produce stronger top-down activation, which would accelerate the reciprocal activation process. On the other hand, Grainger and colleagues have instead argued that the interactive-activation model is ideally suited for explaining inhibitory neighbourhood frequency effects. According to Jacobs and Grainger (1992), the

intralevel inhibition between the lexical units of the model should delay the activation of a word with higher frequency neighbours. More specifically, when a neighbourhood is activated by a target word, each lexical unit begins to inhibit its neighbours. Because higher frequency neighbours have high resting levels of activation, they would be much more powerful inhibitors than lower frequency neighbours. Consequently, the lexical unit of a word with higher frequency neighbours would be subject to more inhibition, which should delay lexical selection. Simulations by Jacobs and Grainger indicate that their implementation of the model does in fact produce such an inhibitory neighbourhood frequency effect. Interestingly, using the same parameter settings, Jacobs and Grainger's attempts to simulate facilitatory neighbourhood size effects were unsuccessful. This result led Jacobs and Grainger to the conclusion that the neighbourhood size effect did not reflect the activity of basic word recognition processes.

In a further attempt to address these issues, more recently, Grainger and Jacobs (1996) have proposed an activation-based model which can apparently accommodate both facilitatory neighbourhood size effects and inhibitory (as well as facilitatory) neighbourhood frequency effects in lexical decision tasks. Grainger and Jacobs's "multiple read-out" model is based on the architecture of the interactive-activation model (McClelland & Rumelhart, 1981), in which a set of lexical and sublexical units accumulates activation over time. The major assumption in the model is that facilitatory neighbourhood size effects in lexical decision do not actually arise during the lexical-selection process, but, rather, are due to a variable response criterion which is sensitive to the degree of overall lexical activation (the  $\Sigma$  criterion). In contrast, Grainger and Jacobs have maintained the assumption that the inhibitory neighbourhood frequency effect is a true lexical selection effect, resulting from intralevel competitive processes which occur during the process of lexical selection. As reported in their paper, with these two mechanisms and certain assumptions about how the nature of the nonwords affects relative use of these mechanisms, the model can be made to simulate both facilitatory neighbourhood size effects (e.g., Andrews, 1989); and inhibitory (e.g., Grainger et al., 1989), as well as facilitatory (e.g., Sears et al., 1995), neighbourhood frequency effects in lexical decision.

Unfortunately, the Grainger and Jacobs (1996) model still has some difficulties because it predicts that an inhibitory neighbourhood frequency effect should occur not just in lexical decision but in any task in which unique word identification is required, such as semantic categorization or perceptual identification. Neither Forster and Shen (1996) nor Sears, Lupker, and Hino (in press) observed such an effect in their semantic categorization experiments. In addition, Sears et al. reported that words with higher frequency neighbours were identified more frequently than

words without higher frequency neighbours in a perceptual identification task (i.e., they observed a facilitatory neighbourhood frequency effect). Thus, this model's ability to accurately simulate neighbourhood effects in tasks other than lexical decision appears to be somewhat limited.

Clearly, the fact that investigators have not yet established the empirical role of higher frequency neighbours makes it difficult to judge which particular model best accounts for the data. The situation is made worse still by the fact that these models do not make as unambiguous predictions as originally thought. For example, the interactive-activation model can accommodate facilitatory neighbourhood size effects or inhibitory neighbourhood frequency effects, depending on whether top-down excitatory feedback or intralevel inhibition is assumed to dominate the model's behaviour (i.e., depending on how the parameter settings are selected). The predictions of serial-search models can be just as ambiguous. Forster (1989), for example, has suggested that by altering some noncrucial assumptions in his version of the serial-search model, it would no longer predict inhibitory neighbourhood size or inhibitory neighbourhood frequency effects. Nonetheless, due to the efforts of previous researchers, our knowledge about the constraints these models must work within when trying to account for orthographic neighbourhood effects has been significantly advanced.

#### ORTHOGRAPHIC NEIGHBOURHOOD EFFECTS IN PARALLEL DISTRIBUTED PROCESSING MODELS

In contrast to the efforts that have been put into understanding how serial-search and activation-based models would account for orthographic neighbourhood effects, those same effects in Seidenberg and McClelland's (1989) parallel distributed processing (PDP) model have received relatively little attention. In this model, there are no abstract units corresponding to words. The representation of a word is encoded in the pattern of activity across an interconnected network of units. Experience with words during training produces changes in the weights between units, such that words which have been presented to the model many times will be better represented in the weights of the model.

To relate these patterns of activation to lexical decision and pronunciation latencies, Seidenberg and McClelland (1989) computed orthographic and phonological error scores, which are measures of how close the model's output is to the desired (correct) output. According to the model, lower orthographic error scores should correspond to shorter lexical decision latencies, and lower phonological error scores should correspond to shorter pronunciation latencies. Orthographic and phonological error scores are, of course, strongly influenced by the model's experience with words. For example, word frequency effects arise because the network is exposed to high frequency words much more often than low frequency words, and thus the

model has more opportunities to encode their orthography and phonology. As a result, the model produces lower orthographic and phonological error scores for high frequency words.

With regard to the issue of neighbourhood size, Seidenberg and McClelland's (1989) model would seem to predict a facilitatory neighbourhood size effect. More specifically, Seidenberg and McClelland reported that the mean phonological error scores for Andrews' (1989) low frequency words with large neighbourhoods were lower than those for her low frequency words with small neighbourhoods. Moreover, for high frequency words no such difference was apparent, which suggests that the model can successfully simulate the interaction between word frequency and neighbourhood size that Andrews reported (see also Andrews, 1992, and Sears et al., 1995). In fact, according to Andrews (1992), facilitatory neighbourhood size effects are a natural byproduct of this model. That is, words that are highly similar to one another would recruit similar units and connections during training, and so the representation of a word with many neighbours would be strengthened by the encoding of its neighbours. Thus, compared to words with small neighbourhoods, which would share connections with few other words, words with large neighbourhoods should exhibit lower phonological and orthographic error scores.

Sears et al. (1995) suggested that reports of facilitatory neighbourhood frequency effects could be explained by the model in a similar fashion. That is, low frequency words would benefit from the existence of higher frequency neighbours because these neighbours would be words whose representations have been encoded by the network many times. The strengthened connections between the units that encode the word's higher frequency neighbours will aid in a low frequency word's identification as well because many of the same units will be recruited by the word itself. Thus, large neighbourhoods and higher frequency neighbours should affect the model in a similar manner — by strengthening the connections among units that represent similar orthographies. In fact, Sears et al. found that the phonological error scores for their low frequency words with higher frequency neighbours were generally lower than those for low frequency words with no higher frequency neighbours.

In spite of these initial analyses, at this point these expectations about the model's behaviour are mainly still speculations because no comprehensive statistical analysis of orthographic neighbourhood effects in the Seidenberg and McClelland (1989) model has been conducted. That is, although previous investigators have reported that phonological error scores for words with large neighbourhoods were generally lower than for words with small neighbourhoods, the generalizability of these observations is limited. This is because the observations were based upon the patterns of error scores for the small sets of stimuli used in those

particular experiments. Moreover, with the exception of the Sears et al. (1995) study, the effects of higher frequency neighbours in the model have not been examined. Consequently, unlike the serial-search and activation-based models, where more specific predictions have been made, the predictions of the Seidenberg and McClelland model are much less clear. In light of the controversy over the effects of orthographic neighbours on the word recognition process and which of these types of models provides the superior account of that process, it would seem important to establish what the predictions of this model are.

Several researchers have, of course, shown that the Seidenberg and McClelland (1989) model has a number of serious problems. In particular, the model has difficulty accurately pronouncing nonwords and certain exception words, and in explaining lexical decision performance in general (Besner, Twilley, McCann, & Seergobin, 1990; Coltheart, Curtis, Atkins, & Haller, 1993; Fera & Besner, 1992). A more recent implementation of the model (Plaut, McClelland, Seidenberg, and Patterson, 1996; Simulation 4) is, however, able to pronounce nonwords as well as skilled readers can. Moreover, the model goes some way towards implementing the lexical-semantic pathway that the Seidenberg and McClelland simulation omitted. Like the Seidenberg and McClelland model, the Plaut et al. (1996) model is a feed-forward network, and an error score — in this case, cross-entropy error — measures how close the model's output is to the correct pronunciation, with lower cross-entropy errors, presumably, corresponding to shorter pronunciation latencies. To our knowledge, the effects of orthographic neighbours in this model have not been examined at all. (Note that the Plaut et al. model does not simulate lexical decision performance, although preliminary efforts to do so have been made; Plaut, 1997. In what follows, we assume that the effects of orthographic neighbours on both naming and lexical decision performance will be quite similar in these types of models. As will be seen, this clearly does turn out to be the case for the Seidenberg & McClelland model.) Thus, the purpose of this investigation was to determine what effect, if any, orthographic neighbours have in both of these models. By pursuing this goal, the models' success in accommodating current and future findings can be better evaluated.

## Method

### STIMULI

The training set for the Seidenberg and McClelland (1989) model consisted of 2,897 monosyllabic words of three or more letters in length. Because most of the previous studies on orthographic neighbourhood effects have used stimuli of four or five letters in length (e.g., Andrews, 1989, 1992; Forster & Shen, 1996; Grainger et al., 1989; Sears et al., 1995), only the error scores for four- and five-letter words

were examined in the following analyses.<sup>1</sup> In addition, because a logarithmic transformation of Kucera and Francis's (1967) normative frequencies was employed in the regression analyses, words with normative frequencies of zero were excluded from the stimulus set. The final set of stimuli used in the present analyses consisted of 2,073 words. For each of these words, the Kucera and Francis normative frequency, the number of orthographic neighbours, and the number of higher frequency neighbours were determined. The statistical properties (mean and range) for each of these variables were as follows: normative frequency (91.4, 1-10,595); number of neighbours (7.03, 0-24); number of higher frequency neighbours (2.61, 0-21).

The training set for the Plaut et al. (1996) model (Simulation 4) consisted of the 2,897 words in the Seidenberg and McClelland model's corpus plus an additional 101 words.<sup>2</sup> To allow direct comparisons between this model and the Seidenberg and McClelland (1989) model, the same 2,073 words selected from the Seidenberg and McClelland corpus were also used in the analyses of the Plaut et al. model.

### Results

#### EFFECTS OF NEIGHBOURHOOD SIZE

In the first analysis, the mean orthographic and phonological error scores from Seidenberg and McClelland's (1989) model for words with large and small neighbourhoods were examined. Consistent with most of the previous literature, in this and the related analyses (although not in the multiple regression analyses), words with less than five neighbours were classified as small neighbourhood words and words with five or more neighbours were classified as large neighbourhood words. Using these criteria, 951 of the words had small neighbourhoods, and 1,122 had large neighbourhoods. The mean orthographic error score for the words with large neighbourhoods (6.49) was significantly lower than the mean orthographic error score for the words with small neighbourhoods (9.82),  $t(2071) = 16.77$ ,  $SE = 0.19$ . (Unless otherwise stated, the  $p$  values for all significant statistics reported in the text are less than .05.) Similarly, words with large neighbourhoods had, on average, lower phonological error scores than words with small neighbourhoods (4.37 versus 5.65),  $t(2071) = 8.66$ ,  $SE = 0.14$ .

Cross-entropy error scores from the Plaut et al. (1996) simulation were submitted to the identical analysis. As was the case with the orthographic and phonological error scores, the mean error score for the words with large neighbourhoods (0.0482) was significantly lower than the mean error score for the words with small neighbourhoods (0.0723),  $t(2071) = 8.95$ ,  $SE = 0.003$ .

<sup>1</sup> We thank Mark Seidenberg and James McClelland for making these data available to us.

<sup>2</sup> We are grateful to David Plaut for making these data available to us.

TABLE 1

Mean Orthographic and Phonological Error Scores From the Seidenberg and McClelland (1989) Model, and Mean Cross-entropy Error Scores From the Plaut et al. (1996) Model, for High Frequency and Low Frequency Words With Small and Large Neighbourhoods

Error score	High frequency words		Low frequency words	
	Small N	Large N	Small N	Large N
Orthographic	4.52	4.31	11.03	7.20
Phonological	3.17	3.48	6.23	4.67
Cross-entropy	0.0264 (119)	0.0258 (181)	0.0836 (758)	0.0558 (827)

Note. Small N = small neighbourhood. Large N = large neighbourhood. Number of cases per group in parentheses.

As previously noted, several investigators have reported interactions between word frequency and neighbourhood size. In Andrews' (1989) experiments, for example, lexical decision and naming latencies for low frequency words with large neighbourhoods were shorter than those for low frequency words with small neighbourhoods. For high frequency words, however, large neighbourhoods had little or no effect on response latencies. Consequently, it is of some interest to determine whether an analogous pattern of data would be found in the Seidenberg and McClelland (1989) and Plaut et al. (1996) simulations' error scores. To this end, the error scores for high frequency and low frequency words with large and small neighbourhoods were submitted to a 2 (Word Frequency: high versus low)  $\times$  2 (Neighbourhood Size: small versus large) analysis of variance. For the purposes of this analysis, words with normative frequencies greater than or equal to 100 were considered high frequency words, and words with normative frequencies less than or equal to 50 were considered low frequency words. Words with frequencies between 51 and 99 (inclusive) were not used in this analysis. (The mean neighbourhood sizes for the low frequency words with small neighbourhoods and the low frequency words with large neighbourhoods were 2.89 and 11.05, respectively. For the high frequency words, the mean neighbourhood sizes were 2.92 and 10.32 for the small neighbourhood and the large neighbourhood words, respectively.) The mean error scores for these stimuli are listed in Table 1.

In the analysis of Seidenberg and McClelland's (1989) orthographic error scores, there was a main effect of word frequency,  $F(1, 1881) = 296.99$ ,  $MSE = 18.11$ , a main effect of neighbourhood size,  $F(1, 1881) = 54.77$ ,  $MSE = 18.11$ , and an interaction between word frequency and neighbourhood size,  $F(1, 1881) = 43.96$ ,  $MSE = 18.11$ . For the phonological error scores, there was also a main effect of word frequency,  $F(1, 1881) = 100.86$ ,  $MSE = 10.88$ , a main effect of neighbourhood size,  $F(1, 1881) = 8.73$ ,  $MSE = 10.88$ , and a significant interaction,  $F(1, 1881) = 19.80$ ,  $MSE = 10.88$ . For the high frequency words, the mean orthographic error scores for the small neighbourhood and large neighbourhood words were

quite similar and not significantly different,  $t(298) = 1.42$ ,  $SE = 0.14$ , as were the mean phonological error scores,  $t(298) = 1.19$ ,  $SE = 0.26$ . In contrast, low frequency words with large neighbourhoods had significantly lower orthographic error scores than low frequency words with small neighbourhoods,  $t(1583) = 16.53$ ,  $SE = 0.23$ . This was true for the phonological error scores as well,  $t(1583) = 9.02$ ,  $SE = 0.17$ .

An analysis of Plaut et al.'s (1996) cross-entropy error scores revealed an identical pattern of results. There was a main effect of word frequency,  $F(1, 1881) = 125.00$ ,  $MSE = 0.004$ , a main effect of neighbourhood size,  $F(1, 1881) = 13.18$ ,  $MSE = 0.004$ , and a significant interaction,  $F(1, 1881) = 12.05$ ,  $MSE = 0.004$ . For high frequency words, there was no significant difference in the mean cross-entropy errors for the words with small and large neighbourhoods,  $t(298) = .29$ ,  $SE = 0.002$ . However, low frequency words with large neighbourhoods had lower cross-entropy error scores than low frequency words with small neighbourhoods,  $t(1583) = 8.37$ ,  $SE = 0.003$ . Thus, the Plaut et al. model, like the Seidenberg and McClelland (1989) model, captures the interaction between word frequency and neighbourhood size reported in the literature. That is, both models predict that low frequency words with large neighbourhoods should be processed faster than low frequency words with small neighbourhoods; however, there should be little evidence of a neighbourhood size effect for high frequency words.

#### EFFECTS OF HIGHER FREQUENCY NEIGHBOURS

To evaluate the effects of higher frequency neighbours in a word's orthographic neighbourhood, error scores for words with and without higher frequency neighbours were examined. Because the existence of higher frequency neighbours is correlated with word frequency (i.e., low frequency words are more likely to have a higher frequency neighbour), separate analyses of the low frequency and high frequency words were conducted. For the purposes of this analysis, words with normative frequencies greater than or equal to 100 were considered high frequency words, and words with normative frequencies less than or equal to 50 were considered low frequency words. Words with frequencies between 51 and 99 (inclusive) were not used in this analysis.

As shown in Table 2, the mean Seidenberg and McClelland (1989) orthographic error score for the low frequency words with higher frequency neighbours was substantially lower than the mean orthographic error score for the low frequency words with no higher frequency neighbours,  $t(1583) = 7.87$ ,  $SE = 0.32$ . Phonological error scores were also lower for low frequency words with higher frequency neighbours,  $t(1583) = 3.98$ ,  $SE = 0.23$ , as were Plaut et al.'s (1996) cross-entropy error scores,  $t(1583) = 3.05$ ,  $SE = 0.004$ .

A less consistent pattern of results emerged in the analysis of the high frequency words. The mean Seidenberg and

TABLE 2

Mean Orthographic and Phonological Error Scores From the Seidenberg and McClelland (1989) Model, and Mean Cross-entropy Error Scores From the Plaut et al. (1996) Model, for High Frequency and Low Frequency Words With and Without Higher Frequency Neighbours

Error score	High frequency words		Low frequency words	
	No HF N	HF N	No HF N	HF N
Orthographic	4.49	4.26	11.17	8.60
Phonological	3.13	3.68	6.19	5.26
Cross-entropy	0.0249 (175)	0.0276 (125)	0.0804 (271)	0.0668 (1314)

Note. No HF N = no higher frequency neighbours; HF N = higher frequency neighbours. Number of cases per group in parentheses.

McClelland (1989) orthographic error scores for high frequency words with and without higher frequency neighbours were not significantly different,  $t(298) = 1.55$ ,  $SE = 1.47$ , nor were the mean Plaut et al. (1996) cross-entropy error scores,  $t(298) = 1.26$ ,  $SE = .002$ . However, high frequency words with higher frequency neighbours had significantly higher Seidenberg and McClelland phonological error scores than high frequency words without higher frequency neighbours,  $t(298) = 2.07$ ,  $SE = .26$ .

With regard to the processing of low frequency words, the basic conclusion that these results suggest is that, according to the models, the presence of higher frequency neighbours in a low frequency word's orthographic neighbourhood should actually be beneficial to processing. Thus, both models will have great difficulty accommodating the inhibitory neighbourhood frequency effects reported by Grainger and colleagues (e.g., Grainger, 1990), because in those studies responses to low frequency words with higher frequency neighbours were slower than the responses to low frequency words without higher frequency neighbours. For the high frequency words the interpretation is not as straightforward, but we will defer any discussion of these findings until after the multiple regression analyses have been presented.<sup>3</sup>

<sup>3</sup> We also examined the data from Plaut et al.'s (1996) Simulation 3 which, unlike Seidenberg and McClelland's (1989) simulation and Plaut et al.'s (1996) Simulation 4, was a fully interactive recurrent network that generated settling times for items rather than error scores. In this simulation, after a word is presented to the network the unit activations eventually settle into a stable pattern representing the phonological output, with the settling times intended to be directly analogous to pronunciation latency. An analysis of these settling times revealed that low frequency words with large neighbourhoods had significantly faster settling times than low frequency words with small neighbourhoods (1.74 versus 1.80),  $t(1566) = 4.79$ ,  $SE = 0.01$ , and that low frequency words with higher frequency neighbours had significantly faster settling times than low frequency words without higher frequency neighbours (1.76 versus 1.81),  $t(1566) = 3.35$ ,  $SE = 0.01$ . Neither neighbourhood size nor neighbourhood frequency had any effect on the settling times for high frequency words. (Note that the items to which the model made an

## MULTIPLE REGRESSION ANALYSES

According to the previous analyses, the Seidenberg and McClelland (1989) and Plaut et al. (1996) models both predict that, for low frequency words, large neighbourhoods and higher frequency neighbours should facilitate word identification and naming. However, because word frequency, neighbourhood size, and neighbourhood frequency are all correlated with one another to varying degrees, converging evidence on these issues should also be obtained through the use of multiple regression analyses. To this end, multiple regression analyses were conducted for each of the three types of error scores.

In the first analysis, the entire data set of 2,073 words was analyzed. The predictor variables were log word frequency, the number of orthographic neighbours, and the number of higher frequency neighbours (the predictor variables were entered simultaneously).<sup>4</sup> Partial correlation coefficients were computed to assess the unique correlation between the models' error scores and each of the predictor variables, and are listed in Table 3. In the multiple regression analysis of Seidenberg and McClelland's (1989) orthographic error scores, 44.6% of the variance was explained by these three variables,  $F(3, 2069) = 556.60$ ,  $MSE = 12.73$ . There were significant negative partial correlations for word frequency, the number of neighbours, and the number of higher frequency neighbours. Specifically, a larger number of neighbours, the existence of higher frequency neighbours, and higher word frequency were associated with lower orthographic error scores when the effects of the other two variables were partialled out.

For Seidenberg and McClelland's (1989) phonological error scores, 18.4% of the variance was explained by these variables,  $F(3, 2069) = 155.76$ ,  $MSE = 9.42$ . Once again, there were significant negative partial correlations between the phonological error scores and word frequency, the number of neighbours, and the number of higher frequency neighbours. Although the magnitude of the partial correlations was smaller than those in the orthographic error scores analysis, the pattern of results was identical. In particular, a larger number of neighbours, the existence of higher fre-

"error response" were not included in any of these analyses.)

<sup>4</sup> Preliminary analyses revealed that word length (four or five letters) was not a significant predictor in any of the regression analyses of the phonological error scores or the cross-entropy error scores. For the orthographic error scores, the partial correlations for word length were statistically significant (and negative), but the increment in the percentage of variance explained by this variable (above and beyond that explained by log word frequency, the number of neighbours, and the number of higher frequency neighbours) was very small (less than 0.5%), which may have been due to the restricted range of this variable (i.e., all the stimuli were either four or five letters in length). For these reasons word length was not used as a predictor variable in any of the regression analyses.

TABLE 3

Partial Correlation Coefficients for the Orthographic and Phonological Error Scores From the Seidenberg and McClelland (1989) Model, and for the Cross-entropy Error Scores From the Plaut et al. (1996) Model

Predictor variable	Error score		
	Orthographic	Phonological	Cross-entropy
Low frequency and high frequency words			
Log WF	-.57	-.35	-.41
N	-.12	-.06	-.08
N HF N	-.22	-.10	-.05
Low frequency words			
Log WF	-.51	-.34	-.41
N	-.13	-.06	-.08
N HF N	-.17	-.10	-.04
High frequency words			
Log WF	-.29	-.01	-.07
N	-.01	-.13	-.15
N HF N	-.14	+.39	+.21

Note. Log WF = Log word frequency. N = number of neighbours. N HF N = number of higher frequency neighbours.

quency neighbours, and higher word frequency were associated with significantly lower phonological error scores. For Plaut et al.'s (1996) cross-entropy error scores, 24.6% of the variance was explained by these variables,  $F(3, 2069) = 226.02$ ,  $MSE = 0.002$ . Again, there were significant negative partial correlations between these error scores and word frequency, the number of neighbours, and the number of higher frequency neighbours.

Because most investigators have focused on orthographic neighbourhood effects for low frequency words, separate regression analyses were conducted on this subset of the stimuli (i.e., words with normative frequencies less than or equal to 50). The analysis of Seidenberg and McClelland's (1989) orthographic error scores revealed that log word frequency, neighbourhood size, and the number of higher frequency neighbours accounted for 40.0% of the variance,  $F(3, 1581) = 351.78$ ,  $MSE = 14.94$ . The partial correlations are listed in Table 3. There were significant negative partial correlations for word frequency, the number of neighbours, and the number of higher frequency neighbours. Similarly, the partial correlations between the phonological error scores and these variables were all significant and negative. Together these variables accounted for 18.0% of the variance,  $F(3, 1581) = 116.01$ ,  $MSE = 10.32$ . In the analysis of Plaut et al.'s (1996) cross-entropy error scores, there were significant negative partial correlations for word frequency and the number of neighbours, but the partial correlation between cross-entropy error and the number of higher frequency neighbours (-.04) was not statistically significant ( $p = .11$ ). Together these variables accounted for 23.4% of the variance,  $F(3, 1581) = 161.82$ ,  $MSE = 0.003$ .

Finally, separate regression analyses were conducted on

the set of high frequency words (i.e., words with normative frequencies greater than or equal to 100). In the analysis of Seidenberg and McClelland's (1989) orthographic error scores, log word frequency, neighbourhood size, and the number of higher frequency neighbours accounted for 8.9% of the variance,  $F(3, 296) = 9.72$ ,  $MSE = 1.45$ . There were significant partial correlations for word frequency and for the number of higher frequency neighbours, but not for the number of neighbours. In contrast, for Seidenberg and McClelland's phonological error scores there was a significant negative partial correlation for neighbourhood size, a significant positive partial correlation for the number of higher frequency neighbours, but no significant partial correlation for word frequency. Together these variables accounted for 16.6% of the variance,  $F(3, 296) = 19.67$ ,  $MSE = 4.35$ . Similarly, for Plaut et al.'s (1996) cross-entropy error scores there was a significant negative partial correlation for neighbourhood size, a significant positive partial correlation for the number of higher frequency neighbours, but no significant partial correlation for word frequency. A total of 6.1% of the variance was explained by these variables,  $F(3, 296) = 6.47$ ,  $MSE = 0.0003$ .<sup>5</sup>

### Discussion

The important findings of this investigation are as follows. First, words with large neighbourhoods had lower orthographic, phonological, and cross-entropy error scores than words with small neighbourhoods. Importantly, only the low frequency words benefitted from the presence of a large neighbourhood, as the error scores for high frequency words with large and small neighbourhoods were not significantly different from one another. Consequently, as noted, both models capture the interaction between word frequency and neighbourhood size that Andrews (1989, 1992) and Sears et al. (1995) have reported.

Second, compared to low frequency words with no higher frequency neighbours, low frequency words with higher frequency neighbours had, on average, lower orthographic, phonological, and cross-entropy error scores. As noted, this result suggests that both models will have difficulties accommodating the inhibitory neighbourhood frequency effects reported by Grainger and colleagues,

although they would appear to be quite consistent with the facilitatory neighbourhood frequency effects reported by Sears et al. (1995, in press).

Third, the regression analyses indicated that, for low frequency words, the number of neighbours and the number of higher frequency neighbours were independently (negatively) correlated with both models' error scores. That is, when the effects of word frequency and the number of higher frequency neighbours were partialled out, larger neighbourhood size was associated with lower orthographic, phonological, and cross-entropy error scores. Similarly, when the effects of word frequency and the number of neighbours were partialled out, the existence of higher frequency neighbours was associated with lower error scores in both of the models. Consequently, both models predict that, for low frequency words, large neighbourhoods and higher frequency neighbours should facilitate word recognition and naming independently of one another.

The results for the high frequency words were slightly different. For high frequency words, the existence of higher frequency neighbours was associated with higher, not lower, phonological and cross-entropy error scores (although this was not the case for the orthographic error scores). According to the models then, pronunciation latencies to high frequency words with higher frequency neighbours should be slower than those to high frequency words without higher frequency neighbours. At present it is difficult to evaluate this prediction because there has been only one published experiment which has examined the effect of higher frequency neighbours for high frequency words in a pronunciation task (Sears et al., 1995; Experiment 2). In that experiment, pronunciation latencies to high frequency words with and without higher frequency neighbours were not significantly different from one another. While this result casts some doubt on the empirical validity of this particular prediction, additional studies will be necessary before any definitive conclusions can be reached.

It is worth noting, however, that the mean phonological error score for Sears et al.'s (1995) high frequency words with higher frequency neighbours (3.05) was lower, not higher, than the mean phonological error score for their high frequency words without higher frequency neighbours (3.15). This was true of the cross-entropy error scores as well (0.0222 versus 0.0241, respectively, for the words with and without higher frequency neighbours). Thus, the naming latencies to this particular sample of high frequency words do not provide a fair test of the models' predictions with regard to neighbourhood frequency effects for high frequency words.

Relatedly, an examination of the models' error scores in experiments that have reported conflicting neighbourhood effects may provide information useful for ascertaining the source of these discrepancies. Consider, for example, Perea and Pollatsek's (1998) Experiment 1, where lexical decision

<sup>5</sup> In the regression analysis of the settling times from Plaut et al.'s (1996) Simulation 3 (excluding "error responses"), for the entire data set, only the partial correlations for word frequency (-.15) and the number of neighbours (-.04,  $p = .05$ ) were statistically significant. Together these variables accounted for 4.2% of the variance,  $F(3, 2046) = 30.27$ ,  $MSE = 0.05$ . In a separate analysis of the low frequency words, word frequency was the only significant predictor (-.14), and accounted for 3.4% of the variance,  $F(3, 1564) = 18.37$ ,  $MSE = 0.05$ . In the analysis of the high frequency words, the partial correlations for the number of neighbours (-.13) and the number of higher frequency neighbours (.21) were statistically significant, but the partial correlation for word frequency was not. Together these variables accounted for 5.0% of the variance,  $F(3, 291) = 5.13$ ,  $MSE = 0.02$ .



latencies to low frequency words with higher frequency neighbours were slower than those to low frequency words without higher frequency neighbours (an inhibitory neighbourhood frequency effect), and Sears et al.'s (1995) Experiment 4a, where a facilitatory neighbourhood frequency was observed. An examination of the orthographic error scores for Perea and Pollatsek's stimuli revealed that their words with higher frequency neighbours had a higher mean orthographic error score (9.45) than their words without higher frequency neighbours (9.09), whereas Sears et al.'s words with higher frequency neighbours had a lower mean orthographic error score (7.32) than words without higher frequency neighbours (7.51). Thus, according to the Seidenberg and McClelland model, and consistent with the findings of these investigators, the neighbourhood frequency effect should have been inhibitory in the Perea and Pollatsek experiment, and facilitatory in the Sears et al. experiment, exactly as observed. (Note that the mean orthographic error scores used in these comparisons are based on a restricted set of words, as the Seidenberg and McClelland model was trained with 51% of the words from Perea and Pollatsek's experiment and 86% of the words from Sears et al.'s experiment.)

Another result of note is the consistency in the pattern of orthographic neighbourhood effects in both models. Although the Plaut et al. (1996) simulation is superior to the Seidenberg and McClelland (1989) simulation in several important ways (most notably its superior performance pronouncing nonwords), the effects of orthographic neighbours in the models is strikingly similar. No doubt this is due to the common principle embodied in the two models — low frequency words with many neighbours, or with higher frequency neighbours, will have their pattern of activity strengthened many times during training, which will facilitate their processing.

The implications of these findings for the two theories are fairly clear. Although there is currently some empirical controversy as to whether higher frequency orthographic neighbours facilitate or inhibit lexical processing, the Seidenberg and McClelland (1989) and the Plaut et al. (1996) models do make very specific predictions in this regard. In this respect, these models will have great difficulty accommodating the inhibitory neighbourhood frequency effects reported by Grainger and colleagues, because for low frequency words, the existence of higher frequency neighbours was associated with lower, not higher, orthographic, phonological, and cross-entropy error scores. On the other hand, these models would seem to be quite compatible with the facilitatory neighbourhood size effects reported by Andrews (1989, 1992), as well as the facilitatory neighbourhood frequency effects reported by Sears et al. (1995, in press). Clearly, any judgements about the models' ultimate success in accommodating orthographic neighbourhood effects must await a resolution of any empirical controver-

sies. In the meantime, investigators will now have a better understanding of what these models have to say about the effects of orthographic neighbours.

We thank Jennifer Chesson for creating the data sets that were analyzed in this study, and Theresa Kline for statistical advice. We also thank Sally Andrews and Ron Borowsky for their very helpful reviews. Correspondence concerning this article should be addressed to Christopher R. Sears, Department of Psychology, University of Calgary, 2500 University Drive, Calgary, Alberta T2N 1N4 (E-mail: sears@ucalgary.ca).

## References

- Andrews, S. (1989). Frequency and neighborhood effects on lexical access: Activation or search? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 802-814.
- Andrews, S. (1992). Frequency and neighborhood effects on lexical access: Lexical similarity or orthographic redundancy? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 234-254.
- Andrews, S. (1997). The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin & Review*, *4*, 439-461.
- Besner, D., Twilley, L., McCann, R. S., & Seergobin, K. (1990). On the association between connectionism and data: Are a few words necessary? *Psychological Review*, *97*, 432-446.
- Carreiras, M., Perea, M., & Grainger, J. (1997). Effects of orthographic neighborhood in visual word recognition: Cross-task comparisons. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 857-871.
- 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., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI* (pp. 535-555). New York: Academic Press.
- Fera, P., & Besner, D. (1992). The process of lexical decision: More words about a parallel distributed processing model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 749-764.
- Forster, K. I. (1976). Accessing the mental lexicon. In R. J. Wales & E. W. Walker (Eds.), *New approaches to language mechanisms* (pp. 257-287). Amsterdam: North Holland.
- Forster, K. I. (1989). Basic issues in lexical processing. In W. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 75-107). Cambridge, MA: MIT Press.
- Forster, K. I., & Shen, D. (1996). No enemies in the neighborhood: Absence of inhibitory neighbourhood effects in lexical decision and semantic categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 696-713.
- Grainger, J. (1990). Word frequency and neighborhood frequency effects in lexical decision and naming. *Journal of Memory and Language*, *29*, 228-244.

- Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review*, 103, 518-565.
- Grainger, J., O'Regan, J. K., Jacobs, A. M., & Segui, J. (1989). On the role of competing word units in visual word recognition: The neighborhood frequency effect. *Perception & Psychophysics*, 45, 189-195.
- Grainger, J., & Segui, J. (1990). Neighborhood frequency effects in visual word recognition: A comparison of lexical decision and masked identification latencies. *Perception & Psychophysics*, 47, 191-198.
- Huntsman, L. A., & Lima, S. D. (1996). Orthographic neighborhood structure and lexical access. *Journal of Psycholinguistic Research*, 25, 417-429.
- 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.
- Kucera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University 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.
- Paap, K. R., Newsome, S. L., McDonald, J. E., & Schvaneveldt, R. W. (1982). An activation verification model for letter and word recognition: The word superiority effect. *Psychological Review*, 89, 573-594.
- Paap, K. R., & Johansen, L. S. (1994). The case of the vanishing frequency effect: A retest of the verification model. *Journal of Experimental Psychology: Human Perception and Performance*, 20, 1129-1157.
- Perea, M., & Pollatsek, A. (1998). The effects of neighborhood frequency in reading and lexical decision. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 767-779.
- Plaut, D. C. (1997). Structure and function in the lexical system: Insights from distributed models of word reading and lexical decision. *Language and Cognitive Processes*, 12, 767-808.
- Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired reading: Computational principles in quasi-regular domains. *Psychological Review*, 103, 56-115.
- Sears, C. R., Hino, Y., & Lupker, S. J. (1995). Neighborhood size and neighbourhood frequency effects in word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 21, 876-900.
- Sears, C. R., Lupker, S. J., Hino, Y. (in press). Orthographic neighborhood effects in perceptual identification and semantic categorization tasks: A test of the Multiple Read-out model. *Perception & Psychophysics*.
- Seidenberg, M. S., & McClelland, J. L. (1989). A distributed developmental model of word recognition and naming. *Psychological Review*, 96, 523-568.

Date of Acceptance: March 23, 1999

## Sommaire

Les voisins orthographiques d'un mot sont les mots qui peuvent être créés en changeant une lettre sans modifier pour autant la position des autres lettres (Coltheart, Davelaar, Jonasson et Besner, 1977). Par exemple, les mots PINE, POLE et TILE sont tous des voisins orthographiques du mot PILE. Un certain nombre de recherches réalisées au cours des dernières années ont cherché à expliquer comment le temps d'attente précédant l'identification d'un mot varie en fonction des diverses caractéristiques des voisins orthographiques d'un mot (voir Andrews, 1997), et il existe aujourd'hui une considérable base de données, bien que contradictoires quelquefois, sur le sujet.

Beaucoup de modèles du processus de reconnaissance d'un mot (p. ex. les modèles de recherche-série, d'activation) présument de façon explicite que les représentations lexicales des voisins orthographiques d'un mot présenté seront générées et joueront un rôle déterminant dans le processus de sélection lexicale. Il est possible par conséquent d'identifier les prédictions spécifiques que ces modèles font à propos des incidences des voisins orthographiques, et de procéder à une analyse empirique d'un certain nombre d'entre elles. Par

opposition, les implications de ces mêmes incidences pour les modèles de traitement parallèle réparti de Seidenberg et McClelland (1989) et de Plaut, McClelland, Seidenberg et Patterson (1996) ont suscité relativement peu d'attention.

Plusieurs analyses statistiques des taux d'erreur associés à ces types de modèles font l'objet du présent document. Voici les principaux résultats obtenus. Premièrement, les mots possédant un vaste voisinage orthographique affichaient un taux d'erreur inférieur pour ce qui concerne l'orthographe, la phonologie et l'entropie réciproque que les mots au voisinage orthographique plus limité. Ces différences, toutefois, n'ont été observées que pour les mots à fréquence peu élevée. Aussi, les deux modèles reflétaient l'interaction entre la fréquence d'un mot et la taille du voisinage orthographique que Andrews (1989; 1992) puis Sears, Hino et Lupker (1995) ont signalée.

Deuxièmement, si on les compare à des mots dont les voisins orthographiques n'ont pas une fréquence élevée, les mots dont le contraire est vrai avaient, en moyenne, un taux d'erreur moindre pour ce qui concerne l'orthographe, la phonologie et l'entropie réciproque. Ce résultat suggère que

les deux modèles auront des difficultés à tenir compte des effets inhibiteurs associés à la fréquence du voisinage orthographique et qu'ont signalés Grainger et ses collègues (Grainger, 1990; Grainger et Jacobs, 1996; Grainger, O'Regan, Jacobs et Segui, 1989; Grainger et Segui, 1990; Jacobs et Grainger, 1992), bien que les modèles sont relativement en mesure de tenir compte des effets facilitateurs associés à la fréquence du voisinage orthographique qu'ont signalés Sears et ses collègues (1995) puis Sears, Lupker et Hino (sous presse).

Troisièmement, les analyses de régression ont démontré que, dans le cas des mots à fréquence peu élevée, il y avait une corrélation négative et indépendante entre, d'une part, le nombre de voisins en général et le nombre de voisins à fréquence plus élevée, d'autre part, les taux d'erreur des deux modèles. En d'autres mots, lorsqu'on élimine les effets de la

fréquence d'un mot et du nombre de voisins à fréquence plus élevée, un plus grand nombre de voisins affichait un faible taux d'erreur pour ce qui concerne l'orthographe, la phonologie et l'entropie réciproque. Par conséquent, les deux modèles prédisent que, règle générale, et les voisinages orthographiques nombreux et les voisins à fréquence plus élevée devraient engendrer un traitement plus rapide des mots à fréquence peu élevée dans pratiquement tous les exercices de reconnaissance d'un mot.

Bien sûr, tout jugement par rapport à la capacité fondamentale des modèles de tenir compte des effets du voisinage orthographique doit attendre la résolution des controverses empiriques courantes mais, d'ici là, les chercheurs auront une meilleure idée des prédictions de ces modèles relativement aux effets des voisins orthographiques.