

The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts

SALLY ANDREWS

University of New South Wales, Sydney, New South Wales, Australia

This paper reviews recent research on the effects of orthographic neighbors on visual word recognition in order to resolve apparently contradictory findings. The review reveals that the empirical evidence is not as contradictory as has been claimed. Neighbors have consistently been reported to facilitate responses to words in naming and lexical decision tasks. Inhibitory effects of neighbors appear to arise from sophisticated guessing strategies in the perceptual identification task or lexical decision strategies adopted in unusual stimulus environments. For English words, there is minimal evidence of competitive influences on lexical retrieval due to higher frequency neighbors. Such effects are more common in such languages as French and Spanish, perhaps because they embody a more consistent relationship between orthography and phonology. These findings provide important constraints on assumptions about the form of lexical representations and the parallel activation mechanisms assumed to underlie lexical retrieval.

The aims of this paper were to review recent research on the effects of orthographic neighborhood structure on tasks requiring word identification and to consider the implications of this evidence for models of visual word-recognition. The review was stimulated by the fact that recent research appears to yield contradictory evidence about how orthographic neighbors affect the word-recognition process. It is important to resolve this contradiction, because neighborhood effects provide critical evidence about lexical retrieval and selection processes. To anticipate the outcome of the review, it appears that the conflict in the existing evidence is more apparent than real. Although there have been some divergent outcomes, particularly when the lexical decision task (LDT) is used, in almost all cases there is an obvious systematic difference between the experiments, yielding contradictory results. In this sense, the determinants of particular effects of neighborhood structure are clear, although the interpretation of these effects remains controversial.

THE ISSUES

Neighborhood Size and Neighbor Frequency

Twenty years ago, as part of an investigation of lexical access procedures, Coltheart, Davelaar, Jonasson, and

Besner (1977) reported the effects of manipulating an orthographic similarity metric they labeled as "N." N is a measure suggested by Landauer and Streeter (1973) to define the number of close neighbors of a stimulus and refers to the number of words that can be created by changing a single letter of a target word. For example, *sand* has many neighbors, including *band*, *send*, *said*, and *sank*, while *club* has only one neighbor—*clue*. Using the LDT, Coltheart et al. found that high-N nonwords were classified more slowly than nonwords with few word neighbors, but that N had no effect on performance for words. They interpreted this result as being consistent with logogen-style activation frameworks (Morton, 1970) in which the strength of activation in individual logogens is determined by sensory input and is insensitive to activity in other logogens. N effects on nonword classification were attributed to a decision mechanism sensitive to overall lexical activation.

Interest in the effects of neighborhood structure was revived by two papers published in 1989 that reported apparently contradictory findings. I reported that N *did* affect responses to words in both lexical decision and naming tasks when the stimuli were selected to orthogonally manipulate N and word frequency (Andrews, 1989). In contrast to the inhibitory effect of N on nonword classifications, high N was associated with *better* performance for words in both tasks. The N effect occurred only for low-frequency words in the LDT, while the naming task showed the same tendency toward a larger effect of N for low- than for high-frequency words but no significant interaction between N and frequency.

Simultaneously, Grainger, O'Regan, Jacobs, and Segui (1989) reported investigations of the conjoint effects of the number and frequency of orthographic neighbors. They

This review was prepared with support from Australian Research Council Grant AC9231195. I am grateful to Ken Forster for organizing the Cognitive Science Lexical Processing Workshop at the University of Arizona in November 1995, for which I initially prepared this review. My thanks are extended to Colin Davis for helping to collate the neighborhood statistics for the CELEX database. Correspondence should be addressed to S. Andrews, School of Psychology, University of New South Wales, Sydney, 2052 Australia (e-mail: s.andrews@unsw.edu.au).

found that, regardless of the number of neighbors, both lexical decision latencies and eye-fixation durations were longer to words with at least one high-frequency neighbor. These results conflict with Andrews both directly, by failing to find a systematic relationship between number of neighbors and lexical classification performance, and indirectly, by finding that words with a higher frequency neighbor are classified more slowly, even though such words are likely to have a larger *number* of neighbors than are words with no higher frequency neighbors.

Since these initial contradictory findings, at least 15 papers have been published investigating the effects of neighborhood size and/or frequency on tasks requiring identification of single words. These are summarized in the section entitled *The Evidence*. The attention to this issue is not simply a function of the empirical conflict; rather, it reflects the centrality of neighborhood similarity effects to evaluation of current models of word recognition.

Lexical Retrieval and Selection

The issue of how a target representation is selected from lexical memory has been critical to models of visual word recognition. Because words are constructed from a small set of features and letters, a target word will often be similar to many lexical representations. Retrieval therefore requires selecting the target representation from alternative lexical candidates. Two major metaphors have been used to describe the selection process in lexical retrieval. Search/verification models (Forster, 1976; Paap, Newsome, McDonald, & Schvaneveldt, 1982) assume that a subset of lexical representations is defined by some means, and that each item in that set is then subjected to a detailed verification process to achieve lexical identification. Parallel activation models assume that all similar lexical representations are activated by input from lower level sublexical representations such as features and letters. To allow selection between the multiple lexical representations that may become active, the interactive activation version of this framework (McClelland & Rumelhart, 1981) that has dominated recent discussions, assumes that lateral inhibition operates between activated word neighbors to achieve a "survival of the fittest" outcome. The lexical representation in which activation first exceeds identification threshold is identified as the target.

Both verification and lateral inhibition mechanisms for achieving selection between similar words predict that neighbors will have an inhibitory effect on performance. In verification models, this is because words with more neighbors will yield larger verification sets. Verification is assumed to operate serially and in order of word frequency, so words with large neighborhoods will be identified more slowly, particularly if they have higher frequency neighbors (Paap & Johansen, 1994). Lateral inhibitory mechanisms lead to competition between activated neighbors, which retards the growth of activation in the target representation and delays identification, although, in the interactive activation model, this inhibition may be offset

by excitatory feedback between different levels of representation (Andrews, 1989).

Because theories of word recognition have been so dominated by these metaphors for lexical selection, researchers have been reluctant to accept evidence that is incompatible with their expectation that competing neighbors will interfere with performance, and even more reluctant to believe that similar neighbors may actually facilitate lexical retrieval. Thus, theoretical intuition has fueled the search for inhibitory effects of neighborhood size and frequency, and alternative explanations of the observed facilitatory effects of neighborhood size have been sought that attribute them to some "task-specific" process that does not undermine the assumption that lexical retrieval involves a competitive process (e.g., Grainger & Jacobs, 1996; Jacobs & Grainger, 1992; Johnson & Pugh, 1994).

The conviction that neighbors will compete for identification was reinforced by a computational implementation of the interactive activation model that confirmed the intuitions derived from conceptual versions of the framework. Jacobs and Grainger (1992) reported that a semi-stochastic implementation of the model, based on the parameters used by McClelland and Rumelhart (1981) to simulate letter identification performance, yielded longer identification times for words with higher frequency neighbors than for items with no such neighbors. Jacobs and Grainger (1992) argued that the compatibility between the interactive activation simulation data and the empirical evidence of inhibitory effects of neighbor frequency confirms the model's assumption of interword competition and implies that facilitatory effects of neighborhood size such as those observed by Andrews (1989, 1992) reflect "situations in which the fluent word recognition process is perturbed" and performance must be based on "strategic long-term and serial processes . . . [which are] beyond the scope of the model" (Jacobs & Grainger, 1992, p. 1485).

Clearly, this attempt to resolve the empirical contradiction with simulation data is circular. The theoretical rationale for the prediction and interpretation of inhibitory effects of neighbor frequency was derived from the model's lateral inhibition assumptions. The confirmation that a computational implementation of the model can, indeed, simulate the conceptual prediction validates the *prediction* but does not provide any independent basis for concluding that the model-compatible data are the more valid reflection of "normal" word-identification processes. Moreover, Jacobs and Grainger's (1992) simulation data reflect only one of a multiplicity of possible outcomes of the interactive activation model that depend on the particular parameter set employed. Coltheart and Rastle (1994) reported simulations of Coltheart, Curtis, Atkins, and Haller's (1993) dual-route cascade (DRC) model in which the lexical procedure consists of a modified interactive activation network. Using a set of parameters determined by attempting to simultaneously simulate a number of word-recognition phenomena, they successfully reproduced facilitatory effects of neighborhood size. Thus, depending

on the particular parameters used and the larger context in which the interactive activation network is embedded, a model including competitive lateral inhibition is capable of simulating *either* facilitatory effects of neighborhood size or inhibitory effects of neighbor frequency.

Grainger and Jacobs (1996) have proposed a possible resolution to this ambiguity by adding an additional parameter to the interactive activation model that reflects summed lexical activity. A word with many activated neighbors will yield higher overall lexical activity, so, in contrast to the dampening effect of competing neighbors on activation of the individual target-word detector, neighbor activation will *increase* the value of the summed activity metric. Grainger and Jacob's "multiple-readout model" assumes that, depending on specific stimulus and task parameters, subjects' responses can be based on either this overall activity criterion or on the activity associated with an individual word detector as proposed in the original version of the model (McClelland & Rumelhart, 1981). Thus, Grainger and Jacobs claim that the apparent contradictions within the empirical literature can be explained by differences in the strategies invoked by particular task contexts and that these different outcomes can be accommodated within a single interactive activation framework.

Lexical Identification Tasks

Consideration of the particular task used to assess word identification is central to evaluating the role of neighborhood activation in lexical retrieval. Effects of neighborhood structure have now been investigated in a wide variety of tasks and using a number of dependent measures ranging from the standard measures of reaction time and accuracy to measures of eye-fixation duration (Grainger et al., 1989) and event-related potential waveforms (Michie, Coltheart, Langdon, & Haller, 1994).

The question of what task provides the "purest" measure of lexical retrieval is controversial. The identification tasks that dominated early research on word recognition fell into disfavor as researchers demonstrated that the accuracy measures obtained in such tasks reflected a variety of sophisticated guessing strategies that could obscure lexical retrieval processes (e.g., Massaro, Taylor, Venezky, Jastrzembski, & Lucas, 1980). Although the LDT was initially touted as a pure measure of the processes required to access a lexical representation (e.g., Forster, 1976; Rubenstein, Garfield, & Millikan, 1970), it too has fallen victim to the charge that performance reflects strategic decision-related processes rather than "normal" retrieval (e.g., Balota & Chumbley, 1984; Gordon, 1983). Word-naming tasks have gained in popularity because they require a practiced skill that is, at least sometimes, part of the normal reading process. But the fact that many stimuli can be accurately pronounced without lexical retrieval means that naming data may also be an unreliable index of lexical retrieval (Coltheart, 1978). Concern with tapping the retrieval processes underlying normal language use has led researchers to argue for the use of semantic categorization tasks that require retrieval of lexical semantics

(Balota & Chumbley, 1984); however, they too are subject to strategic influences (Monsell, Doyle, & Haggard, 1989).

The fact that all of these methods of gaining insight into lexical retrieval processes may be contaminated by processes that are specific to the particular task requirements suggests that the solution lies in seeking evidence of convergence in the data obtained across different task contexts (Andrews, 1989, 1992). All of the tasks mentioned above *can* involve lexical retrieval, but they all allow or require other processes as well. Thus, evidence that a variable influences lexical retrieval is provided by the fact that it exerts similar effects across a variety of tasks, combined with an explanation of how that effect is modulated by the processes specific to particular tasks.

Grainger and Jacobs's (1996) model is explicitly formulated to account for cross-task differences in the effects of neighborhood structure across tasks, but has so far been explicitly extended only to perceptual identification and lexical decision tasks. The present review summarizes the data obtained in all of the major tasks used to assess the effects of neighborhood size and frequency on the speed and accuracy of word identification to provide a comprehensive summary of the evidence. This summary provides the basis for determining which, if any, empirical contradictions need to be resolved by different theoretical accounts.

A final task-related issue that needs to be mentioned concerns the distinction between single-word and priming paradigms. The three investigations of neighborhood structure described so far all used single-word paradigms (Andrews, 1989; Coltheart et al., 1977; Grainger et al., 1989). In such paradigms, the conclusion that performance is sensitive to the number or frequency of neighbors is an inference that rests on the assumption that the stimulus sets selected to differ on a particular neighborhood metric do *not* differ systematically on any other dimension relevant to word identification. That is, the claim that performance differences between words with different neighborhood structures—whether facilitatory or inhibitory—reflect the impact of neighbor activation can never be *directly* validated, because single-word paradigms provide no means of indexing activated neighbors. Accepting such a conclusion requires accepting that the difference in neighborhood structure is the only systematic difference between the stimulus sets, and therefore that neighbor activation provides the only plausible account of the observed outcome. For example, Grainger and Segui (1990) claimed that Andrews's (1989) apparent effects of neighborhood size might actually be due to differences in bigram frequency that were confounded with N, and so may not provide evidence for the influence of lexical neighbors. Andrews (1992) countered this claim by demonstrating that facilitatory effects of N were still evident when high- and low-N stimuli were matched on bigram frequency, ruling out the alternative orthographic redundancy interpretation.

This difficulty in interpretation is the familiar problem of establishing the internal validity of an experimental outcome. Since words vary on a wide variety of dimensions,

there are always a range of possible hypotheses as to the basis of differences between performance for two different sets of words. The validity of a particular conclusion can be evaluated only through a logical evaluation of the plausibility of alternative explanations. Priming paradigms provide a potential solution to this dilemma because they allow estimates of the effect of a particular experimental manipulation to be derived from responses to exactly the same target items when preceded by different prime stimuli. By this means, covarying attributes of the target are held constant and cannot provide an explanation of differential outcomes for different conditions. Reflecting the increase in internal validity that results from this design, priming paradigms are a popular technique for investigating factors influencing word recognition, including the effects of neighborhood structure (e.g., Colombo, 1986; Forster & Davis, 1984; Segui & Grainger, 1990). The priming paradigm is particularly appealing in the context of investigations of neighbor activation because presenting a prime stimulus that is a neighbor of the target seems to provide a relatively direct analogue of the coactivation of neighbors that is presumed to occur in parallel activation models.

This analogy is seductive, but potentially misleading. Priming paradigms can provide insight into *how* coactivation of a neighbor might influence activation of a target representation, but they do not allow for the conclusion that neighbors *are* activated by presentation of a single target word and exert an influence on target retrieval. Moreover, presenting a neighbor as a prime means that the lexical system receives sensory information consistent with *both* the target and its neighbor. This is clearly not the case when a target word alone elicits activation of a neighbor representation. The consequences of coactivation in the priming paradigm are not, therefore, necessarily relevant to whatever coactivation occurs in response to a single target word. Finally, when primes and targets are presented sequentially, as is usually the case, the temporal coactivation functions are clearly different from those that would be obtained if a target word activated the target and its neighbors simultaneously. The complex interplay between lateral inhibition and excitatory feedback between representational levels in the interactive activation framework means that these differences in time course may be critical to the model's output.

Thus, priming paradigms can provide insight into the nature of the similarity relationships between lexical representations and the manner in which direct competition elicited by the sensory properties of a neighbor influences the processes involved in retrieving a particular target representation. However, they cannot tell us whether the sensory features of a single target word result in activation of similar neighbors or whether concurrent activation of representations that partially match the input information help or hinder retrieval of the target representation. To address these issues, it is necessary to compare words differentiated on particular aspects of neighborhood structure. Although such single-word paradigms raise interpretive

problems, they are the only way to address the central issues. The present review is therefore restricted to investigations using these methods.

THE EVIDENCE

The Neighborhood Structure of English Words

Before turning to investigations of the effects of neighborhood structure on word identification, it is appropriate to first consider the relationships between different measures of the construct. To some degree, the conclusion that the empirical evidence of neighborhood effects is contradictory rests on assumptions about the relationship between neighborhood size and neighbor frequency that have not been directly validated. It seems logical that words with more neighbors would, on average, be more likely to have a higher frequency neighbor, but the existence and strength of this relationship has rarely been reported.

There are, in fact, two sets of neighborhood relationships that need to be explored: those that characterize randomly sampled sets of English words and those that hold within the highly selected stimulus sets used in experimental investigations of neighborhood structure. The relationships within English words in general are relevant to specifying the distributional properties of word neighbors and considering their implications for the significance of neighbor coactivation effects on performance. The relationships within restricted samples of experimental items are important in determining whether there are unintended consequences of selection restrictions that distort relationships between different word attributes or qualify interpretation in other ways.

Neighborhood statistics are likely to be very different for words of different lengths. Frauenfelder, Baayen, Hellwig, and Schreuder (1993) reported analyses of the relationship between neighborhood size and word frequency for English and Dutch words of different lengths. In both languages, there was a very strong relationship between length and neighborhood size: Shorter words have more neighbors because there are a smaller number of orthographically permissible letter combinations. Many investigations of effects of neighborhood structure have been restricted to 4-letter words because they allow for very extreme manipulations of neighborhood size. Longer words tend to have no or very few neighbors.

Table 1 illustrates the difference between neighborhood statistics for different-length words by presenting average statistics for a number of word attributes for 4-, 5-, and 6-letter English words. The samples include all words of the relevant length listed in the CELEX database (Burnage, 1988), which contains approximately 30,000 English words. The neighborhood statistics include the total number of neighbors, the number of these neighbors that are higher and lower in frequency than the target word, the highest frequency of neighbors that are higher in frequency than the target, the percentage of words with at least one higher frequency neighbor, and the "spread" of the neighbors (Pugh, Rexer, & Katz, 1994)—the number of letter posi-

Table 1
Means and Standard Deviations of Measures
of Frequency and Neighborhood Structure for
4-, 5-, and 6-Letter Words From the CELEX Database

	<i>M</i>	<i>SD</i>
1,895 Four-Letter Words		
Word frequency	100.3	549.8
Total number of neighbors	7.2	4.9
Spread of neighbors	2.5	1.1
No. of higher frequency neighbors	3.5	3.4
No. of lower frequency neighbors	3.4	3.8
Highest frequency neighbors	651.4	1,615.9
Percent with higher frequency neighbor	80.3	
2,895 Five-Letter Words		
Word frequency	34.2	156.3
Total number of neighbors	2.4	2.3
Spread of neighbors	1.5	1.2
No. of higher frequency neighbors	1.1	1.5
No. of lower frequency neighbors	1.1	1.6
Highest frequency neighbors	126.4	280.5
Percent with higher frequency neighbor	52.0	
4,166 Six-Letter Words		
Word frequency	16.5	62.6
Total number of neighbors	1.1	1.6
Spread of neighbors	0.8	0.9
No. of higher frequency neighbors	0.5	1.0
No. of lower frequency neighbors	0.5	1.1
Highest frequency neighbors	37.1	81.1
Percent with higher frequency neighbor	30.2	

tions at which a letter can be changed to yield a neighbor (e.g., *clue* has a spread of 2—*glue* and *club*—and *sand* a spread of 4—*band*, *send*, *said*, *sank*). As expected, N is systematically related to word length: 4-letter words have 3 times as many neighbors as do 5-letter words, and 6-letter words have an average of only one neighbor. Shorter words also have their neighbors spread across more positions. There is also a systematic relationship between length and frequency. The average word frequency and the frequency of the highest frequency neighbor are considerably higher for 4-letter words and, while most 4-letter words have a higher frequency neighbor, this is true for no more than half of 5- and 6-letter words.

Table 2 presents the correlations between different measures of neighborhood structure for the same three samples of words. Despite the very different N statistics for words of different length, the pattern of relationships between measures is very similar. The data confirm the expected relationship between neighborhood size and neighbor frequency: words with more neighbors are more likely to have high-frequency neighbors. Of course, they are also more likely to have low-frequency neighbors, so the correlation between these variables does not prevent orthogonal manipulation of the two variables (Forster & Shen, 1996; Sears, Hino, & Lupker, 1995). Table 2 also shows that both the number and frequency of word neighbors are correlated with the spread of neighbors across letter positions. Predictably, words with more neighbors are more likely to yield neighbors at a number of letter positions.

Within these large samples of items, none of the measures of neighborhood structure are strongly correlated with word frequency, suggesting that high- and low-frequency words have similar neighbor distributions. This finding conflicts with Landauer and Streeter's (1973) conclusion that there are differences in the similarity structure of high- and low-frequency words, but is compatible with Frauenfelder et al.'s (1993) investigation of relatively exhaustive samples of words of different lengths which found that the correlations between log frequency and neighborhood size were statistically significant, but small ($r = .16$ and $.21$ for English and Dutch, respectively). As for the present samples, this relationship was slightly higher for 4-letter words, and was a little stronger for the Dutch than for the English sample.

As would be expected, higher frequency words tend to have fewer higher frequency neighbors and more lower frequency neighbors than do low-frequency words. The relationships with raw frequency are weak, but stronger correlations are evident for log frequency, indicating that the relationship between word frequency and neighbor distribution is not linear. In general, then, low-frequency words have more low-frequency neighbors and fewer higher frequency neighbors than do high-frequency words.

The pattern of relationships is very similar for a selected sample of 160 4-letter words containing equal numbers of high- and low-frequency words (Table 3). As for the exhaustive word samples, words with more neighbors were more likely to have high-frequency neighbors. This confounding of a word's frequency with its distribution of higher and lower frequency neighbors highlights, rather than reduces, the contradiction between the empirical effects of neighborhood size and frequency. Andrews (1989) found that facilitatory effects of high N were more marked for low-frequency words, particularly in the LDT. Tables 2 and 3 show that low-frequency items are more likely to have higher frequency neighbors. They should, therefore, according to Grainger et al.'s (1989) data, be the items most likely to suffer inhibition from high-frequency neighbors.

Thus, this evaluation of the neighborhood structure provides no resolution to the empirical contradiction between investigations of neighborhood structure. Although the 4-letter words that have dominated research are higher in average frequency and have more neighbors than do longer words, the same pattern of neighborhood relationships is evident for longer words despite their very small number of neighbors. The high-N low-frequency words that Andrews (1989) reported to be facilitated relative to low-N words in LDT and naming tasks are precisely those that should, according to Grainger et al. (1989), suffer most inhibition from higher frequency neighbors. We must, therefore, seek resolution in a detailed consideration of the empirical evidence itself.

Investigations of Neighborhood Size and Neighbor Frequency

Table 4 summarizes the results of 16 published papers reporting investigations of the effects of neighborhood

Table 2
Correlations Between Measures of Frequency and Neighborhood Structure
for 4-Letter, 5-Letter (Italics), and 6-Letter (Bold) Words From the CELEX Database

	Word Frequency	Log Word Frequency	No. of High- Frequency Neighbors	No. of Low- Frequency Neighbors	Highest Frequency Neighbors	Total Neighbors
Log Word Frequency	.45 .52 .62					
No. of High-Frequency Neighbors	-.15 <i>-.11</i> -.10	-.36 <i>-.21</i> -.18				
No. of Low-Frequency Neighbors	.19 <i>.18</i> .14	.70 <i>.55</i> .37	-.11 <i>.01</i> .09			
Highest Frequency Neighbors	.08 <i>.21</i> .07	.09 <i>.19</i> .11	.23 <i>.28</i> .37	.13 <i>.14</i> .19		
Total Neighbors	.03 <i>.05</i> .01	.24 <i>.20</i> .08	.65 <i>.70</i> .73	.67 <i>.69</i> .70	.26 <i>.28</i> .37	
Spread of Neighbors	.03 <i>.03</i> .01	.21 <i>.16</i> .09	.45 <i>.58</i> .64	.49 <i>.54</i> .60	.16 <i>.24</i> .34	.72 <i>.82</i> .87

size or frequency. Most of these papers reported multiple experiments using different tasks, which are presented separately in the summary. These experiments were selected because they used single presentations of words systematically selected to manipulate neighborhood size and/or frequency in tasks that yielded measures of reaction time and/or accuracy.¹ Experiments that investigated neighborhood effects only for nonword stimuli (Laxon, Masterton, Pool, & Keating, 1992; McCann & Besner, 1987) or that investigated unskilled rather than skilled readers (Laxon, Coltheart, & Keating, 1988) are not included. The tasks include perceptual identification tasks in which stimuli were degraded in some manner and an identification response was required, and tasks requiring lexical, speeded naming, and semantic categorization responses to clearly presented stimuli. Investigations relying on measures of eye fixations are not included in the table because they generate more complex data that are not always amenable to a simple summary, but they are discussed in the accompanying text.

The outcomes of each experiment are summarized separately for manipulations of neighborhood size and neighbor frequency. Each outcome is classified according to the

putative influence of neighbor activation indicated by the result. Thus, experiments in which responses to words with more neighbors or higher frequency neighbors are responded to more efficiently than are words with few neighbors or no high-frequency neighbors are classified as "Facilitation" effects because they imply that activation of neighbors benefits performance. Conversely, experiments in which words with many or high-frequency neighbors are responded to less efficiently are classified as implying "inhibition" from activated neighbors. Not all of the experiments included manipulations of *both* neighborhood size and neighbor frequency. Experiments that included a manipulation of either of these factors but found that it exerted no significant effect are summarized as "null," while experiments that included no manipulation of one of the variables are indicated with a dash. Particular conditions of the experiment that differ from those "standard" to each task are listed and discussed in the text, and the summary information for studies using languages other than English are italicized. The results for each task will first be discussed separately, and then the implications of the pattern of results across tasks will be evaluated.

Table 3
Correlations Between Measures of Word Frequency
and Neighborhood Structure for a Sample of 160 4-Letter Words
Selected to Factorially Manipulate N and Word Frequency

	Word Frequency	No. of High-Frequency Neighbors	No. of Low-Frequency Neighbors	Total Neighbors	Spread of Neighbors
Word Frequency	—				
No. of High-Frequency Neighbors	-.53	—			
No. of Low-Frequency Neighbors	.48	-.17	—		
Total Neighbors	.02	.56	.71	—	
Spread of Neighbors	.07	.33	.55	.70	—

Table 4
Summary of Empirical Investigations of the Effects of Neighborhood Structure

Paper	Experiment/Conditions	Neighborhood Size Effect	Neighborhood Frequency Effect
Perceptual Identification			
Carreras et al. (1997)	Experiment 1 ^a	<i>Inhibition</i>	<i>Inhibition</i>
Grainger & Segui (1990)	Experiment 1	—	<i>Inhibition</i>
Grainger & Jacobs (1996)	Experiment 1A	<i>Null</i>	<i>Inhibition</i>
Snodgrass & Minzer (1993)	Experiment 1	Null	—
	Experiment 2	Facilitation	—
	Experiments 3, 4, 5	Inhibition ^a	—
Lexical Decision			
Andrews (1989)	Experiments 1, 2	Facilitation ^a	—
Andrews (1992)	Experiment 1	Facilitation ^a	—
Carreiras et al. (1997)	Experiment 2 ^a	<i>Null</i>	<i>Inhibition</i>
	Experiment 3 ^a : N blocked	<i>Facilitation</i>	<i>Inhibition</i>
Coltheart et al. (1977)	Experiment 2	Null	—
Forster & Shen (1996)	Experiments 1 ^a , 2 ^a , 3 ^a	Facilitation	Null
Grainger et al. (1989)	Experiment 1	—	<i>Inhibition</i>
Grainger (1990)	Experiment 1	—	<i>Inhibition</i>
Grainger et al. (1992)	Experiments 1, 2	—	<i>Inhibition</i> ^b
Grainger & Jacobs (1996)	Experiment 1B: High N NW	<i>Null</i>	<i>Inhibition</i> ^c
	Experiment 1C: Low N NW	<i>Facilitation</i> ^d	<i>Inhibition</i> ^c
	Experiment 1D: High N NW	<i>Facilitation</i> ^d	<i>Null</i>
Huntsman & Lima (1996)		—	<i>Inhibition</i>
Johnson & Pugh (1994)	Experiments 1 ^a , 3 ^a : Legal NW; N blocked	RT: <i>Inhibition</i> Errs: <i>Facilitation</i>	—
	Experiments 2 ^a , 3 ^a : Illegal NW	<i>Facilitation</i>	—
	Experiment 4 ^a : N Mixed	<i>Facilitation</i>	—
	Experiment 5: Legal NW; N blocked	<i>Inhibition</i>	—
	Experiment 6 ^a : Legal NW; N blocked	<i>Facilitation</i>	—
Michie et al. (1994)		<i>Facilitation</i>	—
Perea & Pollatsek (1997)		—	<i>Inhibition</i>
Sears et al. (1995)	Experiment 1	<i>Facilitation</i>	Null
	Experiment 3A	<i>Facilitation</i> ^a	—
	Experiments 4A ^a , 6 ^a	<i>Facilitation</i>	<i>Facilitation</i>
	Experiment 5: High N NW only	<i>Facilitation</i>	Null
Naming			
Andrews (1989)	Experiment 3	<i>Facilitation</i>	—
Andrews (1992)	Experiment 2	<i>Facilitation</i>	—
Carreiras et al. (1997)	Experiment 4 ^a	<i>Facilitation</i> ^d	<i>Inhibition</i> ^c
Grainger (1990)		—	<i>Facilitation</i>
Sears et al. (1995)	Experiment 2	<i>Facilitation</i>	<i>Facilitation</i>
	Experiment 3B ^a	<i>Facilitation</i>	—
	Experiment 4B ^a	<i>Facilitation</i>	<i>Facilitation</i>
Peereman & Content (1995)	Experiment 1 ^a	<i>Facilitation</i>	—
	Experiment 2 ^a	<i>Facilitation</i>	—
Semantic Decision			
Carreiras et al. (1997)	Experiment 5 ^a	<i>Null</i>	<i>Inhibition</i> ^e
Forster & Shen (1996)	Experiments 4 ^a , 5 ^a	Null	Null

^a Only low-frequency stimuli included or effect significant only for low-frequency words. ^b Inhibition significant for items with neighbors at the 4th but not the 2nd position of 5-letter words. ^c Significant only for words from small neighborhoods. ^d Significant only for items with higher frequency neighbors. ^e Significant only for words from large neighborhoods.

Perceptual identification tasks. Four papers have reported investigations of effects of neighborhood structure on identification of degraded words, only one using English stimuli. Snodgrass and Minzer (1993) conducted five experiments in which high- and low-N English words were presented in a series of increasing fragments, and subjects were required to make either successive attempts at identifying the word (Experiments 1 and 2) or a single identification response (Experiments 3, 4, and 5). The successive guessing procedure yielded null effects or weak facilitatory

effects of neighborhood size, but when subjects were required to make a single identification response, accuracy was lower for low-frequency words from large neighborhoods. This was true whether fragments were exposed slowly (1 sec/fragment) or quickly (167 msec/fragment).

Converging results were reported by Grainger and Segui (1990, Experiment 2) for manipulations of neighbor frequency using French stimuli and subjects. They used a “progressive demasking” procedure whereby subjects make a single identification response to a display consist-

ing of interleaved presentations of a target word and a mask in which the length of the target exposure is progressively increased. Paralleling Snodgrass and Minzer's (1993) findings for manipulations of neighborhood size under single response conditions (Experiments 3, 4, and 5), words with a high-frequency neighbor were less accurately identified, particularly when the target was low-frequency. More recently, Grainger and Jacobs (1996) and Carreiras, Perea, and Grainger (1997) have used the same methodology to investigate the independent effects of neighborhood size and neighbor frequency in French and Spanish, respectively. Both investigations replicated the inhibitory effect of neighbor frequency and found a parallel inhibitory effect of neighborhood size, although the latter was not significant in Grainger and Jacobs's data.

Thus, the results obtained in perceptual identification tasks have been relatively consistent. Responses to degraded stimulus presentations are less accurate for low-frequency words that are similar to many words or that have a high-frequency neighbor unless the subject is allowed successive guesses at the target, as they were in Snodgrass and Minzer's (1993) first two experiments. Grainger and Segui (1990) interpreted their results as evidence for lateral inhibition from neighbors and therefore claimed that they "support activation-based accounts of the word-recognition process" (p. 191). Snodgrass and Minzer, however, acknowledged that perceptual identification tasks were subject to response-bias effects. They attributed the interference effects for high-N words observed in their later experiments to the fact that "subjects *guessed* high-frequency neighbors because they had no opportunity to eliminate them in previous responses" (italics added, p. 262), as they did when multiple successive responses were allowed. They concluded that "whether neighborhood effects are facilitatory or inhibitory depends on whether feedback allows subjects to disconfirm initial hypotheses that the target is a high-frequency neighbor" (p. 247).

Although they defended the use of perceptual identification tasks to study word recognition, Snodgrass and Minzer (1993) claimed that their major advantage lay in the information they provided about perceptual and response errors. Thus, people's tendency to guess higher frequency neighbors of the target suggests that neighbors can be sufficiently activated by partial perceptual information to become available as a response. However, the processes and strategies engaged to resolve partial information may be different from those underlying lexical selection when full stimulus information is available. That is, the fact that subjects sometimes guess a neighbor shows that the neighbor was active, but it does not show that its activation interfered with retrieval of the target word, or even that neighbors would necessarily become active when the available perceptual information uniquely identified the target word rather than being compatible with a number of different words, as may be the case for degraded presentations of words with many neighbors.

This conclusion conflicts with Grainger and Jacobs's (1996) multiple-readout account of neighborhood effects because their claim that single-response perceptual identification tasks rely on activation of individual word detectors rather than on overall lexical activity implies that these tasks are "a more direct reflection of the visual word recognition process" (p. 542) than tasks such as lexical decision. The relative validity of these two views of the perceptual identification task will be discussed when the full set of cross-task comparisons are considered.

Lexical decision task. This is the task that has been most frequently studied and, perhaps consequently, the one that has been argued to yield the most contradictory results (e.g., Grainger & Jacobs, 1996). My initial finding of facilitatory effects of N for 4-letter low-frequency words (Andrews, 1989) was argued to be ambiguous because bigram frequency, which is correlated with N, was not controlled (Grainger & Segui, 1990). This interpretation was ruled out in my later experiments, which replicated the facilitatory effects of N in words matched on bigram frequency (Andrews, 1992). An independent replication using substantially the same stimuli as Andrews (1989) was reported by Sears et al. (1995, Experiment 3), and Michie et al. (1994) replicated the facilitatory effect of N using a new word sample that included 4- to 6-letter words. None of these experiments controlled neighbor frequency, but the earlier analyses of neighborhood structure in selected word samples imply that the low-frequency high-N words would have been likely to have more high-frequency neighbors than their low-N counterparts. However, this does not necessarily ensure that the high- and low-N word samples differ in whether or not they have a *single* high-frequency neighbor—the variable that Grainger et al. (1989) found to be critical.

The strongest evidence of facilitatory effects of N for English stimuli are provided by two recent papers reporting experiments that independently manipulated N and neighbor frequency. Sears et al. (1995) found facilitatory effects of N in four lexical decision experiments using three different samples of 4-letter stimuli and one of 5-letter stimuli, all of which included high- and low-N words with and without a higher frequency neighbor. They found no effect of neighbor frequency. Recently, another set of experiments by Forster and Shen (1996) provided a further replication of the differential effects of N and neighbor frequency in three experiments using low-frequency 5- and 6-letter words. Again, they found effects of N but not of neighbor frequency. Thus 11 experiments conducted in four independent laboratories have found facilitatory effects of N on lexical classification performance for English words. The effect tends to be larger for low-frequency words, but this appears to be primarily due to a reduced N effect for very high-frequency words (Sears et al., 1995).

Similar investigations of the independent effects of N and neighbor frequency have been recently reported using French (Grainger & Jacobs, 1996) and Spanish (Carreiras et al., 1997) stimuli. In contrast to the English studies, both

papers reported inhibitory effects of neighbor frequency in combination with either null or facilitatory effects of N. Although the facilitatory effects of N observed by Grainger and Jacobs were significant only for words with higher frequency neighbors, this may be because some of the neighbors of low-frequency high-N target words may be unknown to the subjects, leading to a manipulation of actual neighborhood size that was weaker than it would have been for words with high-frequency neighbors.

Effects of nonword environment in the LDT. Most of the experiments described above have used what might be described as the "standard" lexical decision procedure of randomly intermixing words and nonwords selected to be as equivalent to the words as possible. Nonwords are chosen to be wordlike in the sense that they are phonologically legal and orthographically well structured. They are equated with the words used in the study on a variety of measures of orthographic and phonological structure so that decisions about lexicality must be based on retrieval of lexical information rather than on other structural criteria alone.

Johnson and Pugh (1994) reported a series of studies that varied this standard procedure and investigated changes in the effects of neighborhood size as a function of manipulations of both the nature of the nonword environment and the range of word and nonword stimuli that people were exposed to. They found inhibitory effects of N when words had to be discriminated from legal pronounceable nonwords, but facilitatory effects when illegal nonwords were used instead. They interpreted these data as indicating that lexical retrieval of the target word was inhibited by activated neighbors and that facilitatory effects of neighborhood size were due to response bias. When nonwords were illegal, lexical classifications could be based on lexical activity—whether due to neighbors or the target word itself—without full lexical retrieval, and words from large neighborhoods would therefore be classified more quickly; but when words had to be discriminated from legal nonwords, "the task demands that subjects resolve the cohort to a single candidate (or no candidate) before responding" (Johnson & Pugh, 1994, p. 300) and the interference caused by neighbors would become evident.

Johnson and Pugh's (1994) finding of facilitatory effects of N when illegal nonwords were used is consistent with results of both Andrews (1989) and Forster and Shen (1996), who found a larger and more robust facilitatory effect of N on LDT responses to words when the nonwords were less wordlike. However, their finding of interference for high-N words in an environment of legal nonwords conflicts with the facilitatory results described above. Johnson and Pugh identify the source of the conflicting outcomes as being due to their use of the unusual experimental procedure of blocking stimuli according to neighborhood size and length. When they adopted the more standard method of randomly intermixing stimuli that had been employed in all of the other research on this issue, they found facilitatory effects that were significant in the

accuracy but not the latency data (Experiment 4). Converging evidence that Johnson and Pugh's inhibitory N effects are due to the blocking manipulation rather than to the more difficult discrimination requirements is provided by Sears et al. (1995). They found facilitatory N effects and no effect of neighborhood frequency when all of the nonwords were high N but randomly intermixed with word stimuli. According to Johnson and Pugh's argument, this condition should reduce the feasibility of making lexical classifications on the basis of overall lexical activity alone and result in inhibitory N effects.

Systematic investigations of varying the decision requirements of the LDT have also been reported by Grainger and Jacobs (1996) and Carreiras et al. (1997) using French and Spanish stimuli, respectively. Carreiras et al. found that neighborhood size effects were facilitatory when high- and low-N words were presented in separate blocks in an environment of mixed high- and low-N nonwords, a result that parallels Johnson and Pugh's (1994) findings for English. Grainger and Jacobs compared performance for words when embedded in nonwords that were either high or low in wordlikeness.² When highly wordlike nonwords were used, they found inhibitory effects of neighbor frequency and no effect of N. However, with less wordlike nonwords, N effects were facilitatory and the inhibitory effects of neighbor frequency were reduced. A similar pattern of results was found when speed pressure was applied to discriminations from highly wordlike nonwords. These findings with French stimuli parallel the English results in showing that effects of neighborhood size are facilitatory under easier discrimination conditions, although, in English, N effects remain facilitatory in more difficult decision environments (Sears et al., 1995, Experiment 5). By contrast, Grainger and Jacobs and Carreiras et al. found no effects of neighborhood size for French and Spanish words under these conditions.

Thus, the effects of neighborhood size on lexical decision performance appear to depend on the nature of the word/nonword environment and may vary across languages. For all the languages that have been investigated, a clear, robust advantage for words from large neighborhoods is observed when words must be discriminated from nonwords that are illegal or low in wordlikeness. When English words and legal nonwords varying in neighborhood size are randomly intermixed, the effects are facilitatory but somewhat smaller than in the easier decision environment (Andrews, 1989; Forster & Shen, 1996). But when the same mixture of words and nonwords is presented in blocks of high- or low-N stimuli of a particular length, performance is better for low-N than for high-N words—that is, the effects of neighbor activation appear to be inhibitory (Johnson & Pugh, 1994, Experiment 5). The pattern of results for French and Spanish words also depends on the nonword environment, but facilitatory effects of N appear to occur only in easy discrimination environments or in conditions emphasizing speed of response (Carreiras et al., 1997; Grainger & Jacobs, 1996). The sensitivity of performance to discrimination difficulty clearly

implies that N effects are influenced by decision components of the task. But interpretation of the implications of the evidence regarding N effects on lexical decision performance is controversial and depends on judgments as to how discrimination difficulty affects performance. These issues will be taken up once the effects of neighborhood structure on other tasks have been reviewed.

Even though the interpretation of the LDT data remains a subject of debate, the data themselves are not, ultimately, contradictory. N is associated with facilitatory effects on lexical classification, at least for low-frequency English words, unless stimuli are blocked by neighborhood size. Facilitatory effects of N do not appear to occur in French and Spanish under standard LDT conditions, but can be seen in particular nonword environments. But even here, N does not exert the inhibitory effect predicted by a simple competitive mechanism.

The effects of neighborhood frequency also appear to show variability across languages. All but two of the studies demonstrating inhibitory effects of higher frequency neighbors have used non-English stimuli—French (Grainger & Jacobs, 1996; Grainger et al., 1989; Grainger & Segui, 1990), Dutch (Grainger, 1990), or Spanish (Carreiras et al., 1997). Using English stimuli, both Sears et al. (1995) and Forster and Shen (1996) were unable to replicate these effects despite a number of systematic attempts which relied on the same “single higher frequency neighbor” metric that Grainger employs. The only studies reporting inhibitory effects of N in English are by Huntsman and Lima (1996) and Perea and Pollatsek (in press). Both experiments compared pairs of items that were matched on N and word frequency, but were either among the most or the least frequent words in their respective neighborhoods. The words with higher frequency neighbors were classified more slowly. Unfortunately, neither paper reports details about the neighborhood size of their nonword stimuli. It is therefore impossible to determine whether the discrepancy between these results and those reported by Forster and Shen (1996) and Sears et al. (1995) reflects decision strategies encouraged by the task conditions. Perea and Pollatsek (in press) deliberately emphasized accuracy rather than speed of performance, so their results may reflect a more cautious decision strategy.

Although the effects of neighbor frequency on LDT performance for languages other than English appear to be consistent, in the sense that inhibition has been observed in eight of the nine published experiments, questions remain about their generality. Grainger, O'Regan, Jacobs, and Segui (1992) used 5-letter French words, rather than the 4-letter words used by Grainger et al. (1989) and Grainger (1990), and manipulated the letter position at which a higher frequency neighbor could be formed by changing a letter. Although they found significantly slower average performance for words with a higher frequency neighbor than those with no such neighbor, the effect was only evident for word with neighbors at the fourth position, and not for words with second position neighbors.

Grainger and Jacobs (1996) attribute the variable outcomes of LDT investigations of neighborhood structure

to the fact that lexical decisions can sometimes be based on overall lexical activity rather than “unique word identification.” Like Johnson and Pugh (1994), they attribute facilitatory effects of N to processes specific to the LDT and argue that this conclusion is supported by the various demonstrations that facilitation effects are enhanced under the easier discrimination conditions created by using illegal or low-wordlike nonwords. Unique word identification is, they argue, best indexed by performance in perceptual identification tasks which generally reveal inhibitory neighbor influences. LDT performance can provide insight into word-identification processes, but only in conditions where decisions cannot be based on overall lexical activity. These assertions about the processes underlying performance in different tasks will be reconsidered when the full array of data has been considered.

In relation to the LDT data alone, however, it should be noted that Table 4 provides little basis for the fundamental premise of Grainger and Jacobs's (1996) model: that LDT performance varies in a manner that suggests the involvement of two processes, one of which shows facilitatory effects of coactivated neighbors and the other, inhibitory effects. Despite the influence of nonword discriminability, the effects of neighborhood size are remarkably consistent. Apart from one experiment by Johnson and Pugh (1994, Experiment 1), the effects of manipulations of neighborhood size are all either facilitatory or null, and two of the three null effects occurred in languages other than English. Systematic changes in the magnitude of the N effect *are* observed with changes in nonword environment, but regardless of these decision influences, the overall effect of N is facilitatory. Certainly, there is no evidence of the inhibitory effects for words from large neighborhoods that were predicted from the original interactive activation model (Jacobs & Grainger, 1994) and that Grainger and Jacobs (1996) now attribute to responses based on unique word identification. Thus, there is minimal evidence of the empirical contradictions regarding the effects of neighborhood size that provided part of the motivation for the multiple readout model.

At first glance, the evidence regarding the effects of neighbor frequency seems to contradict this conclusion, because the majority of investigations of this variable have found inhibitory effects. High-N words tend to have more neighbors, so this finding conflicts with the facilitatory effects of neighborhood size. However, closer inspection suggests that inhibitory effects of neighbor frequency are primarily restricted to languages other than English. Only two of the eight studies of effects of neighborhood frequency in English have reported an inhibitory effect. Thus, language appears to be a systematic determinant of whether or not inhibitory effects of higher frequency neighbors occur. Two of the three null effects of neighborhood size also occurred for non-English stimuli. Possible explanations of this language specificity will be considered in the conclusion sections.

Word naming. Word-naming data present a much more consistent picture than the LDT results. *All* of the studies using this task have found facilitatory effects of

neighborhood size at least for low-frequency words. Three of the four studies manipulating neighbor frequency have also yielded facilitatory effects, including Grainger's (1990) investigation of French stimuli and subjects. The single exception is reported by Carreiras et al. (1997) using Spanish stimuli, although this is somewhat ambiguous. There was no overall effect of neighbor frequency, but an interaction with N. N showed a facilitatory effect for words with higher frequency neighbors only; and the effect of neighbor frequency was inhibitory for words with few neighbors, but slightly facilitatory for words with many neighbors. The lack of a significant N effect for stimuli with predominantly low-frequency neighbors may indicate that some of the neighbors of high-N words are not in subjects' vocabularies, leading to a weak manipulation of neighborhood size. Alternatively, the different results for Spanish stimuli may be a further reflection of the language-specific effects that appear to characterize the lexical decision task.

Thus, the empirical evidence regarding word naming is not in contention. With the single possible exception of neighbor-frequency effects in Spanish, naming times are faster for words with both more neighbors and higher frequency neighbors. What *is* the subject of debate is whether neighborhood activation effects on word naming necessarily reflect lexical retrieval. They may instead be due to the effects of orthographic structure on nonlexical naming procedures. High-N words might be faster to name because they contain more common spelling-sound correspondences. Consistent with this possibility, high-N nonwords are named faster than low-N nonwords (e.g., Laxon et al., 1992; McCann & Besner, 1987).

The most direct evidence regarding the contribution of nonlexical naming procedures to N effects is provided by Peereman and Content's (1995) comparison of effects of neighborhood size on word-naming performance as a function of whether or not words were mixed with nonwords. If N effects on word naming are due to nonlexical naming procedures, they should be enhanced when subjects must name nonword as well as word stimuli. This prediction derives from the assumption that nonword naming must rely on nonlexical procedures, because nonword pronunciations are not lexically represented, and that this leads to a bias toward using this procedure to name words as well. Evidence for such a nonlexical bias is provided by findings of enhanced effects of phonological regularity (Monsell, Patterson, Graham, Hughes, & Milroy, 1992) and reduced effects of semantic priming (Baluch & Besner, 1991) for words named in an environment of nonwords. In contrast to the predicted enhancement of N effects in nonword environments, Peereman and Content reported a *smaller* N effect when words were mixed with nonwords rather than words. Peereman and Content concluded that their data were "incompatible with the notion that neighborhood size effects in naming are due to the strength of analytic correspondences [between letters and phonemes]" and support the view that the effects reflect "lexical activation of the neighbors" (p. 416).

Semantic categorization. Forster and Shen (1996) have recently added to the battery of tasks probed for effects of neighborhood size and frequency by conducting two experiments using a semantic categorization task. They argue that this task provides the critical test of whether effects of either neighborhood metric reflect lexical retrieval processes because, in contrast to lexical decision and naming tasks which might be performed using familiarity or nonlexical procedures, respectively, it requires "both identification and access to meaning" (p. 703). Of course, the problem is that this task may also invoke its own specific decision strategies (Monsell et al., 1989). Forster and Shen used a binary animal/nonanimal classification task and investigated N effects for items receiving a "no" categorization response. They claim that this methodology is less vulnerable to a variety of decision artifacts than are other versions of the paradigm. Their results showed no systematic effects of either neighborhood size or frequency, although both experiments showed a significant facilitatory effect in the subjects' analyses which did not generalize over items. Carreiras et al. (1997) employed the same methodology with Spanish stimuli. They also found no significant effect of neighborhood size, but the effects of neighbor frequency were significantly inhibitory for high-N words.

Forster and Shen (1996) concluded that the absence of inhibitory effects in either lexical decision or semantic categorization suggested that "competition does not play a key role in visual word recognition" (p. 696). However, because the facilitatory neighborhood-size effects do not extend to semantic categorization performance, they must reflect task-dependent processes rather than lexical retrieval. Carreiras et al. (1997) interpreted their results within Grainger and Jacobs's (1996) multiple readout model and therefore predicted that semantic categorization should follow the same pattern as perceptual identification, because both are presumed to rely primarily on unique word identification. The absence of clear inhibitory effects of high-frequency neighbors in the semantic categorization task contradicted this prediction.

Eye-movement data. All of the experiments discussed so far used measures of RT and accuracy as dependent variables. Two papers have added to this body of evidence by collecting measures of eye fixations and movements. These data are not summarized in Table 4 because they rely on different dependent measures, but they deserve discussion because they may shed light on some of the apparent contradictions in the behavioral measures of performance.

Grainger et al. (1989) recorded gaze durations for French stimuli in a task requiring judgments of semantic relatedness. Target words varying in N and neighbor frequency were presented with related or unrelated filler words and the total time spent fixating the target was computed. As summarized in Table 4, LDT data for the same stimuli had revealed significantly slower responses to words with higher frequency neighbors than to words with no such neighbors. The magnitude of the inhibitory effect was unaffected by either the number of neighbors overall or the

number of higher frequency neighbors. The gaze-duration data were contradictory. Words with a single higher frequency neighbor were classified more slowly than were words with either no neighbors or with low-frequency neighbors. However, this inhibitory effect was not evident for stimuli with more than one high-frequency neighbor. Since these items were higher in N than were the other groups, this result may reflect a facilitatory effect of number of neighbors over and above the effects of neighbor frequency.

A recent paper has reported a more detailed set of measures of eye movements associated with words with and without higher frequency neighbors (Perea & Pollatsek, in press). The LDT results for these stimuli, which are included in Table 4, represent one of the two findings of inhibitory neighbor-frequency effects observed in English. Perea and Pollatsek presented the same stimuli in sentences and measured initial fixation durations on the target word, as well as measures of fixations to the next two words and regressions back to the target. The latter measures are referred to as "spillover effects" and assumed to reflect "a final selection process" which is required for "full lexical access." Perea and Pollatsek found no differences between the "first-pass" fixation times to words with and without higher frequency neighbors, but measures of total target-fixation time (including regressions) and of fixations on words in the "immediate post-target region" (the two words following the target) were longer for words with higher frequency neighbors.

They interpreted their results within a model of eye movements that assumes two stages of lexical access: (1) the familiarity check stage, in which the signal for an eye movement to the next word is programmed, and (2) full lexical access or "lexical completion" (Reichle, Pollatsek, Fisher, & Rayner, in press). The first stage is identified with first fixation measures and the second with the delayed "spillover effects." Within this model, Perea and Pollatsek's (in press) data suggest that the inhibitory effects of neighbor frequency occur relatively late in lexical access. These findings do not conflict with those of Grainger et al. (1989), because they masked the target word once subjects shifted their gaze to the comparison word and therefore encouraged subjects to complete processing of the target on the first fixation.

Thus, the available eye-movement data are compatible with the view that higher frequency neighbors inhibit activation of the target word, at least when stimuli are selected to be matched on N but at either the "top" or "bottom" of their orthographic neighborhood. This is the manipulation used by Perea and Pollatsek (in press) and by Huntsman and Lima (1996), who both found inhibitory effects of neighbor frequency in the LDT. Forster and Shen (1996) and Sears et al. (1995) factorially manipulated neighbor frequency and N and found either null or facilitatory effects of neighbor frequency. Thus, the neighbor frequency effects on eye movements need to be replicated using other stimulus manipulations to establish their generality.

The further insight provided by the eye-movement data is that the inhibitory effects of neighbors occur late in lexical processing. There are no differences in initial fixation time on words with and without high-frequency neighbors. Perea and Pollatsek (in press) did not manipulate N, so there is no evidence as to its influence on eye-movement measures of early or late lexical processing.

RESOLUTIONS

Empirical Conflicts

There are three main issues to be resolved regarding the empirical effects of neighborhood structure: Is performance affected by neighborhood size, neighbor frequency, or both? Are the effects of size and/or frequency facilitatory or inhibitory? Do the effects reflect lexical retrieval mechanisms that are common to a variety of word-identification tasks or task-specific processes?

Neighborhood size or neighbor frequency? Table 4 summarizes 43 observations of the effects of N on one of the tasks under consideration and 28 that have investigated the effects of neighbor frequency. All but 10 of the investigations of N (77%) have found better performance for words from large neighborhoods than for words with few neighbors. The divergent data can nearly all be attributed to a systematic feature of the stimuli or task. They include the null effects of N reported in Coltheart et al.'s (1977) original investigation of lexical decision performance for stimuli not categorized by word frequency; the inhibitory effects of N observed in perceptual identification tasks requiring a single response (Carreiras et al., 1997; Grainger & Jacobs, 1996; Grainger & Segui, 1990; Snodgrass & Minzer, 1993); lexical classification tasks in which stimuli are blocked by N (Johnson & Pugh, 1994; Grainger & Jacobs, 1996); and semantic categorization experiments (Carreiras et al., 1997; Forster & Shen, 1996).

The results of investigations of neighbor frequency are more mixed. The most common outcome is an inhibitory effect of higher frequency neighbors (54%), but there are also eight null results and five studies reporting a facilitatory effect. The most obvious determinant of this variability is language. Only 2 studies using English stimuli have reported inhibition (Huntsman & Lima, 1996; Perea & Pollatsek, in press) while 11 have reported either null or facilitatory effects in carefully controlled samples of words manipulating both neighborhood size and frequency (Forster & Shen, 1996; Sears et al., 1995). The facilitatory effects with English stimuli are primarily observed in the naming task. Neighbor frequency appears to exert little influence on LDT or semantic categorization performance for English words.

Facilitatory or inhibitory effects? At this overall level, then, the data suggest that identification of English words is more consistently influenced by neighborhood size than by neighbor frequency, and that the effects of this variable are almost always facilitatory except in perceptual identification tasks. A different picture appears to

emerge for French and Spanish, which do appear to be influenced by neighbor frequency and show clear facilitatory effects of N only in word-naming tasks. Notably, however, the effects of N on these languages are still *inhibitory* only in perceptual identification tasks. The LDT shows either null or facilitatory effects, depending, apparently, on non-word environment (Grainger & Jacobs, 1996).

Thus, the contradiction between the empirical effects of neighborhood structure that provided the impetus for the recent flurry of research on this issue seems to be more apparent than real. In English, there is both a relatively consistent pattern of results, and no inherent contradiction between the effects of neighborhood size and neighbor frequency. The effects of N are compatible with the view that activation of orthographically similar neighbors facilitates rather than interferes with word identification, and there is minimal evidence of interference from higher frequency neighbors to undermine these conclusions. However, the nature of neighborhood effects does vary according to task requirements.

The results for French and Spanish do show an apparent conflict between the effects of neighbor size and frequency. Words with many neighbors do not suffer inhibition and can show facilitation in some contexts, but when words are selected according to the presence of higher frequency neighbors, inhibition is evident. Grainger and Jacobs (1996) attribute this data pattern to the contribution of different common and specific processes to performance in different tasks.

Lexical retrieval versus task-specific processes?

The effects of neighborhood structure on the word-naming task summarized in Table 4 are uniformly consistent with a facilitatory influence from activated neighbors. The same pattern holds for the LDT as long as the stimuli are English and either all low frequency or balanced for word frequency (e.g., Andrews, 1989; Forster & Shen, 1996), and as long as the word and nonword stimuli cover the complete range of neighborhood size rather than being blocked by N and length (Johnson & Pugh, 1994). The facilitatory effects of N on lexical decision performance are more consistent for low-frequency words, but it appears that the absence of N effects may pertain only to very high-frequency words (Sears et al., 1995). Semantic categorization performance does not show consistent effects of either neighborhood size or neighbor frequency (Forster & Shen, 1996).

In contrast to the relatively consistent facilitatory effects obtained in tasks using clear stimulus conditions, neighbors appear to exert an inhibitory influence on perceptual identification performance unless subjects are allowed to make repeated responses to the stimulus. This result is hardly surprising under degraded presentation conditions. Reducing the completeness of stimulus information will have more of an effect on words that are similar to many words than on words with few neighbors, because, with only a single opportunity to respond, subjects may guess a high-frequency neighbor of the target and therefore achieve lower overall accuracy than they would for words with few similar neighbors. The fact that neigh-

bors of the target are available as responses confirms that neighbors *can* be activated by incomplete sensory features, and therefore provides evidence that neighbor activation might influence word identification. However, the strategies that subjects use to resolve partial stimulus information over successive degraded presentations may not be the same as those involved when neighbors are activated by a single presentation of the stimulus features of another word. Thus, perceptual identification tasks provide useful insights into the similarity relationships that influence perceptual and response processes (Snodgrass & Minzer, 1993), but the perceptual ambiguity of the stimuli encourages application of strategic guessing processes that may curtail or obscure the mechanisms by which the correct representation is normally selected from multiple possible candidates. The goal of investigating neighborhood effects on word identification is to observe how the word-identification system resolves the ambiguous *consequences* of an unambiguous stimulus event that activates a number of lexical representations. These processes may be very different from those invoked to resolve a perceptually ambiguous stimulus. Data obtained under degraded presentation conditions must therefore be interpreted cautiously. The critical evidence is provided by tasks requiring identification of clearly presented words.

Interpretation of the pattern of effects of neighborhood structure in these tasks depends on assumptions about the role of general lexical retrieval mechanisms and task-specific processes. Traditional serial models of visual word recognition assumed a discrete lexical access procedure that selected the lexical representation that best matched the sensory stimulus and allowed for retrieval of attributes associated with the lexical item, such as its pronunciation or meaning (e.g., Forster, 1976; Paap et al., 1982). According to such frameworks, a common lexical identification process underlies all tasks that require retrieval of a specific lexical representation for successful performance (Monsell et al., 1989).

In more recent models, the assumption of discrete serial processes has given way to the notion of sequential, but overlapping, "cascaded" processes (McClelland, 1979). Such models usually retain the fundamental assumption of a common lexical retrieval procedure. However, the continuous flow of information from one processing stage to another that characterizes cascaded architectures affords the possibility that responses might be able to be based on information apart from the target lexical representation. This is particularly evident in models relying on parallel activation mechanisms (see, e.g., McClelland & Rumelhart, 1981), which assume that a *set* of lexical representations are activated by sensory input. If these multiple candidates are accessible to other cognitive processes, they may provide an alternative basis for achieving task requirements without unique retrieval of the target lexical representation. If processing is unidirectional, then the activated candidates available to influence performance will be determined by their orthographic similarity to the target word. However, in models assuming fully interactive

cascaded processing, such as the complete interactive activation model or recent parallel distributed processing (PDP) frameworks (Plaut, McClelland, Seidenberg, & Patterson, 1996; Seidenberg & McClelland, 1989), it is possible for items semantically related to the target to become activated and influence processing before a single target representation has been uniquely selected.

These consequences of cascaded architectures severely complicate the question of how to investigate lexical retrieval processes. Most fundamentally, they challenge the notion of "lexical access" as an isolable process in visual word identification. There may be no "magic moment" at which the subject has identified the word but not yet gained access to attributes such as its meaning (Balota, 1990). Different task requirements may fundamentally change the processes involved in retrieval and use of word attributes.

These issues have received most attention in discussions of the LDT. This task was initially hailed as providing a relatively direct measure of lexical access because it seemed to *require* lexical access—how else can subjects distinguish between *manliness* and *mantiness*, for example—but involve minimal additional decision-related processes (Rubenstein et al., 1970). But there is clear evidence that lexical decision performance *is* influenced by the discriminability of words and nonwords (Shulman, Hornak, & Sanders, 1978) and it has been argued that decision processes may be a major determinant of lexical decision performance (Balota & Chumbley, 1984; Gordon, 1983).

The evidence concerning the contribution of decision processes to lexical classification performance is contradictory. Balota and Chumbley (1984) argued that lexical classifications were often based on the "familiarity" of a stimulus and reported a larger word-frequency effect in lexical decision than in semantic categorization as evidence for familiarity effects that were specific to the lexical decision process. They therefore questioned the assumption that lexical access was the major determinant of lexical classification performance. This conclusion was challenged by Monsell et al. (1989), who claimed that Balota and Chumbley underestimated the frequency effect in semantic categorization. Monsell et al. reported equivalent frequency effects in categorization and lexical decision when the categorization task required judgments of a single semantic category (people/thing) rather than the changing categories used by Balota and Chumbley. Monsell et al. concluded in favor of a frequency-sensitive lexical identification process that was common to lexical decision, semantic categorization, and naming tasks, but acknowledged that each task was also susceptible to influence from such postidentification processes as retrieval of meaning or such parallel processes as pronunciation assembly. These additional processes might obscure effects due to lexical identification.

With respect to the specific question of how decision processes influence lexical classification, Monsell et al. (1989) distinguished between the independent contribution of a familiarity-sensitive decision process, as pro-

posed by Balota and Chumbley (1984), and the intralexical use of information about the ongoing process of lexical retrieval. They pointed out that activation models allow two sources of information for lexical decision: "occurrence of unique identification" or "the less reliable, but evidently used, at least for nonwords decisions, ... global measure of lexical activation (or change in activation)" (Monsell et al., 1989, p. 48). It is the latter notion of ongoing activation of the lexical cohort of neighbors that underlies both Johnson and Pugh's (1994) interpretation of the facilitatory effects of N on lexical decision performance and Grainger and Jacobs's (1996) account of differences in the pattern of neighborhood effects across LDT and perceptual identification tasks.

Unique word identification versus familiarity. Johnson and Pugh (1994) claimed that the inhibitory effects of N they observed under blocked conditions were a more valid index of lexical identification than the data obtained when stimuli were intermixed. This claim rests on the assumption that, unless stimuli are blocked, subjects required to discriminate words from legal nonwords will tend to base their lexical decisions on overall activation of the cohort of word neighbors rather than on unique identification of the target word. They argued that blocking stimuli by neighborhood size "ensure[s] that cohort resolution would be needed" (p. 288) and reduces the bias to respond "word" to large cohort stimuli that is presumed to underlie the apparent facilitation for high-N words in mixed stimulus conditions. There are at least two problems with this claim.

First, N provides no more useful a basis for classification when the complete stimulus set is randomly intermixed than when high- and low-N word and nonword stimuli are presented separately (Forster & Shen, 1996). Whether stimuli are all high in N, low in N, or mixed, as long as the average N of the words and nonwords is matched, N does not provide a reliable clue to lexicality in any condition. Thus, there seems no a priori basis for Johnson and Pugh's (1994) claim that the inhibitory effects observed when stimuli are blocked reflect lexical retrieval, while the facilitatory effects in mixed conditions are due to bias because there is no more reason to assume that subjects would be biased to rely on N in mixed than in blocked conditions.

Second, blocking of stimuli encourages the development of context-specific strategies. Presenting high- and low-N stimuli in separate blocks allows subjects to vary their decision criteria for the two stimulus classes in a manner that is not possible when high- and low-N stimuli are intermixed. Johnson and Pugh (1994) implicitly acknowledge this by pointing out that the blocking manipulation allows subjects to "easily identify a processing strategy for handling the displays" (p. 288), but fail to consider the possibility that this processing strategy might be different for high-N and low-N stimulus blocks. But, as Forster and Shen (1996) emphasize, "effects obtained only with blocked stimuli are normally thought to be strategic in nature because it is only then that the type of item that is going to

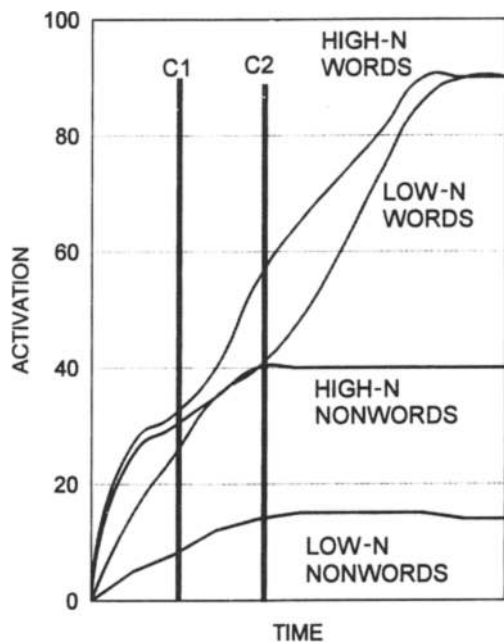


Figure 1. Hypothetical activation functions elicited by high- and low-N words and nonwords. The activation functions for low-N words and nonwords diverge much earlier in processing than do those for high-N words and nonwords. Thus, application of Criterion 1 (C1) in an environment of low-N stimuli would allow lexical discrimination equivalent to Criterion 2 (C2) in a high-N environment. Application of these different criteria for lexical discrimination in the two stimulus environments would yield faster classifications of low-N than of high-N words, even though high-N words are consistently more highly activated than low-N words.

be presented on the next trial can be anticipated" (p. 15). Thus, the slower responses to high-N words in blocked conditions might reflect application of a response criterion when high-N words have to be discriminated from very wordlike nonwords is different from that used when unusual words have to be distinguished from nonwords that are low in similarity to words. For example, the rate of increase in lexical activation for high-N words and nonwords might be similar until relatively late in processing and lead to a strict response threshold. By contrast, the differential activation associated with low-N words and nonwords might be discriminable early in stimulus processing and allow a more lax threshold for classification. Such a difference in criterion would yield longer RT for high-N stimuli even if high-N stimuli had activation levels that were consistently higher than those of low-N words (see Figure 1). This specific account is speculative, but what is clear is that blocking allows subjects to anticipate, and therefore to vary their strategy, for different stimuli in a manner that is not possible with standard mixed-presentation conditions.

Grainger and Jacobs's (1996) account of neighborhood effects in the LDT is similar to that of Johnson and Pugh (1994) in the sense that it attributes neighborhood structure effects in the LDT to a task-specific strategy. By con-

trast, they argue that responses in the perceptual identification task reflect unique word identification that occurs when "the appropriate whole-word orthographic representation reaches a critical level of activation" (Grainger & Jacobs, 1996, p. 521). These assumptions are expressed within the interactive activation model by assuming that LDT performance does not require complete identification and can, instead, be based on "the overall (global) activity in the orthographic lexicon" (p. 522) or, in the case of nonwords, by waiting until a variable deadline has expired. Facilitatory N effects are due to reliance on this intralexical familiarity criterion, while the inhibitory neighborhood effects observed in perceptual identification tasks reflect unique word-identification processes. Variations in the N effect with changes in nonword environment and speed-stress instructions are argued to support the claim that facilitatory N effects are more likely when nonword discriminations are easy or subjects are encouraged to respond quickly.

Grainger and Jacobs's (1996) empirical evidence that N effects in the LDT are influenced by the nature of the stimulus environment is not new. Its compatibility with previous findings in English (Andrews, 1989; Forster & Shen, 1996) demonstrates that, despite the language specificity of neighbor frequency effects, N effects are similar across languages. But, although facilitatory effects of N are largest and most robust when nonwords are illegal or low in wordlikeness, facilitatory effects are also observed in English when words must be discriminated from legal nonwords of equivalent N. It is only when words and nonwords are presented in blocks that are homogeneous with respect to N and length that high-N words suffer a disadvantage relative to low-N words (Johnson & Pugh, 1994, Experiment 5).

The central issue, of course, is which of these conditions provides a more valid index of lexical retrieval uncontaminated by decision artifacts. More particularly, the question concerns which of the processes that are modified by stimulus environment are strategic and, in that sense, specific to the task and/or stimulus context. Johnson and Pugh (1994) argue that perceivers can exert strategic control over their use of neighborhood information such that "anytime the information provided by the status of the cohort is adequate for determining a response decision ... [they] can terminate processing" (p. 319). Thus, the facilitatory effect of N that is assumed to arise from relying on neighbor activation is attributed to a task-specific strategy, and the inhibitory effects observed in blocked conditions are claimed to reflect "true" lexical identification. Similarly, Grainger and Jacobs (1996) argue that LDT responses can be based on a measure of overall lexical activity that is not sufficient for performance in tasks requiring full identification of stimuli.

The fact that facilitatory effects of N on word classifications are enhanced in an environment of low-wordlike nonwords is consistent with the hypothesis that lexical decisions are at least sometimes based on overall lexical activity without full resolution of the target word. But, even

if it is accepted that “easy” discriminations between words and low-N nonwords might reflect the influence of overall lexical activity as a basis for lexical classification (Grainger & Jacobs, 1996), this does not justify the converse assumption that difficult discriminations will provide more valid measures of retrieval. Rather, difficult discriminations, such as those created by blocking stimuli, may cause subjects to adopt additional checking or verification procedures that are not part of “normal” lexical retrieval. Such a strategy has frequently been suggested as a possible explanation of changes in lexical decision performance with discrimination difficulty (e.g., Gordon, 1983; O’Connor & Forster, 1981). The addition of a serial verification process (e.g., Paap et al., 1982) under blocked conditions yields the same general predictions as cohort resolution: increased latency for high-N relative to low-N words and an overall reduction in error rate. Indeed, the two interpretations are conceptually indistinguishable in the sense that both implicate an additional process that is invoked under blocked, but not mixed, stimulus-context conditions and yields a linear inhibitory effect of N. However, they have completely different implications for the locus of facilitatory N effects on lexical decision, because “cohort resolution” is assumed to be an intrinsic component of word identification, but a strategically mediated verification process is task specific.

Measures of eye movements during word and sentence reading may contribute to resolving these issues. Perea and Pollatsek’s (in press) data indicated that the effects of neighbor frequency on eye-gaze data during sentence reading are restricted to measures of “late” lexical processing reflected in fixations *subsequent* to the target word. They assume that these late processes reflect “lexical completion” (Reichle et al., in press) but acknowledge that the precise nature of these processes, and their relationship to the processes involved in single-word retrieval tasks, remains unclear. The fact that the late eye-movement data parallel Perea and Pollatsek’s LDT results for the same stimuli implies that the inhibitory effects of neighbors are not due to a “double-checking process” that is specific to the LDT requirements but does not rule out the possibility that they are due to “verification processes that are [so] slow ... that they only affect lexical decisions ... when a word is relatively low in frequency” (Perea & Pollatsek, in press). Further studies of eye movements to words varying in both N and neighbor frequency in different task conditions are necessary to confirm the generality of Perea and Pollatsek’s findings and provide insight into the nature of the processes reflected in early and late eye-movement effects.

Thus, it is difficult to draw unequivocal conclusions about the influence of stimulus context on lexical decision performance. Both discriminations that are too easy and discriminations that are too difficult can be argued to be influenced by information other than unique word identification: easy tasks may allow decisions to be based on overall lexical activity instead of lexical retrieval, and difficult discriminations may cause subjects to seek additional information beyond lexical retrieval to verify their

decision. Manipulations such as blocking are even more difficult to interpret, because they allow subjects to adjust their decision criteria or the information they use to make a decision for each stimulus class.

The most relevant data are therefore those obtained when words and nonwords match on N but vary across the full range of possible neighborhood sizes characteristic of that stimulus length. Under such conditions, subjects cannot modify their strategy across trials in anticipation of particular stimuli and, because words and nonwords are matched on N, cannot base their lexical classification on neighbor activation alone. Under these standard conditions, facilitatory effects of N are consistently observed in the LDT for English stimuli. French and Spanish stimuli appear to yield null effects in these conditions, but there is no evidence in any language that a large number of activated neighbors cause inhibition.

Within Grainger and Jacobs’s (1996) model, the absence of inhibitory N effects on LDT performance implies that lexical decisions are *never* based on unique word identification. However, this conclusion rests on accepting the perceptual identification task as the touchstone for unique word identification, since this is the only task that yields consistently inhibitory effects of N, at least under single-response conditions. But, as discussed earlier, there is reason to question the relevance of the data obtained under degraded presentation conditions to an understanding of lexical selection processes in clear presentation conditions. Degraded stimuli provide a perceptual input that is compatible with a number of stimuli and therefore, not surprisingly, yield poorer identification performance for words that are similar to other words, unless these alternative responses can be eliminated through repeated guesses. There is no necessary reason that the processes involved in interpreting such ambiguous inputs are the same as those underlying selection between the multiple lexical representations activated by a single clearly presented word.

Thus, although there may be validity to Grainger and Jacobs’s (1996) argument that LDT performance can be influenced by a mechanism that is sensitive to overall lexical activity, their claim that perceptual identification reflects unique word identification is questionable. Moreover, if the perceptual identification data are an unreliable index of lexical selection, then there is no evidence of inhibitory influences of N in any task. While there may be an additional facilitatory effect on LDT performance due to overall lexical activity, the absence of inhibitory effects of N, even under difficult discrimination conditions and even in languages that show inhibitory influences of higher frequency neighbors, makes it very difficult to sustain the claim that the facilitatory effects of overall lexical activity are masking the “real” effects of lateral inhibition from activated neighbors.

Task convergence as a criterion for lexical retrieval. Rather than relying on any single task to assess lexical retrieval, the convergence of effects of a variable across tasks provides evidence that the effect has its locus in lexical identification (Balota & Chumbley, 1984; Monsell

et al., 1989). The consistency of lexical decision and naming data for mixed stimulus sets satisfies this criterion. The conclusion that facilitatory effects of N are due to lexical identification has, however, been resisted both because of the claims that they may reflect decision components of the LDT and because N effects on word naming may reflect sublexical assembly processes rather than lexical identification. The latter conclusion is supported by evidence that N facilitates nonword naming (McCann & Besner, 1987).

However, the only direct evidence on this issue is Peerman and Content's (1995) finding that N effects for French stimuli were reduced rather than enhanced when words were mixed with nonwords. This result is clearly inconsistent with the sublexical assembly account of N effects on word naming. This result needs to be replicated in English to establish its generality.

Thus, it is *possible* to explain the facilitatory effects of N in both lexical decision and naming tasks by assuming that neighbor activation has a facilitatory influence on two different task-specific processes: sublexical assembly of a pronunciation and the decision component of the LDT. There is, however, no direct evidence from either task that demonstrates that the effects are *not* due to processes intrinsic to lexical identification, and the parallel effects across tasks supports a common locus.

Forster and Shen (1996) argued that their semantic categorization data refute this interpretation. Using a semantic categorization task requiring judgments about a single category (animals), they found that "No" responses were not reliably affected by N or neighbor frequency. They therefore concluded that "the component of the LDT that is responsible for the facilitatory effects of density is missing from the categorization task," implying that "this special component must be a decision component, rather than an access component" (p. 707). To reconcile this conclusion with the evidence of neighborhood facilitation in naming tasks, Forster and Shen accepted the claim that N influences pronunciation assembly and therefore may also "have nothing to do with lexical access" (p. 708). Carreiras et al. (1997) drew similar conclusions on the basis of N effects on semantic categorization performance in Spanish.

It is not clear that either experiment justifies such a confident conclusion. In both of Forster and Shen's (1996) semantic categorization experiments, facilitatory effects of N were significant in the subjects' analysis. Although the discrepancy between the outcomes of the subjects' and items' analyses might indicate that the difference between subject means for high-N and low-N words is due to a few aberrant items (Forster & Dickinson, 1976), it might also reflect an uncontrolled item attribute that influences the decision process of the semantic categorization task and leads to variability among items that obscures the effects of N. Carreiras et al. (1997) found a complex pattern of interaction between the effects of neighborhood size and frequency that failed to confirm the predictions that they had derived from Grainger and Jacobs's (1996) claims regarding unique word identification. Thus, the empirical

evidence regarding the effects of neighborhood structure on semantic categorization performance is messy. Further investigations are necessary to justify the conclusion that this task fails to show the facilitatory effects of N that are evident in both lexical decision and naming.

More generally, there is no more reason to accept semantic categorization data as the *sine qua non* of lexical retrieval than any other paradigm. The conflict between Balota and Chumbley (1985) and Monsell et al. (1989) regarding the effects of frequency on this task highlight that it too is an imperfect measure of lexical identification. Like lexical classification, semantic categorization may be achievable without complete identification. Monsell et al. described how activation of a semantic category by the category prime might provide a basis for responses in categorization tasks and reduce sensitivity to unique lexical identification. They assumed that category priming would not influence performance when the same category judgment (person/thing) was required for all items, but it remains possible that, even without presentation of a category cue for each trial, information about the semantic category of the stimulus becomes available prior to complete identification and influences the decision process. Thus, it is just as plausible to argue that the data obtained in the semantic categorization paradigm are an unreliable index of lexical retrieval processes as it is to claim that the absence of reliable effects in the items analysis of this task proves that the robust facilitatory effects observed in LDT and naming tasks should be attributed to processes other than lexical identification.

CONCLUSIONS

This review was stimulated by the reputedly contradictory evidence regarding the effects of neighborhood structure on word identification. The summary of the evidence presented in the section on The Evidence suggests that the contradictions are not as great as has been implied. The inhibitory effects of large neighborhoods observed in the perceptual identification task most likely reflect sophisticated guessing strategies invoked to resolve partial stimulus information. Under standard clear presentation conditions in LDT and naming tasks, large neighborhoods are almost always associated with better performance even when N does not provide a reliable guide to lexicality. Although inhibitory effects of higher frequency neighbors have been observed in lexical classifications of French and Spanish words, such effects are not generally observed in English.

The most parsimonious interpretation of these findings is that the facilitatory effects of N reflect lexical identification processes common to LDT and naming tasks. However, this finding is not, so far, supported by convergent evidence from the semantic categorization task. Attributing N effects to common lexical identification processes does not imply a serial or unidirectional flow of processing, but only that early sensorily driven processing of stimuli is similar regardless of the task to be performed.

Within a symbolic framework like the interactive activation model, this might correspond to the activation of a word detector, while, in the distributed network of a PDP model, it might be defined as a stable network state or pattern of activation elicited by a stimulus acting on the learned network weights (Monsell et al., 1989; Van Orden & Goldinger, 1994).

The conclusion that N has facilitatory effects on lexical retrieval has important implications for assumptions about the representations and processes involved in visual word identification. Discussion will initially focus on the results for English words and then turn to considering the language specificity of some of the effects.

Is the Lexical Selection Process Competitive?

The conclusion that N effects are facilitatory rather than inhibitory seems to undermine the assumptions of parallel activation mechanisms of lexical retrieval. The notion of competition between similar words implemented as lateral inhibition between activated representations is central to the interactive activation framework and has provided the basis for interpretation of a variety of phenomena in the word-recognition literature. The dominance of this metaphor presumably accounts for the continued reluctance of researchers to accept the conclusion that facilitatory N effects are due to lexical retrieval even when this necessitates invoking different task-specific mechanisms to account for the same empirical outcome in different tasks.

Of course, the interactive activation framework does not require that the *net* outcome for high-N words be inhibitory. Lateral inhibition between word representations may be counteracted by excitatory activation between letters and words (Andrews, 1989, 1992). Supporting this possibility, Coltheart and Rastle (1994) found facilitatory effects of N when the interactive activation model was used as the lexical component of the DRC model, suggesting that a net facilitatory effect of N can emerge from particular parameter combinations. The conclusion that lexical retrieval is faster for words with many neighbors than for words with few neighbors does not, therefore, necessarily rule out the contribution of lateral inhibition to selecting between activated representations.

Consistent with this possibility, I have recently reported that words that have a neighbor that is identical except for the transposition of two adjacent letters (e.g., *salt/slat*; *quiet/quite*) are responded to more slowly than matched control words in both the LDT and the naming task (Andrews, 1996). This suggests that this specific form of neighbor similarity *does* reveal competition between neighbors. Importantly, regression analyses of this data set simultaneously showed inhibitory effects of transposed-letter confusability and facilitatory effects of N, demonstrating that the different effects of lexical similarity are unlikely to be due to changes in subjects' response criteria or decision strategies for different words. For example, it is difficult to see how the results could be explained by Grainger and Jacobs's (1996) multiple read-out model, because all stimuli were presented in a single block and the N and TL

effects cannot, therefore, be attributed to different response criteria. Further simulations of different versions of the interactive activation model are necessary to determine whether it is possible to find a parameter set capable of simultaneously mimicking both effects of lexical similarity.

The question of whether facilitatory neighborhood effects can be accommodated by the more recent PDP models of word-recognition models is also unclear. PDP models have neither explicit inhibitory connections between representations nor localist representations of words to compete for identification. Words are defined by patterns of activation in a distributed network. Through training, the weights on connections between the elements of the network are adjusted to allow determination of the correct output for each input in the training set. Once an optimal set of weights has been learned, a particular stimulus will elicit the same pattern of network activation whenever it is presented. "Selection" of the pattern that best matches the input is achieved when the network settles into a stable state or "attractor" corresponding to a trained pattern.

Within this framework, words with similar input features will elicit similar activation patterns because they rely on similar connection weights. These shared weights provide a potential mechanism for explaining facilitation for words from large neighborhoods because the relevant connections will have been strengthened by exposure to multiple training items. Consistent with this intuition, Seidenberg and McClelland's (1989) PDP implementation successfully simulated Andrews's (1989) facilitatory effects of N on word naming. However, it can be argued that this simulation finesses the critical issue of lexical selection. Because of computational limitations, the trained model was tested in a single sweep without the stochastic processing that would allow the model to settle into a stable attractor state in a fully interactive simulation. Performance estimates were based on the magnitude of the output error (the difference between the actual and desired output pattern). This metric relies on the assumption that stimuli that initially elicit patterns closer to the target output would also achieve resolution more quickly in a stochastic simulation. But this assumption may be invalid. Even if high-N words initially elicit more accurate output patterns than do low-N words because of the stronger connection weights associated with more frequently occurring patterns, the overlap between the patterns for similar words may mean that this initial advantage is not reflected in faster *resolution* of high-N words. That is, even though PDP models do not use a lateral inhibition mechanism to select between representations in the manner of symbolic interactive activation frameworks, the overlap between the attractors for similar words may have the functionally equivalent consequence of delaying the time taken to achieve a stable network state for high-N words.

Plaut et al. (1996) have implemented a stochastic PDP model that provides the basis for evaluating this possibility. But they have not, so far, investigated the effects of N or extended the simulation to tasks other than word naming. So it remains to be seen whether PDP models will

prove to effectively simulate facilitatory effects of N on lexical decision and naming performance.

Thus, the finding of superior performance for words with many neighbors does not necessarily disprove the involvement of lateral inhibitory mechanisms. Within the interactive activation framework, the impact of such inhibition might be obscured by excitatory activation between representations for letters and words, and in PDP models, there may be a functionally equivalent tradeoff between the stronger connections developed for more frequently occurring patterns and the overlap of the attractors for similar words.

This issue cannot be resolved by simulation alone. The interactivity of the models means that it is possible to produce a wide variety of outcomes by varying the parameters governing processing. Systematic examinations of how a model's performance changes with modifications to its architecture or training regime (e.g., Plaut et al., 1996; Plunkett & Marchman, 1991) go some way toward establishing the "design principles" of the model (Van Orden & Goldinger, 1994) and providing a basis for theoretically constraining the plausible range of parameter variations. However, the most effective constraint on model implementation is achieved by requiring that a particular implementation be capable of successfully simulating a *set* of interrelated phenomena presumed to reflect similar underlying mechanisms (Coltheart & Rastle, 1994). The facilitatory effects of N in both LDT and naming tasks, in combination with the lack of effect of neighbor frequency, and the evidence of effects of specific similarity such as that between transposed-letter confusable word pairs (Andrews, 1996) provide a powerful constraint on the precise form of interaction between excitatory and inhibitory processes that can guide evaluation of computational implementations of different models.

Grainger and Jacobs's (1996) computational simulations provide a laudable example of a systematic approach to investigating how cross-task differences might be accounted for by reliance on different mechanisms within a single interactive activation architecture. The flaw in their approach is their reliance on the perceptual identification task as the index of word identification. This leads them to reify inhibitory effects of N as markers of "true" lexical retrieval. But the present review indicates that facilitatory effects of N are the norm in English, and that even in French and Spanish N never exerts an inhibitory effect in any task except perceptual identification. There are inhibitory effects of higher frequency neighbors in French and Spanish, but they rarely occur in English.

Defining Lexical Similarity

The apparent contradictions in the evidence of effects of neighborhood structure, combined with the centrality of the notion of lateral inhibition to current models, has encouraged a focus on the implications of neighborhood effects for lexical retrieval processes. However, these issues about retrieval processes are intimately intertwined with assumptions about the form and organization of lexical

knowledge that have been relatively neglected in discussion of the implications of effects of neighborhood structure.

In all of the research discussed so far, the construct of a lexical neighbor has been operationalized according to Coltheart et al.'s (1977) N metric. The fact that manipulations of N account for significant variability in performance suggests that it captures a construct that is *relevant* to similarity between representations, but does not justify the conclusion that N is the correct or only way of conceptualizing lexical similarity.

The N metric defines neighbors as words that can be created by changing one letter of the target. This definition appears to imply two general assumptions about the representations underlying lexical retrieval: that words are coded in terms of position-specific letter representations, and that word representations are coded for length so that the cohorts of possible neighbors are restricted to words of the same length as the target. These are the assumptions built into the interactive activation model that is implemented with the even stronger assumption that letter positions are entirely independent so that there is no greater similarity between the same letter in different positions than between two different letters. These assumptions are, to a large degree, a reflection of pragmatic limitations on the process of implementing the simulation. For example, the original interactive activation simulation was restricted to a vocabulary of 4-letter words, so restricting neighbors to words of the same length was an implementational rather than a conceptual decision. However, the model's reliance on a slot-based coding scheme in which each letter position is treated independently means that extending the model to other lengths would require separate input nodes for each letter position in words of every length. Modifying this slot-coding assumption to allow for sensitivity to letter identity across different within-word positions, or to similarity between words of different lengths, would dramatically change the architecture of the model and, potentially, the effects of similarity on performance. Thus, it is not sufficient for the proponents of computational models to claim that they are not theoretically committed to a particular set of representational assumptions but have adopted them to achieve computational tractability (e.g., Seidenberg & McClelland, 1989), because the manner in which inputs are parsed and representations are organized critically determines the set of words that will be activated by a particular target.

One might be tempted to argue that the fact that performance is sensitive to N justifies the assumption of a "slot-based" input coding scheme. But this temptation must be resisted. There are other ways of capturing lexical similarity that are likely to be correlated with N, but which have very different implications for representational organization. For example, Pugh et al. (1994) found that a measure of the number of positions at which a neighbor can be formed accounts for unique variance over and above N.

More generally, the fact that neighbors are not evenly distributed across letter positions raises the possibility that N is related to other similarity constructs. For example, a

disproportionate number of 4-letter words' neighbors are created by changes at the first position. In the sample of 1,895 words described in Table 1, there were an average of 3.35 neighbors at the first position compared with 0.88, 1.43, and 1.57 for positions 2, 3, and 4, respectively. Thus, an average of 46% of 4-letter word neighbors share an orthographic body. A similar dominance by first-position neighbors is evident for 5- and 6-letter words: for 5-letter words, the average number of first-position neighbors is 0.92 and an average of 0.27 for other positions; the corresponding statistics for 6-letter words are 0.52 and 0.12.

This suggests that N effects may be due primarily to the effects of neighbors that share a body with the target word. There is considerable evidence that bodies play an important role in word recognition (see Treiman, 1992, for a review), so N effects may, in part, reflect correlations between N and body-defined neighborhoods. If effects of N were really due primarily to items with common bodies, they might implicate representational assumptions that were relatively different from those reflecting sensitivity to neighbors at all letter positions. Taft (1991), for example, has suggested supplementing the interactive activation model with a level of body representations.

One way of interpreting the correlation between N and body neighborhoods is to assume that N effects are really due to phonological rather than orthographic structure. Body neighbors usually rhyme, so N effects may be attributable to phonological rather than orthographic similarity. This is not to say that neighborhood effects are simply a surrogate of effects of phonological consistency. Presumably high-N words are more likely to have inconsistent neighbors, yet they are responded to more quickly than are words with few neighbors.

Rather, body units might reflect the interface between orthography and phonology. Treiman, Mullennix, Bijeljac-Babic, & Richmond-Welty (1995) reported a systematic analysis of the orthographic-phonological relationships embodied in English monosyllabic words, which showed both that orthographic bodies were more redundant than other within-word units and that the pronunciation of ambiguous vowel units in monosyllabic words were more strongly predicted by the consonants that followed them than by those that preceded them. That is, body units are more consistently pronounced than are either vowels alone or C(C)V units. In English, then, orthographic body units are useful functional units because they provide systematic cues to inconsistent pronunciations. Since most neighbors are body neighbors, the neighborhood advantage may reflect the functional status of body units in word identification.

Cross-Language Differences in Lexical Similarity

The speculation that neighborhood effects in English may be due, at least in part, to words with similar bodies suggests a possible explanation of the language specificity of neighborhood effects. The level of phonological inconsistency in French and Spanish is much lower than in Eng-

lish. Spanish is a highly regular language with a shallow relationship between orthography and phonology, so there may be little need for the reader to develop sensitivity to relationships higher than the grapheme-phoneme level to determine the mapping from orthography to phonology. Although French has a deeper, or more opaque, mapping relationship, it has been estimated that 95% of French words are consistent in the sense that they could be correctly pronounced using context-sensitive grapheme-phoneme rules (Gak, 1976).

Facilitatory effects of neighborhood size in English may arise because orthographic bodies play a more important role in lexical retrieval than they do in languages with a more consistent orthographic-to-phonological mapping. There is considerable evidence that young children learning English acquire sensitivity to the rime and body units of words in the course of reading development (e.g., Treiman, 1992). The consistency of pronunciation of body units remains a significant predictor of adult naming performance and accounts for variability over and above that due to grapheme-phoneme correspondences (Treiman et al., 1995).

Treiman et al. (1995) argue that rime/body units come to play an important role in reading English words because the "lexical statistics" of the language are such that they provide "a potential way for readers to deal with the vagaries of the English writing system" (p. 130). Their analysis of monosyllabic English words showed that vowels have a wider variety of pronunciations than consonants, and that the pronunciations of vowels are better predicted by the final than by the initial consonants. Thus, the recognition of body units "helps to regularize the links between spelling and sound in the English writing system" (p. 130). Consistent with the view that bodies play a role independently of grapheme-phoneme regularity or consistency, Treiman, Goswami, and Bruck (1990) reported that both children and adults pronounce nonwords like *tain* and *goach*, which have many body neighbors, more accurately than they do nonwords like *goan* and *taich*, in which the bodies are uncommon, even when the items are constructed from the same graphemes and are both regular and consistent.

Thus, facilitatory N effects in English may be primarily due to the influence of words that share a body with the target word. The "special role" (Treiman et al., 1995) played by these neighbors may be relatively specific to English because it reflects particular properties of the orthographic-phonological mapping of the language: the differential ambiguity of vowels and consonants, and the greater statistical redundancy of body as compared with initial CV units. Languages with a shallower orthography or with a different statistical structure would not necessarily lead to the emergence of bodies as functional units. For example, the ambiguities in the orthographic-phonological mapping in French are different from those in English: the pronunciations of final consonants are more ambiguous than those of vowels, and body/rime units make little contribution to resolving inconsistencies because of

the predominance of open syllables (Content, 1991). Consistent with the view that bodies do not play the important role in processing French that they do in English, Taft and Radeau (1995) reported that French subjects' pronunciations of ambiguous nonwords did not show the influence of body units that is characteristic in English (Andrews & Scarratt, in press; Patterson & Morton, 1985) but appeared instead to be based on a syllabic parsing strategy.

The differences between the orthographic-phonological mapping of Spanish and English are even more salient. Besides its shallow mapping relationship, Spanish is a syllabically structured language and almost all words have at least two syllables. Thus, all of the words in the sample used by Carreiras et al. (1997) were bisyllabic even though they had only 4 or 5 letters. Multisyllabic words have rarely been included in English investigations of neighborhood structure. Given these properties of Spanish, it seems likely that syllabic units may play a more important role than body units. Consistent with this possibility, Perea and Carreiras (in press) reported that LDT and naming performance for Spanish stimuli showed inhibitory effects of *syllable* neighbors. Thus, like French, the determinants of similarity between different words may be different from those that characterize English.

To some degree, this speculation as to the basis of the cross-language differences in the effects of orthographic neighborhood structure rests on the assumption that sensitivity to body units arises as a consequence of exposure to orthography rather than because of the phonological knowledge that children acquire through spoken language experience. Rimes are seen as the natural phonological constituent of the spoken syllable and are therefore salient in spoken words across all languages (Treiman, 1988). Sensitivity to body units might, therefore, be expected to arise in all languages that systematically map between orthography and phonology, because of a tendency to parse the orthographic form into units corresponding to its phonological constituents. Thus, if body units play a special role in English, it must arise because of the specific mapping of orthography to phonology and therefore depend on exposure to orthography.

Treiman et al.'s (1990; Treiman et al., 1995) investigations of this issue suggest that sensitivity to body units is evident very early in reading development—at least by the end of first grade. But they acknowledge that their evidence does not conclusively demonstrate that body units “derive their importance from rime units in spoken words” (Treiman et al., 1990, p. 565). In any case, even if it is the salience of phonological rimes that triggers attention to body units, it remains possible that these units acquire special status only when they facilitate the efficiency of mapping from orthography to phonology. This may be more true for English than for other alphabetic languages.

Thus, the apparent differences in the facilitatory effects of N in English by comparison with French and Spanish may arise because orthographic redundancy plays a greater role in English because of its higher level of phonological

inconsistency. One level of orthographic redundancy that is useful in English is captured by body units, and, because N is highly correlated with number of body neighbors, this metric successfully accounts for variability in performance. It is a less effective predictor of performance in other languages either because shallower orthographies do not require the same degree of use of orthographic redundancy to achieve successful retrieval and/or because the N metric is a less effective summary of the dimensions of orthographic redundancy that *are* relevant to that language. Investigations of the lexical statistics of languages like French and Spanish, comparable to Treiman et al.'s (1995) analysis of English, and systematic investigations of sensitivity to body units in these languages, are needed to evaluate these speculations.

The preceding arguments that body units play a special role in processing English might be taken as support for “multiple level” models that assume a representational level corresponding to the body that intervenes between the letter or letter cluster and the word level (Taft, 1991). However, this is not a necessary assumption. “Body units” may be an emergent property of a distributed representational system that also allows for the representation of any other systematic relationships between orthography and phonology (Seidenberg & McClelland, 1989). Moreover, any such representational level may be superfluous in a system that parses words into overlapping multiletter units such as “Wickelgraphs” (Wickelgren, 1969) or syllabic segments such as onsets, vowels, and codas (Plaut et al., 1996). Converging empirical and computational approaches are needed to attempt to evaluate the validity and implications of these different solutions. Constraints derived from investigations of neural mechanisms may contribute to this pursuit (McClelland, McNaughton, & O'Reilly, 1995).

Generality of Neighborhood Effects

A final issue that requires consideration concerns how much of an impact neighborhood structure is likely to have on visual word recognition. As demonstrated in Table 1, there is a strong relationship between N and word length, because words with more than 4 letters have very few neighbors. The majority of the research on effects of neighborhood structure has focused on 4-letter words because, in a sense, N affects only short words—long words do not have neighbors by the N definition. It might, therefore, be claimed that determining the basis of these effects is of little relevance to understanding word recognition, because the majority of words do not have neighbors. There are a number of arguments against this view.

First, even if it were accepted that only short words were susceptible to influence from neighbors, the fact that these words dominate text means that they *do* comprise a substantial proportion of the words a reader encounters.

Second, the validity of the conclusion that only short words are vulnerable to the influence of neighbors depends on how similarity is defined. If, for example, words

were parsed into syllabic units early in processing (Spoehr & Smith, 1973; Taft, 1979), then long words might be susceptible to influence from syllabically defined neighbors via mechanisms similar to those underlying N effects for monosyllabic words.

Third, and most critically, even if neighborhood effects are restricted to short words, their importance lies not in how often these effects would impact on reading, but rather in the constraints they impose on the lexical retrieval process. Short words present the word-recognition system with its most difficult selection task because these words are necessarily more similar to each other than are longer words, at least at the level of shared letters. Thus, the organization of the word-recognition system must, to a degree, have been driven by the need to make these most difficult discriminations between lexical symbols. In the same way that the need to discriminate between similar words may have shaped the human word-recognition system, empirical data on the effects of lexical similarity provide essential constraints to govern the development of models of this process.

REFERENCES

- ANDREWS, S. (1989). Frequency and neighborhood effects on lexical access: Activation or search? *Journal of Experimental Psychology: Learning, Memory, & 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, & Cognition*, **18**, 234-254.
- ANDREWS, S. (1996). Lexical retrieval and selection processes: Effects of transposed-letter confusability. *Journal of Memory & Language*, **35**, 775-800.
- ANDREWS, S., & SCARRATT, D. R. (in press). Rule and analogy mechanisms in pronouncing nonwords: Hough dou peapel gnew wirts. *Journal of Experimental Psychology: Human Perception & Performance*.
- BALOTA, D. A. (1990). The role of meaning in word recognition. In D. A. Balota, G. B. Flores d'Arcais, & K. Rayner (Eds.), *Comprehension processes in reading* (pp. 9-32). Hillsdale, NJ: Erlbaum.
- BALOTA, D. A., & CHUMBLEY, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception & Performance*, **10**, 340-357.
- BALOTA, D. A., & CHUMBLEY, J. I. (1985). The locus of word-frequency effects in the pronunciation task: Lexical access and/or production? *Journal of Memory & Language*, **24**, 89-106.
- BALUCH, B., & BESNER, D. (1991). Visual word recognition: Evidence for strategic control of lexical and nonlexical routes in oral reading. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **17**, 644-652.
- BURNAGE, D. (1998). *CELEX: A guide for users*. Nijmegen: Centre for Lexical Information.
- CARREIRAS, M., PEREA, M., & GRAINGER, J. (in press). Effects of orthographic neighborhood in visual word recognition: Cross-task comparisons. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **23**, 857-871.
- COLOMBO, L. (1986). Activation and inhibition with orthographically similar words. *Journal of Experimental Psychology: Human Perception & Performance*, **12**, 226-234.
- COLTHEART, M. (1978). Lexical access in simple reading tasks. In G. Underwood (Ed.), *Strategies of information processing* (pp. 151-216). New York: Academic Press.
- COLTHEART, M., CURTIS, B., ATKINS, P., & HALLER, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review*, **100**, 589-608.
- COLTHEART, M., DAVELAAR, E., JONASSON, J. T., & BESNER, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI* (pp. 535-555). Hillsdale, NJ: Erlbaum.
- COLTHEART, M., & RASTLE, K. (1994). Serial processing in reading aloud: Evidence for dual-route models of reading. *Journal of Experimental Psychology: Human Perception & Performance*, **20**, 1197-1211.
- CONTENT, A. (1991). The effect of spelling-to-sound regularity on naming in French. *Psychological Research*, **53**, 3-12.
- FORSTER, K. I. (1976). Accessing the mental lexicon. In R. Wales & E. Walker (Eds.), *New approaches to language mechanisms* (pp. 257-287). Amsterdam: North-Holland.
- FORSTER, K. I., & DAVIS, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **10**, 680-698.
- FORSTER, K. I., & DICKINSON, R. G. (1976). More on the language-as-fixed-effect fallacy: Monte Carlo estimates of error rates for F1, F2, F and min F. *Journal of Verbal Learning & Verbal Behavior*, **15**, 135-142.
- FORSTER, K. I., & SHEN, D. (1996). No enemies in the neighborhood: Absence of inhibitory effects in lexical decision and semantic categorization. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **22**, 696-713.
- FRAUENFELDER, U. H., BAAYEN, R. H., HELLWIG, F. M., & SCHREUDER, R. (1993). Neighborhood density and frequency across languages and modalities. *Journal of Memory & Language*, **32**, 781-804.
- GAK, V. G. (1976). *L'orthographe du francais* [French orthography]. Paris: SELAF.
- GORDON, B. (1983). Lexical access and lexical decision: Mechanisms of frequency sensitivity. *Journal of Verbal Learning & Verbal Behavior*, **22**, 24-44.
- GRAINGER, J. (1990). Word frequency and neighbourhood frequency effects in lexical decision and naming. *Journal of Memory & 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., O'REGAN, J. K., JACOBS, A. M., & SEGUI, J. (1992). Neighborhood frequency effects and letter visibility in visual word recognition. *Perception & Psychophysics*, **51**, 49-56.
- GRAINGER, J., & SEGUI, J. (1990). Neighborhood frequency effects in visual word recognition. *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 & Performance*, **18**, 1174-1188.
- JACOBS, A. M., & GRAINGER, J. (1994). Models of visual word recognition—Sampling the state of the art. *Journal of Experimental Psychology: Human Perception & Performance*, **20**, 1311-1334.
- JOHNSON, N. F., & PUGH, K. R. (1994). A cohort model of visual word recognition. *Cognitive Psychology*, **26**, 240-346.
- LANDAUER, T. K., & STREETER, L. A. (1973). Structural differences between common and rare words: Failure or equivalence assumptions for theories of word recognition. *Journal of Verbal Learning & Verbal Behavior*, **12**, 119-131.
- LAXON, V. J., COLTHEART, V., & KEATING, C. (1988). Children find friendly words friendly too: Words with many orthographic neighbours are easier to read and spell. *British Journal of Educational Psychology*, **58**, 103-119.
- LAXON, V., MASTERTON, J., POOL, M., & KEATING, C. (1992). Nonword naming: Further exploration of the pseudohomophone effect in terms of orthographic neighborhood size, graphemic changes, spelling-sound consistency, and reader accuracy. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 730-748.
- MASSARO, D. W., TAYLOR, G. A., VENEZKY, R. L., JASTRZEMBSKI, J. E., & LUCAS, P. A. (1980). *Letter and word perception: Orthographic structure and visual processing in reading*. New York: North-Holland.

- MCCANN, R. S., & BESNER, D. (1987). Reading pseudohomophones: Implications for models of pronunciation assembly and the locus of word-frequency effects in naming. *Journal of Experimental Psychology: Human Perception & Performance*, **13**, 14-24.
- MCCLELLAND, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review*, **86**, 287-330.
- MCCLELLAND, J. L., MCNAUGHTON, B. L., & O'REILLY, R. C. (1995). Why are there complementary learning mechanisms in the hippocampus and neocortex? Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, **102**, 419-437.
- 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.
- MICHIE, P. T., COLTHEART, M., LANGDON, R., & HALLER, M. (1994). *Effects of orthographic neighborhood size on visual word recognition: Behavioral, electrophysiological and computational evidence*. Unpublished manuscript, Macquarie University.
- MONSELL, S., DOYLE, M. C., & HAGGARD, P. N. (1989). Effects of frequency on visual word recognition tasks: Where are they? *Journal of Experimental Psychology: General*, **118**, 43-71.
- MONSELL, S., PATTERSON, K. E., GRAHAM, A., HUGHES, C. H., & MILROY, R. (1992). Lexical and sublexical translation of spelling to sound: Strategic anticipation of lexical status. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 452-467.
- MORTON, J. (1970). A functional model for memory. In D. A. Norman (Ed.), *Models of human memory* (pp. 203-254). New York: Academic Press.
- O'CONNOR, R. E., & FORSTER, K. I. (1981). Criterion bias and search sequence bias in word recognition. *Memory & Cognition*, **9**, 78-92.
- 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 & Performance*, **26**, 1129-1157.
- 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.
- PATTERSON, J., & MORTON, K. E. (1985). From orthography to phonology: A new attempt at an old interpretation. In K. E. Patterson, J. C. Marshall, & M. Coltheart. (Eds.), *Surface dyslexia* (pp. 335-359). Hillsdale, NJ: Erlbaum.
- PEEREMAN, R., & CONTENT, A. (1995). Neighborhood size effect in naming words and pseudowords. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 409-421.
- PEREA, M., & CARREIRAS, M. (in press). Effects of syllable frequency and syllable neighborhood frequency in visual word recognition. *Journal of Experimental Psychology: Human Perception & Performance*.
- PEREA, M., & POLLATSEK, A. (in press). The effects of neighborhood frequency in reading and lexical decision. *Journal of Experimental Psychology: Human Perception & Performance*.
- 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.
- PLUNKETT, K., & MARCHMAN, V. (1991). U-shaped learning and frequency effects in a multi-layered perception: Implications for child language acquisition. *Cognition*, **38**, 43-102.
- PUGH, K. R., REXER, K., & KATZ, L. (1994). Evidence of flexible coding in visual word recognition. *Journal of Experimental Psychology: Human Perception & Performance*, **20**, 807-825.
- REICHLER, E., POLLATSEK, A., FISHER, D. L., & RAYNER, K. (in press). Towards a model of eye movements in reading. *Psychological Review*.
- RUBENSTEIN, H., GARFIELD, L., & MILLIKAN, J. A. (1970). Homographic entries in the internal lexicon. *Journal of Verbal Learning & Verbal Behavior*, **9**, 487-494.
- SEARS, C. R., HINO, Y., & LUPKER, S. J. (1995). Neighborhood size and neighborhood frequency effects in word recognition. *Journal of Experimental Psychology: Human Perception & Performance*, **21**, 876-900.
- SEGUI, J., & GRAINGER, J. (1990). Priming word recognition with orthographic neighbors: The effects of relative prime-target frequency. *Journal of Experimental Psychology: Human Perception & Performance*, **16**, 65-76.
- SEIDENBERG, M. S., & MCCLELLAND, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, **96**, 523-568.
- SHULMAN, H. G., HORNAK, R., & SANDERS, E. (1978). The effects of graphemic, phonetic, and semantic relationships on access to lexical structures. *Memory & Cognition*, **6**, 115-123.
- SNODGRASS, J. G., & MINZER, M. (1993). Neighborhood effects in visual word recognition: Facilitatory or inhibitory? *Memory & Cognition*, **21**, 247-266.
- SPOEHR, K. T., & SMITH, E. E. (1973). The role of syllables in perceptual processing. *Cognitive Psychology*, **104**, 21-34.
- TAFT, M. (1979). Lexical access via an orthographic code: The basic orthographic syllabic structure (BOSS). *Journal of Verbal Learning & Verbal Behavior*, **18**, 21-39.
- TAFT, M. (1991). *Reading and the mental lexicon*. Hillsdale, NJ: Erlbaum.
- TAFT, M., & RADEAU, M. (1995). The influence of the phonological characteristics of a language on the functional units of reading: A study of French. *Canadian Journal of Experimental Psychology*, **49**, 330-346.
- TREIMAN, R. (1988). The internal structure of syllables. In G. Carlson & M. Tanenhaus (Eds.), *Linguistic structure in language processing* (pp. 27-52). Dordrecht: Kluwer.
- TREIMAN, R. (1992). The role of intrasyllabic units in learning to read and spell. In P. B. Gough, L. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp. 65-106). Hillsdale, NJ: Erlbaum.
- TREIMAN, R., GOSWAMI, U., & BRUCK, M. (1990). Not all nonwords are alike: Implications for reading development and theory. *Memory & Cognition*, **18**, 559-567.
- TREIMAN, R., MULLENNIX, J., BIJELJAC-BABIC, R., & RICHMOND-WELTY, E. D. (1995). The special role of rimes in description, use, and acquisition of English orthography. *Journal of Experimental Psychology: General*, **124**, 107-136.
- VAN ORDEN, G. C., & GOLDINGER, S. D. (1994). Independence of form and function in cognitive systems explains perception of printed words. *Journal of Experimental Psychology: Human Perception & Performance*, **20**, 1269-1291.
- WICKELGREN, W. A. (1969). Context-sensitive coding, associative memory, and serial order in (speech) behavior. *Psychological Review*, **76**, 1-15.

NOTES

1. A paper by Paap and Johansen (1994) was not included because, although it reported regression analyses including N and neighbor frequency as a predictor of lexical decision latency, the data were obtained in an unusual procedure in which lexical decisions were delayed until after an intervening letter-identification judgment. The lexical decision responses made to words presented more than 2 sec previously and after an intervening response are an unreliable index of on-line lexical retrieval.

2. Wordlikeness was defined by assessing the amount of lexical activity generated by each nonword in the interactive activation simulation.

(Manuscript received August 20, 1996;
revision accepted for publication June 2, 1997.)