An Electrophysiological Study of the Effects of Orthographic Neighborhood Size on Printed Word Perception

Phillip J. Holcomb[1], **Jonathan Grainger**[2], **and Tim O'Rourke**[1]

Abstract

■ In two experiments participants read words and pseudowords that belonged to either large or small lexical neighborhoods while event-related brain potentials (ERPs) were recorded from their scalps. In Experiment 1, participants made speeded lexical decisions to all items, while in Experiment 2 they engaged in a go/no-go semantic categorization task in which the critical items did not require an overt behavioral response. In both experiments, words and pseudowords produced a consistent pattern of ERP effects: items with many lexical neighbors (large neighborhoods) generated larger N400s than similar items with relatively fewer lexical neighbors (small neighborhoods). Reaction time (RT, Experiment 1), on the other hand, showed a different pattern consistent with previous behavioral studies. While words tended to produce a facilitation in RT for larger neighborhoods, pseudowords produced an inhibition effect. The findings are discussed in terms of recent theories of word recognition and the functional significance of the N400. ■

INTRODUCTION

In one of the earliest studies of the effects of neighborhood size in printed word perception, Coltheart, Davelaar, Jonasson, and Besner (1977) demonstrated that the time to reject a string of letters as not being a word in the lexical decision task (speeded word/pseudoword classification) was related to the number of real words that it resembled orthographically; pseudowords that resembled a relatively large number of real words (i.e., came from large lexical neighborhoods), took longer than pseudowords with relatively few real word neighbors. Andrews (1989) replicated this pseudoword neighborhood size effect and also found neighborhood effects for real words, but only when word frequency was relatively low. Interestingly, unlike pseudowords, words from dense neighborhoods were actually categorized as words more quickly than words from sparser neighbors. In other words, there was a facilitatory effect of neighborhood size on words and an inhibitory effect for pseudowords. This precise pattern of effects obtained in the lexical decision task has since been replicated in several different studies (Carreiras, Perea, & Grainger, 1997; Forster & Shen, 1996; Grainger & Jacobs, 1996; Sears, Hino, & Lupker, 1995; Johnson & Pugh, 1994; Andrews, 1989, 1992).

At a general level of theorizing, it is often assumed that the presentation of a word from a large neighborhood results in activation of the target item's representation in memory as well as the representations of other items in the lexical neighborhood. So, for example, when a word such as "dime" is encountered, it results in activation in the lexical representation for "dime" itself, as well as partial activation for the representations of "time," "lime," "mime," "dine," "dame," "dome," etc. More debate has been generated around the issue of what precise mechanism underlies the facilitatory effect of neighborhood density observed with word stimuli in the lexical decision task. Andrews (1989, 1992, 1997), for example, has argued that the facilitatory effect of neighborhood size on responses to word stimuli reflects basic processes in printed word perception. One possible mechanism is word-to-letter feedback: Words with many neighbors produce higher levels of resonance between word and letter representations, hence facilitating processing of these stimuli. Advocates of this and similar accounts (e.g., Sears, Lupker, & Hino, 1999) explain the inhibitory effect observed for pseudoword stimuli in terms of an increased "word-likeness" of pseudowords with large numbers of word neighbors. The more word-like a pseudoword, the harder it will be to respond negatively to such stimuli (although the authors in question typically do not describe a specific mechanism for this).

Grainger and Jacobs (1996) offered an alternative explanation for the observed dissociation in neighborhood effects to word and pseudoword stimuli in the lexical decision task. These authors argued that the

---

[1] Tufts University, [2] CNRS, Aix-Marseille, France

facilitatory effect of neighborhood size on responses to word stimuli reflects the operation of a mechanism that is specific to the lexical decision task. In line with Andrews' account, Grainger and Jacobs assume that words with a large neighborhood lead to the partial activation of a large number of word representations in memory. Novel in this account, however, is the idea that lexical decision responses could be based on a measure of the summed activation of all positively activated word representations, a notion similar to the familiarity evaluation mechanism in Balota and Chumbley's (1984) model of lexical decision. More precisely, Grainger and Jacobs suggested that participants in a speeded lexical decision experiment could use either of two decision criteria to trigger positive responses: a standard criterion based on activity in individual word representations and a criterion set on global lexical activity.

This theoretical account of variations in neighborhood effects across tasks was implemented as the multiple read-out model within the general framework of Rumelhart and McClelland's (1982) interactive activation model. Simulations run with the multiple read-out model have shown that responses based on individual word unit activity generally show inhibitory effects of neighborhood density: Words with large numbers of orthographic neighbors suffer more within-level lateral inhibition from their simultaneously activated neighbors, thus slowing down the processing of such words. On the other hand, responses based on global lexical activity show facilitation, because words with many neighbors generate high levels of global lexical activation. Thus, in the lexical decision task, subjects can use the extra activity associated with partial activation of all items in the neighborhood to speed their "yes" word response. For small neighborhoods, less activity in the neighborhood translates into slower reaction times (RTs) because of lower overall lexical activity. According to Grainger and Jacobs (1996), the same mechanism that causes RTs to be faster for large neighborhood words also causes pseudowords from large word neighborhoods to be responded to more slowly. This is because pseudowords with many real word neighbors also generate substantial lexical activity. In the multiple read-out model, summed lexical activity is used to set a negative response criterion in the form of a variable deadline mechanism (respond "no" if a yes response has not been triggered before a given time limit). Higher global activation in early phases of stimulus processing generates a longer deadline, hence producing longer correct negative RTs to pseudoword stimuli with more orthographic neighbors.

The opposing pattern of neighborhood size effects for word and pseudoword stimuli in the lexical decision task is a well-documented, robust finding (e.g., Carreiras et al., 1997; Forster & Shen, 1996; Grainger & Jacobs, 1996; Sears et al., 1995; Johnson & Pugh, 1994; Snodgrass & Mintzer, 1993; Andrews, 1989, 1992). One important aspect of the multiple read-out model's account of this dissociation is that the same core phenomenon (e.g., heightened lexical activation with increasing numbers of word neighbors) can lead to distinct patterns of behavioral effects via task-specific mechanisms that translate core processes into task-relevant responses. Thus words and pseudowords in a lexical decision task can lead to diametrically opposed effects by the way global lexical activation is used to set positive and negative response criteria for that specific task. Furthermore, variations in effects of neighborhood size obtained to the same set of word stimuli tested in different experimental tasks (e.g., Forster & Shen, 1996) can also be captured within the same theoretical framework via task-specific response-generation mechanisms operating on the same core processes. It is this general hypothesis that is put to test in the present study.

EXPERIMENT 1

In Experiment 1, we first sought further evidence for Grainger and Jacobs' (1996) hypothesis that the facilitatory effects of neighborhood size on RTs to words and the inhibitory effects on RTs to pseudowords in the lexical decision task are due to the same mechanism: enhanced lexical activity when neighborhood size is large. However, rather than using RT as the principle dependent variable, we instead recorded event-related brain potentials (ERPs) to words and pseudowords while participants classified letter strings as words or pseudowords. We reasoned that if neighborhood size affects the processing of words and pseudowords in a similar manner, then a measure of word processing that is less sensitive to strategic or decision-related factors might produce a consistent pattern of effects across words and pseudowords.

There is reason to believe that one ERP component, the N400, might fit the bill for such a measure. Prior work with this ERP component has shown that while it is very sensitive to semantic aspects of word processing (e.g., Kutas & Hillyard, 1980, 1984), it is relatively insensitive to decision or response strategies that frequently influence RT (e.g., Kounios & Holcomb, 1992). Of particular relevance here is the observation that the N400 is larger whenever a word is associated with more semantic information. For example, Kounios and Holcomb (1992) demonstrated that N400 amplitude was larger when target words had more semantic associations (see also Holcomb, Kounios, Anderson, & West, 1999). Extended to the case of neighborhood size, one prediction is that words from larger neighborhoods should result in greater semantic activation and therefore generate larger N400s. This could occur because in addition to activating their own semantic representations, word stimuli might also partially activate the semantic representations of their orthographic neighbors. This difference in total semantic activation should be reflected in the size of the N400—larger N400s are needed for more activation. It should be noted that for the present purposes we need not draw a distinction between arguments based at the level of word forms (i.e., orthographic whole-word representations) and arguments based at the level of semantic representations. In a cascaded activation network, such as the multiple read-out model, these two measures will be highly correlated.

Although they have no semantic representation of their own, word-like pseudowords might partially activate the semantic representations of their real-word lexical neighbors. If this occurs then, by the same logic as for words, pseudowords from large neighborhoods should also produce larger N400s than pseudowords from small neighborhoods. Indirect evidence consistent with this latter hypothesis already exists. Several studies (e.g., Holcomb & Neville, 1990; Rugg & Nagy, 1987) have reported that word-like pseudowords elicit large N400s. However, this observation has always been somewhat difficult to rectify with semantic explanations of the functional significance of the N400. Why should items without semantic representations produce sizable N400s? In light of the neighborhood hypothesis, the most parsimonious explanation is that N400s to pseudowords reflect activation of semantic information associated with real word neighbors. Therefore, finding a neighborhood size effect on the N400 to pseudowords would also provide additional evidence for the semantic interpretation of pseudoword N400s.

This experiment manipulated neighborhood size for low-frequency words and pronounceable pseudowords in the lexical decision task. Overt responses were collected so that ERPs could be compared directly to behavioral data and discussed in relation to the standard findings discussed above. Word and pseudowords came from either large neighborhoods or small neighborhoods. Neighborhoods were defined using the *N*-metric proposed by Coltheart et al. (1977). That is, a lexical item was considered a neighbor if it shared the same number of letters and was identical to the target except for one of the *N* letters. Although this is only one possible operational definition of orthographic similarity, it has been used extensively in the literature and therefore seemed the most appropriate definition to use in this initial study. However, because the number of neighbors for a word decreases as word length increases (Forster, 1987), we employed only four- and five-letter words (in line with the vast majority of studies reported in the literature).

Results

Behavioral Findings

Mean RT and proportion of errors were analyzed separately. RT and proportion of error means and standard deviations are displayed in Table 1. Overall, 8% of the trials resulted in errors. In the RT analysis, all three factors produced main effects: stimulus type, $F(1,23) = 38.62$, $p < .00001$, $MSE = 10810$; neighborhood size, $F(1,23) = 23.38$, $p < .0001$, $MSE = 225$; number of letters, $F(1,23) = 13.23$, $p < .005$, $MSE = 616$. Overall, RT was faster for words than for pseudowords (mean = 693 vs. 787 msec). Stimuli with large neighborhoods had faster RTs than stimuli with small neighborhoods (mean = 735 vs. 745 msec). Stimuli with four letters had faster RTs than stimuli with five letters (mean = 733 vs. 746 msec). These effects were qualified by a significant interaction between stimulus type and neighborhood size, $F(1,23) = 59.72$, $p < .00001$, $MSE = 866$. Simple effect tests examining neighborhood size within levels of stimulus type indicated that words from large neighborhoods were responded to significantly more quickly than words from small neighborhoods, $F(1,23) = 73.16$, $p < .00001$, $MSE = 307$. Conversely, pseudowords from large neighborhoods were responded to significantly more slowly than pseudowords from small neighborhoods, $F(1,23) = 25.16$, $p < .00001$, $MSE = 238$. No other interactions were significant.

Table 1. Behavioral Data

	Reaction Time		*Proportion Errors*	
	Mean	*SD*	*Mean*	*SD*
Four-letter words, large *N*	668	80.7	0.10	0.04
Four-letter words, small *N*	708	89.0	0.17	0.09
Five-letter words, large *N*	676	84.4	0.09	0.05
Five-letter words, small *N*	724	104.6	0.12	0.10
Four-letter pseudowords, large *N*	796	122.7	0.04	0.06
Four-letter pseudowords, small *N*	764	111.0	0.04	0.04
Five-letter pseudowords, large *N*	800	132.4	0.09	0.06
Five-letter pseudowords, small *N*	788	117.4	0.06	0.05

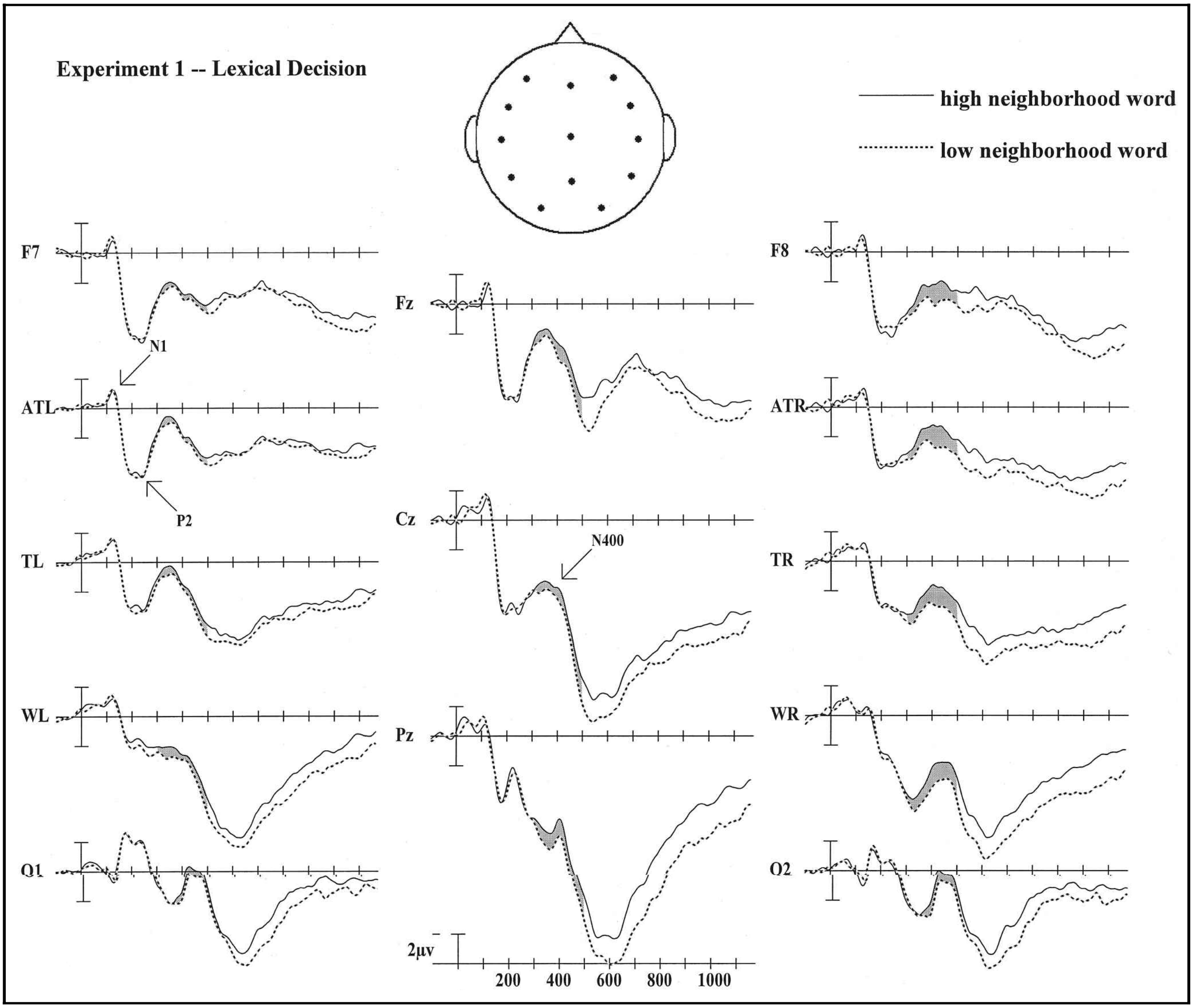

Figure 1. Plotted in this figure are the grand average ERPs from 24 subjects for words stimuli in Experiment 1. The solid lines are ERPs from large neighborhoods (i.e., high neighborhood density) and the dashed line is for ERPs from small neighborhoods (low neighborhood density). Note that stimulus onset is represented by the vertical microvolt calibration bar and that negative voltages are plotted in the upward direction.

The error analysis (see Table 1) revealed one main effect, stimulus type, $F(1,23) = 22.04$, $p < .00001$, $MSE = .03$. More errors were made to words than to pseudowords (0.12 vs. 0.06). All of the two-way interactions were significant. Neighborhood size and stimulus type interacted, $F(1,23) = 61.09$, $p < .00001$, $MSE = .001$, in a way that was consistent with the RT interaction. Simple effects showed pseudowords from large neighborhoods had more errors than pseudowords from small neighborhoods, $F(1,23) = 6.32$, $p < .05$, $MSE = .001$. Words had more errors for small neighborhoods than for large neighborhoods, $F(1,23) = 15.28$, $p < .005$, $MSE = .001$. Neighborhood size and number of letters also interacted, $F(1,23) = 5.16$, $p < .05$, $MSE = .01$. This was due to relatively more errors occurring to small than larger neighborhood stimuli for four letter compared to five letter items, $F(1,23) = 7.71$, $p < .05$, $MSE = .001$. Finally, the number of letters and stimulus type interacted, $F(1,23) = 19.10$, $p < .00001$, $MSE = .01$, because of a bigger difference in error between words and pseudowords for four-letter stimuli, $F(1,23) = 43.57$, $p < .0001$, $MSE = .01$, than for five-letter stimuli, $F(1,23) = 4.53$, $p < .05$, $MSE = .01$. The three-way interaction was not significant.

Electrophysiological Findings

The ERP grand mean waveforms for words are plotted in Figure 1 while the grand mean waveforms for pseudowords are plotted in Figure 2. Approximately 10% (mean = 9.55) of the correct trials were rejected because of eye blinks, horizontal eye movement, or amplifier blocking. For all the ERPs, the first visible component was a negative-going deflection occurring

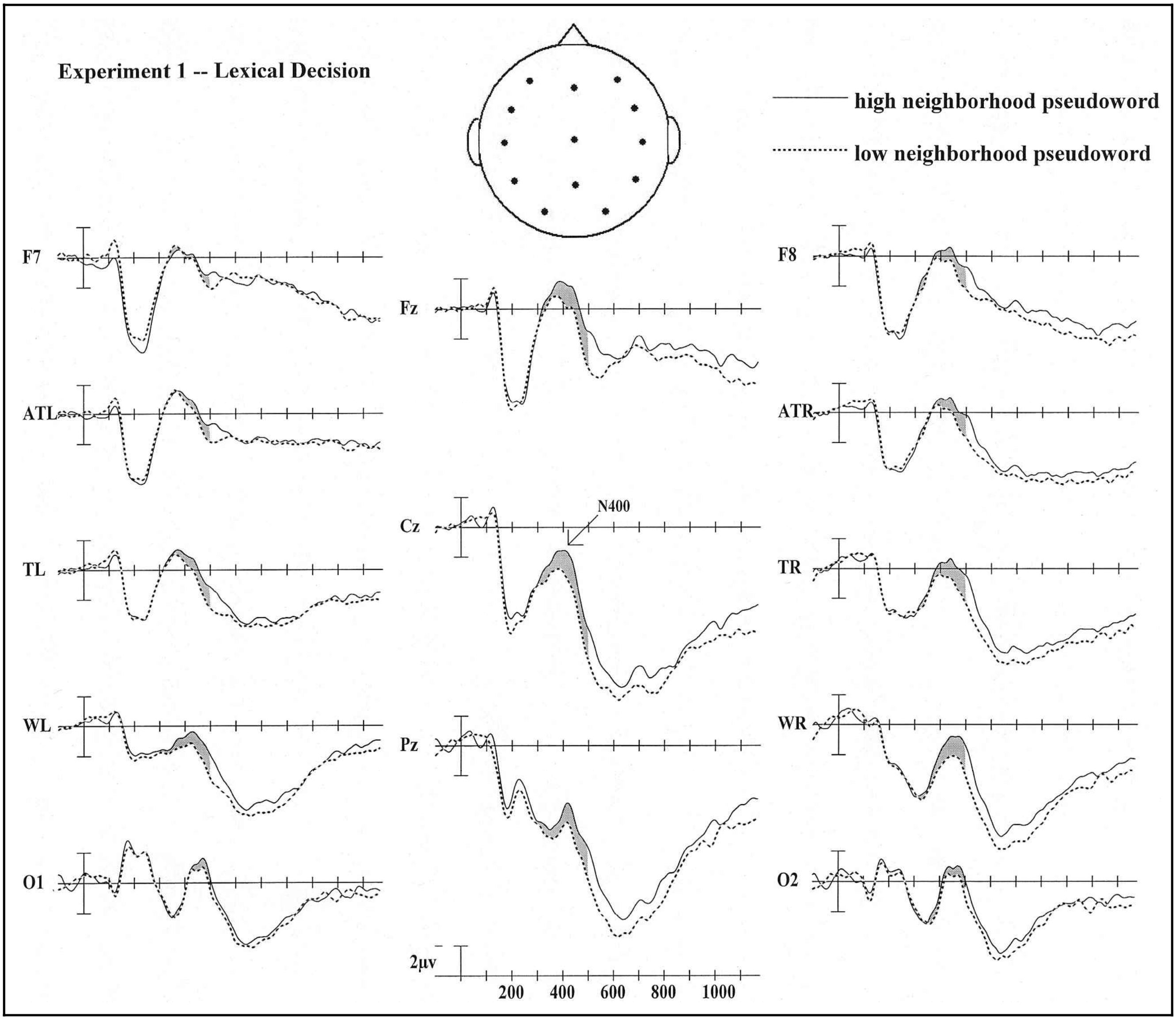

Figure 2. Plotted in this figure are the grand average ERPs from 24 subjects for pseudowords stimuli in Experiment 1. The solid lines are ERPs from large (high-density) neighborhoods and the dashed lines are for ERPs from small (low-density) neighborhoods.

between 100 and 150 msec from stimulus onset (N1). This was followed by a positive deflection occurring at approximately 200 msec (P2). An N400-like broad negativity followed the P2 and peaked at approximately 400 msec. Finally, a broad positivity (P3) occurred starting around 600 msec.

Analyses between 150 and 350 msec from stimulus onset revealed no effects of neighborhood size or any interactions of any factor with neighborhood size for either the midline or the lateral analyses. Several effects, however, were significant between 350 and 550 msec from stimulus onset. There was a main effect of neighborhood size [midline: $F(1,23) = 27.62$, $p < .00001$, $MSE = 6.33$; lateral: $F(1,23) = 6.94$, $p < .015$, $MSE = 16.41$], with stimuli from larger neighborhoods generating greater negativity than stimuli from smaller neighborhoods. Stimulus type also produced a main effect [midline: $F(1,23) = 34.92$, $p < .00001$, $MSE = 30.86$; lateral: $F(1,23) = 25.43$, $p < .00001$, $MSE = 46.41$], with pseudowords producing greater negativity than words. The main effect of stimulus type was also significant in the interval between 550 and 800 msec [midline: $F(1,23) = 7.05$, $p < .05$, $MSE = 16.65$; lateral: $F(1,23) = 5.12$, $p < .0334$, $MSE = 17.06$], but the neighborhood size effect was only significant in the midline analysis [midline: $F(1,23) = 4.37$, $p < .05$, $MSE = 14.46$; lateral: $p > .16$]. There were no significant effects of number of letters in any measurement window.

Finally, and most relevant for the comparison to RT results, there was an interaction between stimulus type and neighborhood size in the N400 window [midline: $F(1,23) = 6.42$, $p < .05$, $MSE = 3.72$; lateral: $F(1,23) = 3.52$, $p < .07$, $MSE = 6.32$]. Simple effects tests revealed that large neighborhoods produced greater negativity than small neighborhoods for words at midline sites [midline: $F(1,23) = 8.47$, $p < .01$, $MSE = 24.67$; lateral:

$p > .20$] and for pseudowords in both midline and lateral analyses [midline: $F(1,23) = 27.62$, $p < .00001$, $MSE = 35.64$; lateral: $F(1,23) = 9.85$, $p < .0046$, $MSE = .60$]. Therefore, the significant interaction between stimulus type and neighborhood size was due to the effect of neighborhood size being larger for pseudowords than for words, not to an opposite pattern for words and pseudowords as with RT.

Discussion

The most important result from Experiment 1 was the interaction between neighborhood size and stimulus type. For RT and error rates, there was a clear and strong cross-over interaction replicating the standard findings (e.g., Andrews, 1989) of a facilitatory effect of neighborhood size for words and an inhibitory effect for pseudowords. In other words, words from large neighborhoods were responded to significantly faster than words from small neighborhoods while the opposite was true for pseudowords. The ERP results were qualitatively different from the behavioral findings. The effect of neighborhood size for both words and pseudowords was in the same direction, with items from large neighborhoods producing greater negativity between 350 and 900 msec than items from smaller neighborhoods. This result is in line with our predictions. We argued that the same core mechanism, operating on global lexical activity, is at the basis of both the facilitatory and the inhibitory effects of orthographic neighborhood size on behavioral responses to word and pseudoword stimuli in the lexical decision task. It was then argued that a measure of processing that directly reflects variations in global lexical activation should show effects of neighborhood size that are in the same direction for word and pseudoword stimuli. This is exactly what the electrophysiological findings of Experiment 1 demonstrate.

At a more general level, this study suggests that neighborhood size as defined by the N-metric (Coltheart et al., 1977) is a legitimate operational definition of lexical similarity because reliable neighborhood effects were found for stimuli with four and five letters. Although some behavioral differences were obtained between four- and five-letter stimuli, this factor did not interact with neighborhood size, nor did it produce any effects on the ERP results.

The ERP results from this experiment are also consistent with the semantic interpretation of the N400 component. According to this account, larger N400s were found for words from larger neighborhoods because these items result in activation of more semantic information. Unlike previous studies, which have demonstrated that it is the semantic information directly associated with the target word itself that produces larger N400s (e.g., as in the concreteness effect, West & Holcomb, 2002; Holcomb et al., 1999), here it appears that the N400 was sensitive to the sum of semantic information from the target word as well as from its lexical neighbors. A similar pattern of sensitivity to neighborhood size for pseudowords helps make two points. First, this finding is consistent with the view that the semantic properties of neighbors contribute to the N400, as in this case lexical neighbors are the only possible source of semantic activity. Second, the finding that pseudowords are sensitive to neighborhood size supports the hypothesis put forward in the Introduction that these items produce N400s, at least in part, because they activate representations for their real word lexical neighbors.

Finally, two aspects of the effect of neighborhood size on the N400 were troubling and specifically motivated the second experiment. First, the overall magnitude of the effect was relatively small, particularly for words. The largest difference observed at any individual electrode site between large and small neighborhoods was slightly more than 1.5 μV. Second, the N400 neighborhood effect for words was only significant in the midline analyses. Furthermore, it could be that the results of Experiment 1 were due to the use of the lexical decision task itself. That is, the stimulus type and neighborhood size effects on the N400 may have resulted from a task-specific word processing strategy. If, on the other hand, the N400 is tapping into task-independent processes involved in the perception of printed word stimuli, then we expect to observe similar effect patterns across tasks.

EXPERIMENT 2: SEMANTIC CATEGORIZATION

A dominant feature in both Figures 1 and 2 was a late positive component starting during the N400 interval and extending throughout much of the end of the recording epoch. One possible reason for this positivity may be that it is associated the overt decision that subject has had to make to every item. Numerous studies have reported similar late positivities when decisions of this type are required (see Donchin & Coles, 1988). Because the late positivity overlaps the N400, it is possible that its presence may have attenuated or obscured the preceding negativity in the region of the N400. Moreover, the effects of neighborhood size also extended late into the epoch of this component. This finding could indicate that neighborhood size effects both the N400 and a subsequent ERP component or that the underlying effect is on some component other than the N400 itself (e.g., one that starts in the N400 time period, but extends beyond it). Therefore, one goal of Experiment 2 was to determine if neighborhood size effects could be obtained in the absence of late positivities and if the effect could be restricted to the traditional N400 time window (300–600 msec).

Another goal of this study was to determine if the neighborhood size effect found in Experiment 1 could

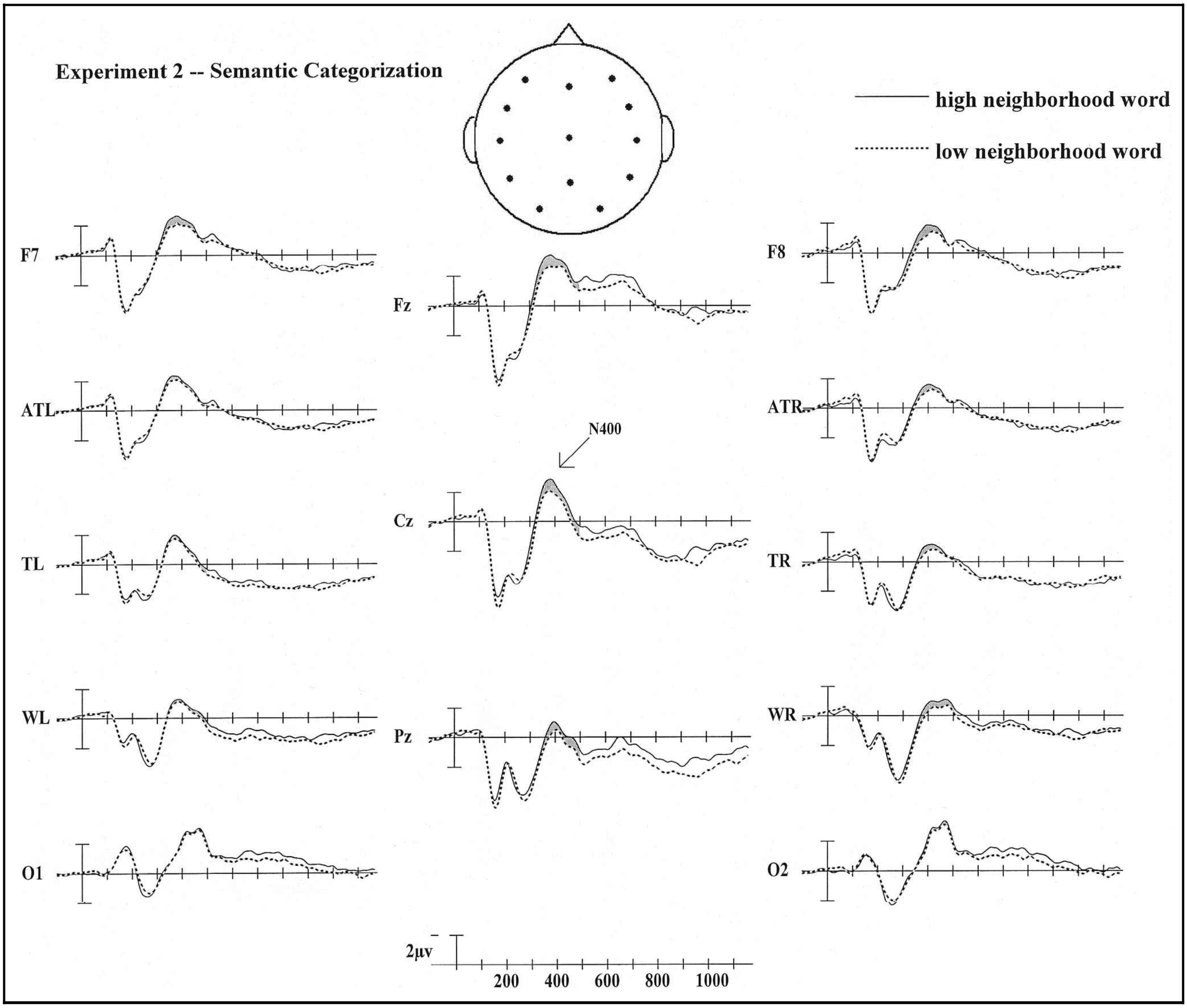

Figure 3. Plotted in this figure are the grand average ERPs from 24 subjects for words stimuli in Experiment 2. The solid lines are ERPs from large (high-density) neighborhoods and the dashed lines are for ERPs from small (low-density) neighborhoods.

also be found in a different type of task: semantic categorization. Prior research with such a task (Sears et al., 1999; Carreiras et al., 1997; Forster & Shen, 1996) has led to a mixed set of results. In all of these studies subjects were requested to decide as rapidly as possible whether a given word was or was not an animal. The critical stimuli were not animal names, and hence elicited negative responses from subjects. The main dependent variable was correct negative RT. Forster and Shen (1996) reported a facilitatory effect of neighborhood size that was not significant in the item analyses. In the Sears et al. (1999) study, this facilitatory effect was significant by subjects and by items, while Carreiras et al. (1997) reported a null effect of this variable (although the precise pattern was complicated by an interaction between neighborhood density and the relative frequency of these neighbors). However, when one reflects on how subjects might actually go about making semantic categorization decisions, then this variability in the data is not so surprising. Carreiras et al. argued that the semantic categorization task involves a task-specific component where subjects monitor activity in a task-relevant semantic dimension. Negative responses can then be generated by a deadline mechanism that would be sensitive to the extent to which a target word causes activation in the relevant semantic dimension. In line with this hypothesis, Carreiras et al. showed that the rated ‘‘animalness’’ of their nonanimal target words correlated significantly with RTs in that experiment.

Given the difficulties in measuring the influence of neighborhood size in overt semantic categorization responses, it seemed critical to test whether an effect could be observed in ERP recordings. The critical prediction for the present study is that contrary to behavioral measures, some ERP measures (e.g., N400) might not be sensitive to task variations, but tap core processes that operate independently of task. Thus in Experiment 2, subjects viewed the same stimuli used in Experiment 1, but only made overt responses to members of a single semantic category (body parts). Words in this category were not part of the experimental cohort of items.

Results

Approximately 11% of the trials were rejected because of eye blinks, horizontal eye movement, or amplifiers blocking. Because overt responses were not made to the stimuli of interest, only ERP data were available for analysis. Figure 3 displays ERPs for words with large neighborhoods and small neighborhoods. Figure 4 shows ERPs for pseudowords with large neighborhoods and small neighborhoods. These ERPs begin with a standard N1–P2 complex but are followed by a sustained negativity in the region of the N400. The most salient feature of these figures, relative to the ERPs from the first study, is the comparative absence of positivity late in the waveforms.

Both the main effects of stimulus type and neighborhood size were significant between 150 and 300 msec from stimulus onset [midline: $F(1,23) = 14.22$, $p < .001$, $MSE = 4.81$; lateral: $F(1,23) = 12.76$, $p < .002$, $MSE = 4.90$; midline: $F(1,23) = 9.69$, $p < .01$, $MSE = 4.21$; lateral: $F(1,23) = 3.96$, $p < .06$, $MSE = 5.87$]. Within this early window, pseudowords produced a greater negativity than words and stimuli from large neighborhoods, which produced a greater negativity than stimuli from small neighborhoods. These main effects were maintained between 350 and 550 msec from stimulus onset [stimulus type, midline: $F(1,23) = 6.34$, $p < .02$, $MSE = 8.38$; lateral: $F(1,23) = 3.19$, $p < .09$, $MSE = 9.50$;

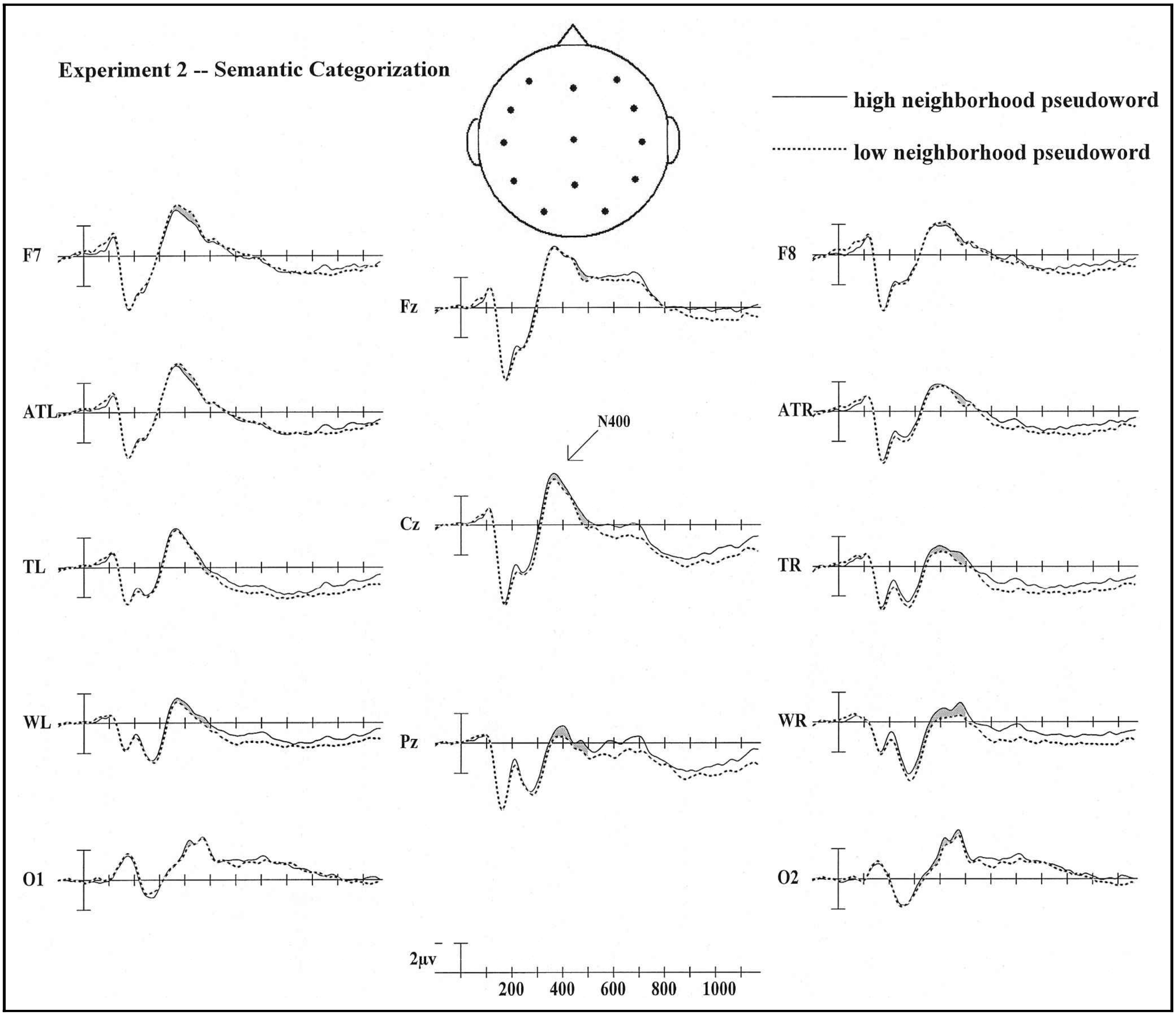

Figure 4. Plotted in this figure are the grand average ERPs from 24 subjects for pseudowords stimuli in Experiment 2. The solid lines are ERPs from large (high-density) neighborhoods and the dashed lines are for ERPs from small (low-density) neighborhoods.

neighborhood size, midline: $F(1,23) = 22.06$, $p < .0001$, $MSE = 4.29$; lateral: $F(1,23) = 16.07$, $p < .0006$, $MSE = .39$]. Finally, the main effect of neighborhood size was maintained between 550 and 850 msec from stimulus onset [midline: $F(1,23) = 10.27$, $p < .01$, $MSE = 8.2$; lateral: $F(1,23) = 10.27$, $p < .004$, $MSE = 8.20$]. Again, stimuli with more neighbors produced a larger negativity. No other effects were significant.

Discussion

Experiment 2 demonstrated that ERPs for both words and pseudowords were more negative for stimuli from larger neighborhoods than for stimuli from small neighborhoods when subjects were required to evaluate each stimulus for membership in a semantic category. This finding extends the pattern of results found in Experiment 1 using a lexical decision task to a second, qualitatively different task.

The results of this experiment, while consistent with those of Experiment 1, are also more straightforward in that there were significant neighborhood effects at both midline and lateral sites and this effect did not interact with lexical status (stimulus type). Moreover, removing the requirement of an overt decision and response from the critical items successfully eliminated the overlapping late positive component that was seen in Experiment 1 (compare Figures 1 and 2 with Figures 3 and 4). However, this did not eliminate the later phase of the neighborhood size effect seen in the waveforms of Experiment 1 between 550 and 850 msec. In Experiment 2, there continued to be a neighborhood size effect in this late epoch. In addition, a similar effect, not seen in Experiment 1, emerged during the earliest epoch in Experiment 2 (150–300 msec).

GENERAL DISCUSSION

The two experiments reported here examined the effects of orthographic similarity across printed word stimuli using ERP measures of stimulus processing. The behavioral literature has repeatedly demonstrated processing differences across stimuli that vary in terms of the size of their orthographic neighborhood. Thus, in the lexical decision task (speeded word/pseudoword classification), word stimuli with a large number of neighbors are responded to faster than words with few orthographic neighbors (see Andrews, 1997, for a review). On the other hand, RTs to pseudoword stimuli in the lexical decision task are longer when these stimuli have many word neighbors. One interpretation of this observed dissociation is that increased global lexical activation during stimulus processing, resulting from the presence of large numbers of orthographic neighbors, leads to diametrically opposite effects through the operation of distinct decision mechanisms (Grainger & Jacobs, 1996; Johnson & Pugh, 1994). For word stimuli, fast positive responses can be generated via a response criterion set on a measure of global lexical activation. Hence, words that generate higher levels of global activation, which is the case when number of orthographic neighbors is increased, will be responded to more quickly. For pseudoword stimuli, a similar increase in global lexical activation will lead to the raising of a negative response criterion, thus slowing responses to stimuli that generate such activation.

Experiment 1 of the present study tested one prediction derived from the above account of the observed dissociation in behavioral data: That similar effects of neighborhood size for words and pseudowords would be observed in ERP components reflecting basic, response-independent processing of stimuli. The lexical decision results of this experiment replicated the strong cross-over interaction between neighborhood size and stimulus type in RTs, while showing neighborhood size effects in the N400 amplitudes that go in the same direction for words and pseudowords. Stimuli with many neighbors generated stronger negativity than stimuli with few neighbors, and this effect was significantly stronger for pseudowords compared to word stimuli. The fact that the effects of neighborhood size were in the same direction for word and pseudoword stimuli suggests that ERPs are tapping into some basic processing of these stimuli that is affected by number of orthographic neighbors.

As well as the classic dissociation in neighborhood effects obtained to word and pseudoword stimuli in the lexical decision task, the behavioral literature also shows very different patterns of neighborhood effects with the same set of word stimuli presented in different experimental tasks. Continuing the logic developed above, it can been argued that different influences of a given variable across different tasks reflects the different ways in which activation generated by a stimulus is translated into a response specific to a given task. Thus, in a semantic categorization task, although word stimuli with many neighbors may be generating higher levels of lexical activation, this increase in global activation need not necessarily lead to faster responding in that task. On the other hand, if the N400 is capturing a response-independent process that is sensitive to neighborhood size then it was predicted that the ERP data should be consistent across tasks. This was tested in Experiment 2 with a semantic categorization task. Because subjects did not have to respond to the critical stimuli in this experiment, late positivities associated with speeded behavioral responses were therefore eliminated.

The results of Experiment 2 provide further confirmation that the ERP measures studied here reflect stimulus processing that is affected by neighborhood size, and that the precise processes that are being influenced by the neighborhood manipulation are being tapped independently of task- and response-specific requirements. In both experiments of the present study, the effects of neighborhood size were robust within the standard window of the N400 (350–550 msec) and continued to be significant in later components. Furthermore, ERP recordings in the semantic categorization experiment showed a significant influence of neighborhood size in earlier components (150–350 msec).

The most straightforward interpretation of the present findings is that stimuli with larger numbers of neighbors lead to increased levels of activation associated with the processing of printed strings of letters. This increase in activation level could be associated with some measure of global lexical activity, either at the level of form (orthographic and phonological) representations and/or at the level of semantic representations. Given current evidence in favor of a semantic interpretation of the N400 (e.g., Brown & Hagoort, 1993; Holcomb, 1993), it seems likely that the effects observed in this specific ERP component reflect differences in overall semantic activation generated by different stimuli. It is interesting to note that some recent eye movement studies have shown an inhibitory influence of orthographic neighborhood on reading words in sentence contexts (Pollatsek, Perea, & Binder, 1999; Perea & Pollatsek, 1998). Readers spent significantly more time inspecting words with many orthographic neighbors than words with few neighbors when reading for meaning. This increase in eye fixation duration could reflect an increased difficulty in integrating a specific word meaning into a sentence-level interpretation, given the coactivation of meanings from orthographic neighbors.

A seemingly curious finding that cropped up in both experiments was the main effect of stimulus type. Pseudowords, irrespective of neighborhood size, produced larger N400s than real words. A number of prior studies have also found larger N400s for pseudowords than real words (e.g., Holcomb, 1993; Holcomb & Neville, 1990; Bentin, McCarthy, & Wood, 1985). One obvious question is why should pseudowords, which were matched with the words for neighborhood size, generate larger N400s? Should they not produce N400s equivalent to or possibility even smaller than those to their matched real words? Within the framework discussed earlier (Grainger & Jacobs, 1996), such a finding might indicate that for real words, the reader is able to efficiently resolve upon a single semantic representation, even when an item is from a large neighborhood. This could result from a process like lateral inhibition, where initial semantic activity from lexical neighbors would be high (producing an N400 proportional to neighborhood size), but as processing continues lexical activity from the actual target item would tend to dominate and therefore suppress the activity from other items in the neighborhood. One possibility is that it is this eventual domination of one item that results in the termination of the N400. Now, consider pseudowords. Initially, these items would also activate the semantic representations of their real word neighborhood (also producing an N400 proportional to neighborhood size). However, because no one lexical item would tend to dominate, the semantic representations of real word neighbors would not be suppressed and the N400 would tend to grow unfettered, eventually exceeding that of real words. In other words, pseudoword N400s are larger than real word N400s because semantic activity from lexical neighbors is not efficiently suppressed.

One problem for this explanation is the question of why N400s to pseudowords are in fact not even larger than those observed here? If no suppression occurs, then N400 should continue to build indefinitely. Grainger and Jacobs (1996) offered one possible solution to this problem. In their multiple read-out model, summed lexical activity is used to set a negative response criterion in the form of a variable deadline mechanism. Although they proposed this mechanism for explaining the greater duration of "no" responses in lexical decision, it can easily be extended to the case of semantic processing and termination of the N400. In this view, if an item is not resolved within a given time limit then semantic processing is terminated.

If the above account of N400 effects to pseudowords is correct, then holding neighborhood size constant, while manipulating another index of word similarity, should also produce changes in pseudoword N400 amplitude. In future research, we intend to further examine pseudoword/word similarity by systematically manipulating the number of letters shared by words and pseudowords. Pseudowords differing from a real word by a single letter as opposed to two or more letters should better activate the semantic representation of its real word target, thus resulting in larger N400s than items that differ by two or more letters.

Another line of future experimentation could be to examine ERPs to the stimuli tested in the present study when presented in sentence context. This could be done while varying other factors know to affect N400 amplitude, such as whether or not the target word is semantically anomalous given the preceding context (e.g., Holcomb et al., 1999; Kutas & Hillyard, 1980). This additional manipulation should help specify the precise nature of the neighborhood size effect observed in the present study.

METHODS

Experiment 1

Subjects

Twenty-four volunteers (15 women and 9 men) from the Tufts University community between the ages of 18 and 20 years (mean = 18.7) participated in the study. All were right-handed native speakers of English with normal or corrected-to-normal vision. Nine subjects had at least one left-handed relative in the immediate family (see Kutas, Van Petten, & Besson, 1988).

Stimuli

The stimuli for this experiment were selected from a lexical database containing 150,000 words listing normative data for 26 psycholinguistic attributes (Wilson, 1988). Three hundred and twenty stimuli were used in a 2 × 2 × 2 factorial within-subjects design. The factors were "stimulus type" (words and pseudowords), "number of letters" (four and five letters), and "neighborhood size" (large and small). One letter in each of 160 words was changed to create 160 pseudowords. Stimuli were counterbalanced into two lists with different pseudorandomizations using the constraints that each condition was equally represented and that a pseudoword and the word it was made from never appeared on the same lists. Lexical neighborhoods were defined using the *N*-metric (Coltheart et al., 1977). Conditions were generated such that within each letter and stimulus type large neighborhoods had significantly more members than small neighborhoods (see Table 2). Following Andrews (1992), orthographic redundancy, or bi-gram frequency, was controlled as much as possible. Bi-gram was calculated using the 150,000-word database. Holding the number of letters constant, bi-gram was defined as the sum of all two consecutive letter counts within each stimulus (there are 3 two-letter combinations in four-letter words and 4 two-letter combinations in five-letter words). For three of the four comparisons, bi-gram frequency was not significantly different between large and small neighborhoods. In the one case where there was a significant difference, the mean difference was less than one standard deviation (see Table 2). For word stimuli, word frequency (Francis & Kucera, 1982) was held constant between large and small neighborhoods (see Table 2).

Procedure

A PC style computer was used for stimulus delivery. All stimuli were displayed as white letters on a black screen. A trial consisted of a single stimulus being presented in the center of a computer monitor in lower case letters for 300 msec. Each stimulus was preceded by a 500-msec warning signal (+) presented in the same location as the stimuli and followed by a blank screen for 2 sec. The intertrial interval was 3 sec and coincided with the presentation of a capital letter "B" in the center of the computer monitor. Subjects were instructed to press one button (using one thumb) labeled "yes" if the stimulus was an English word and to press another button (using the other thumb) labeled "no" if the stimulus was not an English word. The hand used for each response was counterbalanced across subjects. Speed and accuracy were stressed equally. Subjects were asked to refrain from moving (except for the button press) or blinking during the presentation of each stimulus and for the interval that the screen was blank following each stimulus. Short breaks were provided approximately every 40 trials. The experimental sessions lasted approximately 20 min.

EEG Procedure

Subjects were seated in a comfortable chair, and an elastic cap (Electro-Cap International) containing 13 tin electrodes was fitted to the scalp. The scalp sites included seven standard international 10–20 system locations (O1, O2, F7, F8, Pz, Cz, and Fz) and six nonstandard locations, including Wernicke's area and its right hemisphere homologue (30% of the interaural distance lateral to a point 13% of the nasion–inion distance posterior to Cz: WL and WR), left and right temporal (33% of the interaural distance lateral to Cz: TL and TR), and left and right anterior temporal (one-half of the distance between F7 and F8 and between T3 and T4: ATL and ATR). Four additional electrodes were attached over the left and right mastoids (the right was recorded actively and the left served as reference for the rest of the sites), below the left eye (for monitoring vertical eye movements and blinks), and to the right of the right eye

Table 2. Stimulus Characteristics

	Neighborhood Size		*Bi-gram Frequency*		*Word Frequency*	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*
Four-letter words, large *N*	14.05	2.26	70.70	9.67	8.98	7.83
Four-letter words, small *N*	3.71	1.54	69.33	13.80	9.11	8.91
Five-letter words, large *N*	7.26	1.60	193.09	36.19	8.89	7.59
Five-letter words, small *N*	1.20	.83	189.56	45.30	7.85	8.08
Four-letter pseudowords, large *N*	14.76	3.27	72.75	16.55		
Four-letter pseudowords, small *N*	4.08	1.41	63.41	16.37		
Five-letter pseudowords, large *N*	6.79	1.81	186.16	55.08		
Five-letter pseudowords, small *N*	1.80	1.06	180.29	57.41		

(for monitoring horizontal eye movements). All impedances were maintained at less than 5 k. . The 16 active sites were interfaced to a Grass Model 12 amplifier system (bandpass 0.01–100 Hz, 60-Hz notch) and the EEG was digitized (12-bit resolution) continuously on-line (200 Hz) throughout the experiment on a second PC-style computer. Averaging was performed off-line after the experimental run.

Data Analysis

Mean RT and proportion of errors were calculated for all stimuli. Only trials with correct responses and latencies between 200 and 2000 msec were included. Additionally, five stimuli had to be excluded from analyses due to errors in stimulus preparation. Average ERPs were formed for each of the eight conditions from correct response trials that were free of ocular and amplifier saturation artifact (less than 15% per condition). The ERP data were quantified by calculating the mean amplitudes (relative to a 100-msec prestimulus baseline) in three latency windows (150–350, 350–550, and 550–850 msec). These three windows were chosen because they roughly correspond to the latency ranges of the N1–P2 complex, N400, and P3 waves reported in previous language studies (see Kutas & Van Petten, 1988). ERPs from midline and lateral sites were analyzed in separate repeated measures analyses of variance (ANOVA; BMDP2V). In addition to stimulus type, number of letters, and neighborhood size, lateral site ERP analyses included the factors electrodes site (occipital, Wernicke's, temporal, anterior temporal, and frontal) and hemisphere (right and left); midline site analyses included an electrode factor (Fz, Cz, and Pz). Only the results from the midline analysis are reported unless midline and lateral analyses differed or hemispheric differences occurred. The Geisser–Greenhouse correction (Geisser & Greenhouse, 1959) was applied to all repeated measures containing more than one degree of freedom in the numerator.

Experiment 2

Subjects

Twenty-four new subjects (14 women and 10 men) from the Tufts University community between the ages of 18 and 20 years (mean = 18.75) volunteered for the study. All were right-handed native speakers of English with normal or corrected-to-normal vision. Six subjects had at least one left-handed relative in the immediate family.

Stimuli and Procedure

Forty additional four- and five-letter words that are in the semantic category of "body parts" (e.g., knee, heart) were added to each of the two lists used in the first experiment. Subjects were instructed to respond with either the left or right thumb (counterbalanced across subjects) when they saw a member of this semantic category. Accuracy was stressed. Three of 160 words used in Experiment 1 were excluded from analyses in Experiment 2 because they were members of the semantic category. The timing, ordering of materials (with the random addition of the fillers), the location of scalp electrodes, and analysis procedures were identical to the first study.

Reprint requests should be sent to Dr. Philip J. Holcomb, Department of Psychology, Tufts University, 490 Boston Ave., Medford, MA 02155, USA, or via e-mail: pholcomb@tufts.edu.

REFERENCES

Andrews, S. (1989). Frequency and neighborhood effects on lexical access: Activation or search? *Journal of Experimental Psychology. Learning, Memory, and Cognition, 15,* 802–814.

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

Andrews, S. (1997). The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin and Review, 4,* 439–461.

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 and Performance, 10,* 340–357.

Bentin, S., McCarthy, G., & Wood, C. C. (1985). Event related potentials, lexical decision and semantic priming. *Electroencephalography and Clinical Neuropsychology, 60,* 343–355.

Brown, C. M., & Hagoort, P. (1993). The processing nature of the N400: Evidence from masked priming. *Journal of Cognitive Neuroscience, 5,* 34–44.

Carreiras, M., Perea, M., & Grainger, J. (1997). Effects of the orthographic neighborhood in visual word recognition: Cross-task comparisons. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 23,* 857–871.

Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance IV* (pp. 535–555). Hillsdale, NJ: Erlbaum.

Donchin, E., & Coles, M. G. H. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Science, 11,* 355–372.

Forster, K. I. (1987). Form priming with masked primes: The best match hypothesis. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading.* Hillsdale NJ: Erlbaum.

Forster, K. I., & Shen, D. (1996). No enemies in the neighborhood: Absence of inhibitory neighborhood effects in lexical decision and semantic categorization. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 22,* 696–713.

Francis, W. N., & Kucera, H. (1982). *Frequency analysis of English usage.* Boston: Houghton Mifflin.

Geisser, S., & Greenhouse, S. (1959). On methods in analysis of profile data. *Psychometrika, 24,* 95–112.

Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review, 103,* 518–565.

Holcomb, P. J. (1993). Semantic priming an stimulus degradation: Implications for the role of the N400 in language processing. *Psychophysiology, 30,* 47–61.

Holcomb, P. J., Kounios, J., Anderson, J. E., & West, W. C. (1999). Dual-coding, context-availability, and concreteness effects in sentence comprehension: An electrophysiological investigation. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 25,* 721–742.

Holcomb, P. J., & Neville, H. J. (1990). Semantic priming in visual and auditory lexical decision: A between modality comparison. *Language and Cognitive Processes, 5,* 281–312.

Johnson, N. F., & Pugh, K. R. (1994). A cohort model of visual word recognition. *Cognitive Psychology, 26,* 240–346.

Kounios, J., & Holcomb, P. J. (1992). Structure and process in semantic memory: Evidence from event-related brain potentials and reaction time. *Journal of Experimental Psychology. General, 121,* 459–479.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207,* 203–205.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307,* 161–163.

Kutas, M., & Van Petten, C. (1988). Event-related brain potential studies of language. In P. K. Ackles, J. R. Jennings, & M. G. H. Coles (Eds.), *Advances in psychophysiology* (vol. 3, pp. 139–187). Greenwich, CT: JAI Press.

Kutas, M., Van Petten, C., & Besson, M. (1988). Event-related potential asymmetries during the reading of sentences. *Electroencephalography and Clinical Neurophysiology, 62,* 203–208.

Perea M., & Pollatsek A. (1998). The effects of neighborhood frequency in reading and lexical decision. *Journal of Experimental Psychology. Human Perception and Performance, 24,* 767–779.

Pollatsek, A., Perea, M., & Binder, K. (1999). The effects of neighborhood size in reading and lexical decision. *Journal of Experimental Psychology. Human Perception and Performance, 25,* 1142–1158.

Rugg, M. D., & Nagy, M. E. (1987). Lexical contribution to nonword-repetition effects: Evidence from event-related potentials. *Memory and Cognition, 15,* 473–481.

Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review, 89,* 60–94.

Sears, C. R., Hino, Y., & Lupker, S. J. (1995). Neighborhood frequency and neighborhood size effects in visual word recognition. *Journal of Experimental Psychology. Human Perception and Performance, 21,* 876–900.

Sears, C. R., Lupker, S. J., & Hino, Y. (1999). Orthographic neighborhood effects in perceptual identification and semantic categorization tasks: A test of the multiple read-out model. *Perception and Psychophysics, 61,* 1537–1554.

Snodgrass, J. G., & Mintzer, M. (1993). Neighborhood effects in visual word recognition: Facilitory or inhibitory? *Memory and Cognition, 21,* 247–266.

West, W. C., & Holcomb, P. J. (2002). Event-related potentials during discourse-level semantic integration of complex pictures. *Cognitive Brain Research, 13,* 363–375.

Wilson, M. D. (1988). MRC psycholinguistic database: Machine-useable dictionary, version 2.0. *Behavioral Research Methods, Instrumentation, and Computers, 20,* 6–10.