Friends in low-entropy places: Letter position influences orthographic neighbor effects in visual word identification

Sahil Luthra (sahil.luthra@uconn.edu)
Psychological Sciences and Connecticut Institute for the Brain and Cognitive Sciences
Storrs, CT 06269-1020 USA

James S. Magnuson (james.magnuson@uconn.edu)
Psychological Sciences and Connecticut Institute for the Brain and Cognitive Sciences
Storrs, CT 06269-1020 USA

Abstract

In visual word recognition, having more orthographic neighbors (words that differ by a single letter) generally speeds access to a target word. But neighbors can mismatch at any letter position. In light of evidence that information content varies between letter positions, we consider how neighbor effects might vary across letter positions. Results from a word naming task indicate that response latencies are better predicted by the relative number of positional friends and enemies (respectively, neighbors that match the target at a given letter position and those that mismatch) at some letter positions than at others. In particular, benefits from friends are most pronounced at positions associated with low *a priori* uncertainty (positional entropy). We consider how these results relate to previous accounts of position-specific effects and how such effects might emerge in serial and parallel processing systems.

Keywords: visual word recognition; orthographic neighbor; letter position; friend; enemy

Introduction

A hallmark of visual word identification is that presentation of a target word entails not only accessing the target but also activating a number of related words. A word's *orthographic neighbors* (henceforth, *neighbors*) are typically defined as words that differ from the target word by the substitution of a single letter (Coltheart et al., 1977). For instance, the neighborhood for *LAKE* includes *BAKE, LIKE, LACE* and *LANE*. In general, words from larger neighborhoods are recognized more quickly than words with fewer neighbors (Andrews, 1997), though the precise way in which neighbors influence the dynamics of lexical access is fairly complex and controversial.

As an illustration, consider the interactive model of word recognition of McClelland and Rumelhart (1981), which includes excitatory bottom-up connections from letter units to word units, reciprocal top-down connections from word units to letter units, and lateral inhibitory connections at both the letter level and word level. In such an architecture, a neighbor (e.g., *LIKE*) directly inhibits a target (*LAKE*) through lateral connections. At the same time, the neighbor reinforces its constituent letters through top-down connections, and these reinforced letters (*L*, *K* and *E*) in turn "confirm" the lexical prediction, boosting activation of the target through bottom-up connections. Thus, neighbors can inhibit and/or enhance the activation of a target word.

The present work is motivated by the idea that in assessing the influence of neighbors on word recognition, it may be useful to consider not only the size of the neighborhood but also its composition. To this end, every neighbor can be classified as a *friend* or *enemy* of a given letter position, depending on whether it respectively matches or mismatches the target word at that position.[1] For instance, *CAKE* is an enemy of *LAKE* at the first letter position but a friend at positions two, three and four.

A consideration of friends and enemies allows for more nuanced characterizations of many phenomena in letter and word identification. Consider the word superiority effect (Reicher, 1969), which is the finding that letter identification is facilitated if the target letter (e.g., *S*) is presented in a word context (*SHIP*) relative to being presented in isolation (*S*). In interactive architectures, this effect is attributable to the critical letter receiving top-down support from the word layer. By appealing to friends and enemies, we are able to capture the slight differences in the word superiority effect when *S* is presented in a context like *SINK* compared to a context like *SHIP*. The word *SINK* has relatively many enemies at the first position (e.g., *LINK*, *MINK*, *PINK*, *RINK*, *WINK*), whereas the word *SHIP* only has two first-position enemies (*CHIP* and *WHIP*). Because there are relatively more possibilities for the first letter of *_INK* than *_HIP*, the word superiority effect is attenuated in the *SINK* context compared to the *SHIP* context, particularly in visual conditions where the input is slightly degraded (Broadbent & Gregory, 1968; Johnston, 1978; McClelland & Rumelhart, 1981).

While the influences of friends and enemies on letter identification are relatively subtle, previous findings suggest that the characteristics of the neighborhood may have a relatively important influence on the dynamics of word identification. When multiple enemies mismatch a target

---

[1] Here, we use the terms *friend* and *enemy* as they are used by McClelland and Rumelhart (1981). Note that this is distinct from another use of these terms in the literature, in which they refer to sets of words with similar or dissimilar phoneme-grapheme correspondence, respectively (e.g., Kay & Bishop, 1987).

word at a given letter position, these enemies form a "gang" and mutually reinforce each other through their interactions with letter nodes; for instance, all the words that mismatch the target *LAKE* in word-initial position constitute an *_AKE* gang. Because of mutual reinforcement among gang members, words in large gangs are relatively more activated than words in small gangs (McClelland & Rumelhart, 1981). Additional evidence for the influence of enemies on word recognition comes from a study by Pugh, Rexer, Peter and Katz (1994), who found that delaying access to a letter impairs visual word recognition if multiple candidates are possible at that position (i.e., if there is at least one enemy to the target word at that position) but not when only one candidate is possible. More generally, such effects suggest that the relative number of friends and enemies at each position has important consequences for the activation dynamics in word identification. A heretofore unanswered question is whether it matters *which* letter positions have particular ratios of friends to enemies. That is, does the facilitative influence of orthographic neighbors depend on *where* those neighbors mismatch the target word?

The idea that it may be important to consider the relative location of friends and enemies has its roots in a substantial body of literature indicating that some letter positions may be more important for visual word identification than others. For instance, a number of studies suggest that access to the initial letter positions in a word is particularly important; subjects fixate on these letters early on in word naming tasks, and visual word recognition is speeded when a target is preceded by a prime sharing word-initial letters more than if the prime shares word-final letters (Forster, 1979; Inhoff & Tousman, 1990; O'Regan, 1981). These biases for word-initial positions appear to reflect the fact that these positions often have less predictable (i.e., more informative) letters rather than being pre-lexical in nature. Indeed, participants employ different fixation strategies when they know the words they are viewing have higher information content in word-final positions, and the enhanced priming from word-initial letters disappears when differences in letter frequency across letter positions are controlled (Grainger & Jacobs, 1993; O'Regan et al., 1984).

Recent work by Blais et al. (2009) supports the idea that a word's information content is not evenly distributed over its letter positions and thus offers a helpful context for thinking about how neighbor effects might differ across letter positions. In their study, participants completed a speeded naming task with five-letter words in French. On each trial, a movie of semi-transparent "bubbles" was overlaid on the word, such that different letter positions were briefly obscured at different points in time. By sampling across a range of trials, the authors were able to ascertain that early access to positions one, three and four was particularly important for correct naming of five-letter words. The authors also conducted "ideal reader" analyses to determine the letter positions where readers should seek to prioritize information extraction. These analyses used a model that iteratively identified the best letter position to process next based on the information identified from previously processed positions. Note that a critical assumption that follows from this task and analysis is that information is sampled serially, letter-by-letter, but not in a simple left-to-right or right-to-left fashion. Using this approach, Blais et al. computed a *relative importance* metric for each letter position in four-, five-, six- and seven-letter words in both French and English; this metric can be thought of as reflecting the importance of accessing a particular letter position early in processing. Finally, the authors compared recognition of five-letter French words in the bubbles task to the ideal reading strategy their analysis derived for such words, observing that that reader recognition was most impaired when access to "more important" positions was impaired. They thus suggested that readers may process letters in accordance with their relative importance. Such an interpretation is striking in light of prominent theoretical accounts of visual word recognition, which assume that letters are processed either in parallel (e.g., McClelland & Rumelhart, 1981) or serially, from left to right for languages like English (e.g., Coltheart & Rastle, 1994; Whitney, 2008).

In the present investigation, we first ask whether the *relative importance* metric of Blais et al. (2009) might be approximated without assuming any degree of serial processing; in particular, we consider the *a priori* amount of uncertainty (or *positional entropy*) associated with each letter position. We then ask how the distribution of friends and enemies across different letter positions affects performance on a speeded word identification task, using a database of trial-by-trial naming data collected for the English Lexicon Project (ELP; Balota et al., 2007). Finally, we consider how the degree to which friends at different letter positions facilitate word recognition relates to the positional entropy at each letter position.

Methods

A full overview of the ELP database is provided by Balota et al. (2007). The database includes lexical characteristics (word frequency, neighborhood size, etc.) for 40,480 words as well as trial-by-trial data from a speeded naming task conducted across multiple universities. For each word, we calculated the number of friends and enemies at each letter position, as well as the ratio of friends to neighbors at each position, resulting in a position-specific measurement of the *relative number of friends* at position *p* (*RF_p*):

$$RF_p = \frac{F_p}{N},$$

where F= friends and N = all neighbors. Equivalently, since any neighbor that is not a friend at position *p* is an enemy at position *p*:

$$RF_p = \frac{F_p}{F_p + E_p}.$$

472 participants contributed to the ELP dataset. We limited our analyses to 4-, 5-, 6- and 7-letter words that had at least one neighbor; the latter constraint was a consequence of our RF calculation. A total of 10,730 words met these criteria and were included in analyses. Trials were

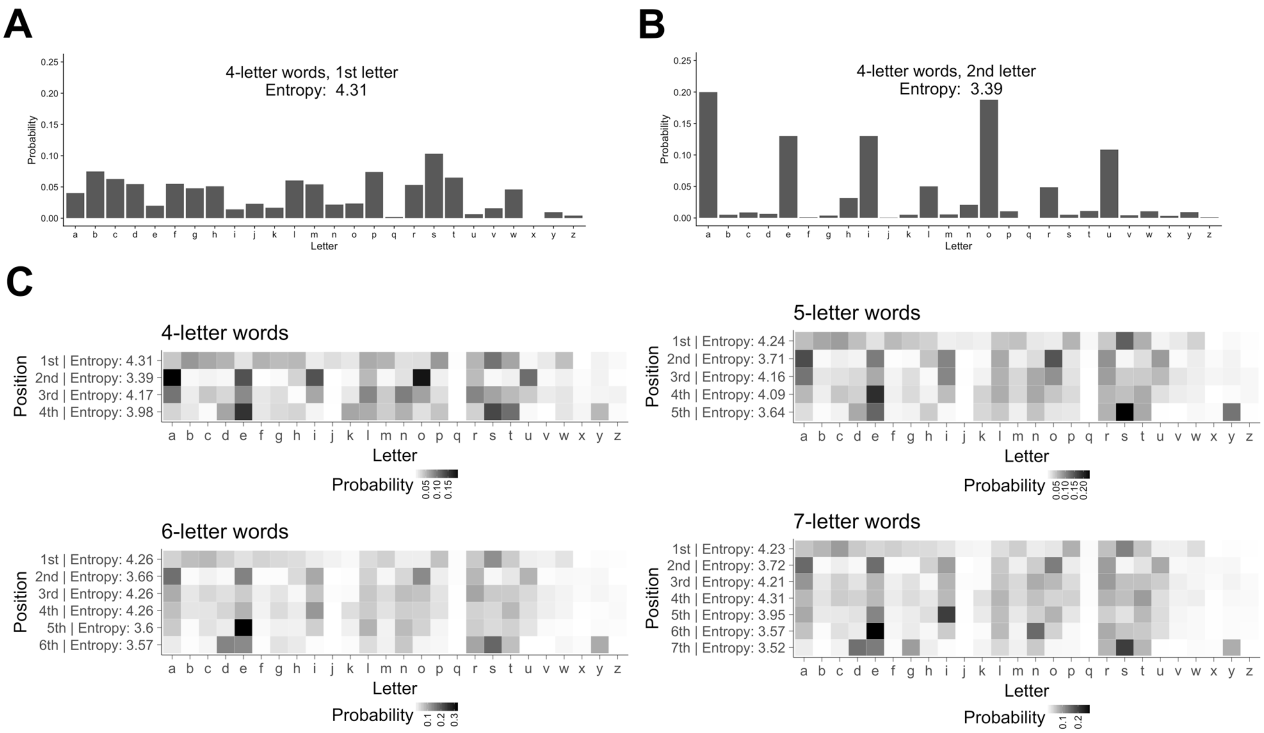

Figure 1: The positional entropy at each letter position for words of varying length. In panels A and B, bar plots represent the probability of each of 26 letters in specific positions for 4-letter words. In panel C, these probabilities are represented more compactly. For example, in the top left diagram in C, the top row corresponds to the bar chart in panel A and the second row corresponds to panel B. In the diagrams in C, darker cells indicate higher probability. Row labels include entropy; a low entropy value corresponds to probability being amassed on a small number of letters, whereas high entropy indicates probability distributed across many letters.

only included if participants self-reported that they had pronounced the word correctly, yielding an average of 27.8 (SD = 2.9) observations per item.

Results

Positional entropy

The *a priori* uncertainty about letter identity differs across letter positions. We quantified uncertainty by computing the Shannon entropy at each letter position based on all the words of that length in the ELP. In particular,

$$Positional\ Entropy = -\sum p(x_i) \cdot log_2 p(x_i),$$

where x_i represents each possible letter at a given position (Figure 1). The smaller the value, the less uncertainty there is about the letter's identity (given only information about word length). If only one letter were possible at a given position (e.g., if every word in the English language began with an *e*), then the entropy at that position would be 0. Similarly, if all letters were equally probable at a given position, the entropy would be ∞.

In computing their *relative importance* values for each letter position, the ideal reader model used by Blais et al. (2009) sampled letters serially, and the relative importance of each letter was conditioned on the identity of other letters in the word. By contrast, our positional entropy metric offers a coarser measurement of relative importance, as it does not assume serial processing and probabilities are computed given only the length of the word. As shown in Figure 2, even this coarse positional entropy metric can

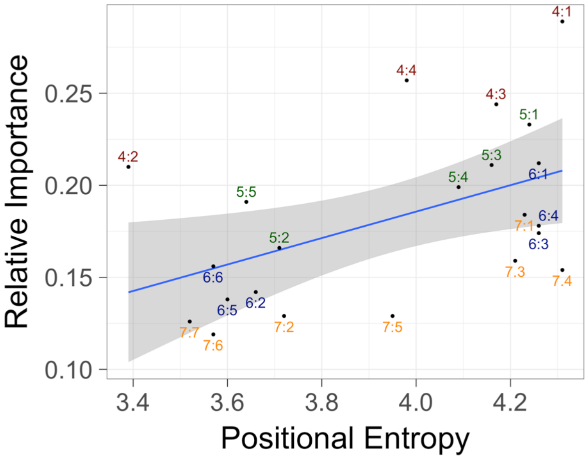

Figure 2: The *a priori* positional entropy at each letter position approximates the (serial) relative importance metric from Blais et al. (2009). Labels (e.g., 4:2) first indicate word length then letter position.

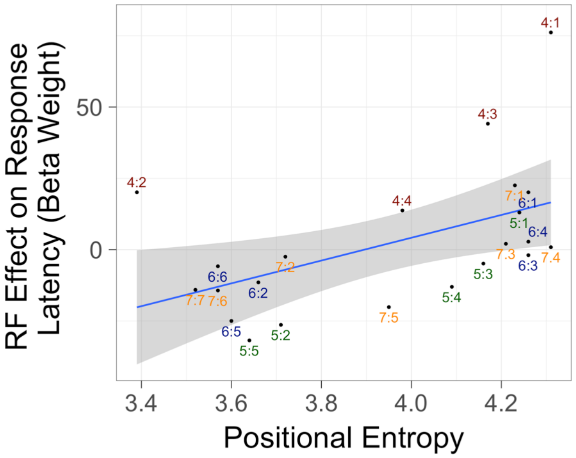

Figure 3: The relative number of friends has a facilitative effect on response latency (negative beta weight) at low-entropy positions.

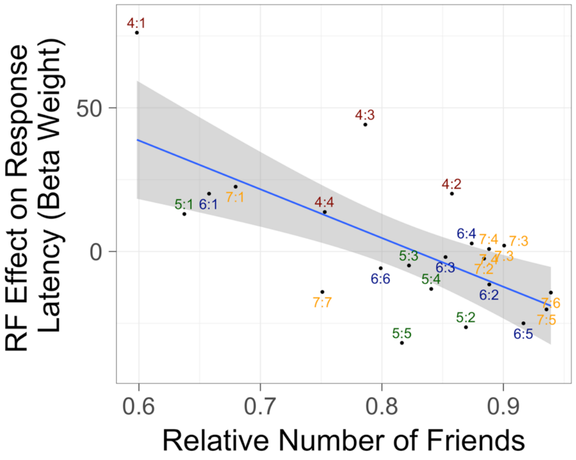

Figure 4: The effect of RF on response latency is more pronounced at positions where there tend to be relatively more friends.

approximate the relative importance values determined by Blais et al. (2009) [$t(20) = 2.49$, $p = 0.022$, $r = 0.487$, $R^2 = 0.237$]. In this and subsequent figures, the label on each point indicates first the length of the word and then the letter position within the word. For instance, the label 6:2 refers to the second position of a 6-letter word.

RF effects by position

We used a series of regression analyses to assess the influence of RF_p on response latency in the speeded naming task[2] conducted by Balota et al. (2007). Because position-specific statistics are contingent on word length, separate models were used for words of different length. Log-normalized word frequencies from the HAL corpus (Lund & Burgess, 1996) were included as a nuisance regressor in each model. Model fit improved as word length increased, with adjusted R^2 values of 0.111, 0.140, 0.141, and 0.154 for the 4-, 5-, 6- and 7-letter word analyses, respectively. Results are summarized in Tables 1-4.

Table 1: Regression analysis for 4-letter words.

	β estimate	SE	*t* value	*p* value
(Intercept)	629.51	104.88	6.00	< 0.001
Log Freq.	-10.59	0.84	-12.64	< 0.001
RF_1	76.17	35.46	2.15	0.03
RF_2	20.12	35.58	0.57	0.57
RF_3	44.16	35.13	1.26	0.21
RF_4	13.74	35.32	0.39	0.70

Table 2: Regression analysis for 5-letter words.

	β estimate	SE	*t* value	*p* value
(Intercept)	855.41	48.28	17.72	< 0.001
Log Freq.	-16.07	0.79	-20.35	< 0.001
RF_1	13.04	12.52	1.04	0.30
RF_2	-26.37	13.22	-2.00	0.05
RF_3	-4.88	13.00	-0.38	0.71
RF_4	-13.04	12.93	-1.01	0.31
RF_5	-31.82	12.99	-2.45	0.01

Table 3: Regression analysis for 6-letter words.

	β estimate	SE	*t* value	*p* value
(Intercept)	828.32	39.42	21.01	< 0.001
Log Freq.	-16.06	0.74	-21.76	< 0.001
RF_1	20.11	8.61	2.34	0.02
RF_2	-11.49	9.42	-1.22	0.22
RF_3	-1.94	9.03	-0.22	0.83
RF_4	2.79	9.25	0.30	0.76
RF_5	-25.00	9.25	-2.70	0.01
RF_6	-5.83	8.80	-0.66	0.51

Table 4: Regression analysis for 7-letter words.

	β estimate	SE	*t* value	*p* value
(Intercept)	843.11	29.98	28.13	< 0.001
Log Freq.	-16.76	0.81	-20.72	< 0.001
RF_1	22.54	6.35	3.55	< 0.001
RF_2	-2.48	7.52	-0.33	0.74
RF_3	2.05	7.77	0.26	0.79
RF_4	0.86	7.50	0.11	0.91
RF_5	-20.16	8.43	-2.39	0.02
RF_6	-14.31	8.20	-1.75	0.08
RF_7	-14.08	6.39	-2.21	0.03

Beta estimates correspond to effects of having relatively more friends at a particular position on reaction time, where negative beta weights indicate a facilitative friend effect. The variability in the beta weights indicates that the relative number of friends has a more pronounced influence on reaction times at some letter positions than at others.

---

[2] We acknowledge that these analyses do not establish causality, but for convenience we adopt the common practice of discussing "effects" of lexical properties on performance measures.

In order to understand why the influence of friends differs across positions, we compared the size of the RF_p effect to the *a priori* amount of uncertainty (i.e., the positional entropy) associated with each letter position. As shown in Figure 3, the influence of the relative number of friends is facilitative (i.e., negative beta weights) in low-entropy positions [$t(20) = 2.65$, $p = 0.015$, $r = 0.510$, $R^2 = 0.260$].

We also compared the size of the RF_p effect on response latency to the mean number of relative friends at each letter position. As depicted in Figure 4, the RF_p benefit for response times is highly pronounced at positions that have relatively more friends [$t(20) = -4.10$, $p = 0.001$, $r = -0.676$, $R^2 = 0.456$]. That is, if there tend to be many friends (on average) at a particular letter position, then words with more friends at that position are recognized more quickly than words with relatively fewer friends at that position.

Discussion

A substantial body of research indicates that a word's substitution neighbors play a pivotal role in the process of word identification, and words with many neighbors are typically recognized more quickly than words with fewer (e.g., Andrews, 1997). McClelland and Rumelhart (1981) noted that every neighbor can be described as an *enemy* (competitor) to the letter position where it mismatches the target and a *friend* (supporter) to all other letter positions. The authors further noted that classifying neighbors as positional friends and enemies might yield a more nuanced characterization of word and letter identification. Here, we investigated how differences in the relative number of friends at each letter position predicted response latencies in a speeded word naming task (Balota et al., 2007), with particular interest in comparing these results to recent work suggesting that some letter positions are more informative than others are (e.g., Blais et al., 2009).

Our results indicate that the facilitative influence of friends is more pronounced at some letter positions than at others. In particular, friend benefits appear to be most pronounced at positions where there is low *a priori* uncertainty about the identity of the letter, as indexed by the Shannon entropy of the letter position (Figure 3). Furthermore, the more friends there tend to be (on average) at a particular letter position, the greater the benefit of having an additional friend at that position (Figure 4).

To clarify the relationship between friend benefits and positional entropy, it may be useful to explicitly state the relationships between entropy, friends and enemies. At low-entropy positions, probability tends to be amassed on relatively few letters, so it is unlikely for a target word to have many enemies at these positions; these positions thus tend to have a relatively large proportion of friends (high RF_p). By contrast, when probability is distributed among relatively many letters (high entropy), enemies are more common, so RF_p values tend to be lower. Friends have a particularly facilitative effect on word recognition when they appear at positions with relatively high *RF* values.

The present results are useful to consider in conjunction with a broader literature suggesting that some letter positions are more informative than others. In one such study, Blais et al. (2009) ascertained the "relative importance" of each letter position using an ideal reader model that assumed that positions are processed in order of their information content; after processing one letter, the model would determine which position would be most informative and process that position next. Blais et al. further suggested that readers may prioritize particular letter positions during visual word recognition, as naming accuracy was impeded to a greater degree when relatively important positions were obscured. In the present work, we demonstrate that the *a priori* degree of uncertainty at each letter position (positional entropy) approximates the Blais et al. measure of the relative importance of different letter positions, suggesting that uncertainty about letter identity may drive differences among positions in visual word recognition. Notably, the entropy of each letter position is calculated independently from other positions (e.g., letter probabilities for the second position are not conditioned on letter identities in the first position).

The present results may inform future investigations about the computational mechanisms underlying visual word recognition. On the basis of their findings, Blais et al. (2009) suggested that visual word recognition entails either a serial processing strategy in which positions are processed in order of their importance or a partially parallel strategy in which only letters in the more important positions are processed simultaneously. It seems possible that the positional effect observed in the current analyses – namely, that there are relatively pronounced facilitation effects of friends in low-entropy places – could emerge in either such architecture. If readers prioritize extracting information from letter positions where there is high prior uncertainty about letter identity, as Blais et al. (2009) suggest, then friends may be particularly helpful at letter positions that are not being prioritized, where they can support a reader's predictions about letter identity. The strong support of these friends may facilitate the extraction of information from high-entropy positions, where there are relatively more enemies and where it is thus particularly helpful to prioritize bottom-up feature extraction. Alternatively, the enhanced relative friend effect observed here might also emerge in a fully parallel processing system for the simple reason that low-entropy positions will tend to have more friends, resulting in greater lexical feedback to these positions and thus position-specific benefits. Computational modeling is needed to dissociate between these possible mechanisms for the emergence of position-specific friend effects.

Finally, while entropy offers a useful way to understand how friend benefits differ across letter positions during word identification, it is worth remarking that our entropy metric considers only orthography. Word naming tasks like the one used by Balota et al. (2007) also require mapping from orthography to phonology, and it is striking that the letters that are most probable in low-entropy positions tend to be ones with highly irregular grapheme-phoneme mappings (e.g., vowels). As such, it may be the case that the relative priority given to some letter positions reflects not the information content of the letter position but rather the likelihood that the position contains a letter that is easy to map to phonology. That is, readers given a word naming task may prioritize positions that tend to have regular phoneme-grapheme correspondence, which may facilitate word naming by providing useful constraints on how to produce the irregular letters. The notion that the relative importance of different letter positions may arise from the orthography-to-phonology mapping process is in line with a literature examining the relationship between orthography and phonology in visual word recognition. For instance, work by Adelman and Brown (2007) suggests that neighborhood effects may be a consequence of print-sound conversion processes rather than of top-down and bottom-up interactions between the word and letter layers (as is argued in interactive accounts of visual word recognition).

In summary, the present investigation brings together work on neighborhood effects in visual word recognition with research suggesting that various letter positions may be differentially important in word identification. Specifically, we focused our analysis on the number of friends (neighbors that match the target at a given letter position) relative to the number of enemies (neighbors that mismatch at the given letter position). Our analyses indicate that the relative number of friends at a given letter position is a useful predictor of word identification latency and that friend benefits for response latencies are most pronounced at low-entropy letter positions, where there is little *a priori* uncertainty about letter identity. We defer to future investigation the question of whether these effects emerge through a system that entails (at least partially) serial processing, whereby processing of high-entropy positions is prioritized, or through a fully parallel processing system.

Acknowledgments

The work pursued here was directly inspired by feedback on a poster we presented at the 2017 Cognitive Science Conference in London. We thank attendees from that conference as well as members of the Neurobiology of Language program at the University of Connecticut for the conversations that motivated this investigation. SL was supported by NSF IGERT DGE-1144399.

References

Adelman, J. S., & Brown, G. D. (2007). Phonographic neighbors, not orthographic neighbors, determine word naming latencies. *Psychonomic Bulletin & Review, 14*(3), 455-459.

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

Balota, D. A., Yap., M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English lexicon project. *Behavior Research Methods*, *39*, 445-459.

Blais, C., Fiset, D., Arguin, M., Jolicoeur, P., Bub, D., & Gosselin, F. (2009). Reading between eye saccades. *PLoS One*, *4*(7), e6448.

Broadbent, D. E., & Gregory, M. (1968). Visual perception of words differing in letter digram frequency. *Journal of Verbal Learning and Verbal Behavior*, *7*(2), 569-571.

Coltheart, M., Davelaar, E., Jonasson, T., & Besner, D., (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and Performance VI*. 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 and Performance*, *20*(6), 1197-1211.

Forster, K. I. (1979). Levels of processing and the structure of the language processor. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett*. Hillsdale, NJ: Erlbaum.

Grainger, J., & Jacobs, A. M. (1993). Masked partial-word priming in visual word recognition: Effects of positional letter frequency. *Journal of Experimental Psychology: Human Perception and Performance*, *19*(5), 951-964.

Inhoff, A. W., & Tousman, S. (1990). Lexical priming from partial-word previews. *J. Experimental Psychology: Learning, Memory, and Cognition*, *16*, 825-836.

Johnston, J. C. (1978). A test of the sophisticated guessing theory of word perception. *Cognitive Psych.*, *10*, 123-154.

Kay, J., & Bishop, D. (1987). Anatomical differences between nose, palm, and foot, or, the body in question: Further dissection of the processes of sub-lexical spelling-sound translation. In M. Coltheart (Ed.), *Attention and performance 12: The psychology of reading*. Hillsdale, New Jersey: Erlbaum.

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers*, *28*(2), 203-208.

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*(5), 375-407.

O'Regan, J. K. (1981). The convenient viewing position hypothesis. In D. F. Fisher, R. A. Monty, & J. W. Senders (Eds.), *Eye movements, cognition, and visual perception*. Hillsdale, New Jersey: Erlbaum.

Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology*, *81*(2), 275-280.

Pugh, K. R., Rexer, K., Peter, M., & Katz. L. (1994). Neighborhood effects in visual word recognition: Effects of letter delay and nonword context difficulty. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(3), 639-648.

Whitney, C. (2008). Supporting the serial in the SERIOL model. *Language and Cognitive Processes*, *23*, 824-865.