A Computational Analysis of the Constraints on Parallel Word Identification

Erik D. Reichle (erik.reichle@mq.edu.au)
Department of Psychology, Macquarie University, NSW 2109, Australia

Elizabeth R. Schotter (eschoter@usf.edu)
Department of Psychology, University of South Florida, FL 33620, USA

Abstract

The debate about how attention is allocated during reading has been framed in as: *Either* attention is allocated in a strictly serial manner, to support the identification of one word at a time, *or* it is allocated as a gradient, to support the concurrent processing of multiple words. The first part of this article reviews reading models to examine the feasibility of both positions. Although word-identification and sentence-processing models assume that words are identified serially to incrementally build larger units of representation, discourse-processing model allow several propositions to be co-active in working memory. The remainder of this article then describes an instance-based model of word identification, Über-Reader, and simulations comparing the identification of single words and word pairs. These simulations indicate that, although word pairs can be identified, accurate identification is restricted to short high-frequency words due to the computational demands of both memory retrieval and limited visual acuity.

Keywords: attention; computational modeling; reading; sentence processing; Über-Reader; word identification

Introduction

The role of attention during reading has been debated because models of eye-movement control in reading (see Table 1) alternatively posit that the attention required to support lexical processing is either limited to one word at a time (e.g., *ASM*: Reilly, 1993; *E-Z Reader*: Reichle, Pollatsek, Fisher, & Rayner, 1998; *EMMA*: Salvucci, 2001), or alternatively, that it can be allocated to support the concurrent processing of several words (e.g., *Glenmore*: Reilly & Radach, 2003; *OB1-Reader*; Snell, van Leipsig, Grainger, & Meeter, 2018; *SWIFT*; Engbert, Nuthmann, Richter, & Kliegl, 2005). Although this debate has motivated many experiments to adjudicate between the two positions, the question has not been resolved because the empirical findings are subject to alternative interpretations, and because the models instantiating the two positions provide equally good accounts of eye-movement control during reading.

As we will argue here, however, this debate has been almost exclusively framed around models of eye-movement control and the eye-movement experiments that they have motivated, with little consideration of what is known about other components of reading. In the remainder of this article, we will redress this limitation by considering the serial-vs.-parallel debate within the larger context of what is known about word identification, sentence processing, and the representation of discourse. More specifically, we consider the role of attention from the perspective of what models of each of the aforementioned processes suggest about the constraints that skilled reading imposes on how words are identified and then used to construct larger units of meaning (e.g., the phrases, sentences, and propositions of a text). In doing this, we remove the debate about attention from its current "either-or" framing by showing what the parallel lexical processing of words might actually entail.

Table 1 lists some of the most influential models of reading. Although this list is not exhaustive, the models are representative of the alternative approaches to understanding how readers (1) use the visual features of words to access their spellings, pronunciations, and meanings from memory; (2) use the meanings of words to construct larger representations of sentences and discourse; and (3) coordinate the movement of their eyes and attention to do the aforementioned processing with some degree of speed and accuracy.

Table 1: Models of reading.

Domain	Model	Primary Reference	Design Influences?
Word Identification	IA	McClelland & Rumelhart (1981)	
	Activation-Verification	Papp et al. (1982)	
	Triangle	Seidenberg & McClelland (1989)	
	Multiple-Levels	Norris (1994)	IA
	Multiple Read-Out	Grainger & Jacobs (1996)	IA
	Multiple-Trace Memory	Ans et al. (1998)	
	CDP	Zorzi et al. (1998)	IA
	DRC	Coltheart et al. (2001)	IA
	SERIOL	Whitney (2001)	
	ACT-R LDT	Van Rijn & Anderson (2003)	
	Bayesian Reader	Norris (2006)	
	CDP+	Perry et al. (2007)	IA, DRC, CDP
	Overlap	Gomez et al. (2008)	
	SCM	Davis (2010)	IA
Sentence Processing	Garden-Path	Frazier & Rayner (1990)	
	SG	St. John & McClelland (1990)	
	SRN	Elman (1990)	
	CC-Reader	Just & Carpenter (1992)	Reader
	Probabilistic Parser	Jurafsky (1996)	
	Attractor-Based	Tabor et al. (1997)	SRN
	Constraint-Based	Spivey & Tanenhaus (1998)	IA
	DLT	Gibson (1998)	
	Cue-Based Parser	Van Dyke & Lewis (2003)	
	Activation-Based	Lewis & Vasishth (2005)	Cue-Based Parser
	Surprisal	Levy (2008)	
Discourse Representation	CI	Kintsch (1988)	IA
	State-Space Search	Fletcher & Bloom (1988)	
	Situation-Space	Golden & Rumelhart (1993)	
	3CI-Dynamic	Goldman & Varma (1995)	CI
	Landscape	van den Broek et al. (1996)	
	Resonance	Myers & O'Brien (1998)	CI
	Langston-Trabasso	Langston & Trabasso (1999)	
	DSS	Frank et al. (2003)	Situation-Space
Reading Architecture	ASM	Reilly (1993)	
	E-Z Reader	Reichle et al. (1998)	
	EMMA	Salvucci (2001)	E-Z Reader
	Glenmore	Reilly & Radach (2003)	IA
	SWIFT	Engert et al. (2005)	
	SERIF	McDonald et al. (2005)	
	OB1-Reader	Snell et al. (2018)	IA, Glenmore

Word-identification models. Although these models instantiate "word identification" in a variety of ways, it is important to note that "…one of the most important aims of

the lexical access process is to make the meaning of the word available to the sentence comprehension system" (Taft, 1991, p. 2). Although the mental processes needed to do this might appear straightforward, the computations required to rapidly and reliably convert the visual features of a word into its spelling, pronunciation, and meaning are complex and prone to error. As demonstrated below, the computational demands of lexical access are severe enough that the models listed in Table 1 adopt specific assumptions to help guarantee its accuracy, with one of the chief assumptions being that individual words and/or their subcomponents are processed serially, one at a time.

For example, the highly influential *interactive-activation model* (*IA*) of McClelland and Rumelhart (1981) assumes that words are identified via a process whereby visual features of letters activate a layer of nodes representing individual letters, which then activate nodes representing words. This activation propagates between these layers of representation across processing cycles, until the word node that best matches the visual input comes to dominate the others through a set of mutually inhibitory connections. This latter assumption is critically important for the present discussion because it is specifically intended to ensure that only one word is identified at any given point in time. This point is illustrated in Figure 1, which shows how input from the word "cat" drives activity in the model in two time steps. First, input from the letter nodes "c," "a," and "t" partially activate the word nodes for "cat," "catch," and "sat." Then, because the letters match "cat," its node comes to dominate the others via an inhibitory 'winner-take-all' competition. This example shows precisely why the inhibitory connections are necessary; without them and the assumption that only one word is identified at a time, words would often be misidentified as their similarly spelled 'neighbors.'

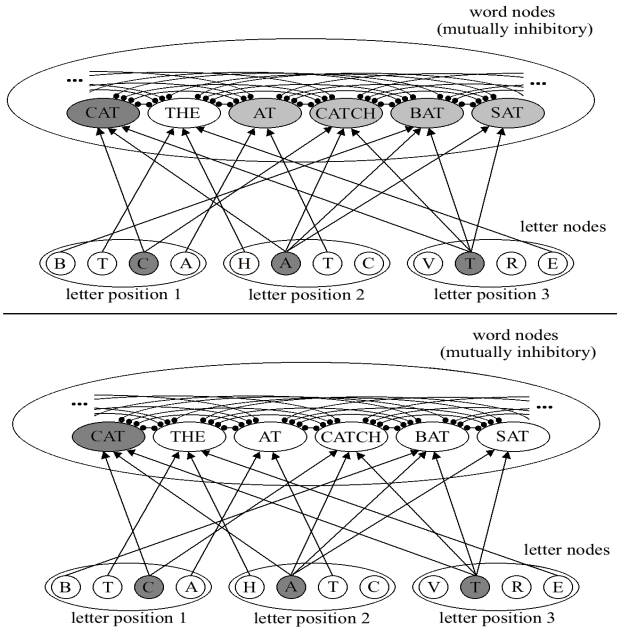

Figure 1: Word identification in the IA model.

As Table 1 shows, this example is important because variants of the IA model are the 'cores' of several more recent word-identifications models. In these models, the nodes represented individual words are interconnected via mutually inhibitory connections for the sole purpose of ensuring that, during any given point in time, only one word can be identified. Thus, like the IA model, its many progeny are specifically designed to enforce the serial identification of words.

Several of the other models also include mechanisms that enforce serial word identification or—in some instances—the serial processing of sub-lexical constituents. For example, the *Activation-Verification model* (Paap, Newsome, McDonald, & Schvaneveldt, 1982) identifies words in two stages: an initial stage in which a cohort of possible "matches" to a word are activated, followed by a stage in which the these candidates are verified one at a time in a frequency-ranked order. Similarly, the *ACT-R LDT* model (Van Rijn & Anderson, 2003) simulates lexical-decision (i.e., binary word vs. non-word decisions to letter strings) by adopting a core assumption of the cognitive architecture from which the model was developed—that only one "chunk" of declarative knowledge (corresponding to a word) can be active during any given 50-ms processing cycle. Finally, four of the remaining models assume a sub-lexical processing route in which letters (*DRC*, *SCM*, *SERIOL*) and/or syllables (*Multiple-Trace Memory model*: Ans, Carbonnel, & Valdois, 1998) are processed serially.

Of course, the assumption that words and/or their sub-lexical constituents are processed serially does not necessarily preclude additional assumptions that might afford the parallel identification of words. For example, the models that assume serial processing of letters might be augmented with the assumption that, at any given point in time, multiple processing 'streams' allow serial letter processing within multiple words, with attention perhaps being instrumental in keeping track of which letters are being processing within each stream. Such possibilities have been suggested, for example, by two of the current models of eye-movement control in reading: Glenmore (Reilly & Radach, 2003) and OB1-Reader (Snell et al., 2018). Both models incorporate variants of the IA model as their word-identification cores, and both models assume that, with the limits of the perceptual span, the letters from spatially adjacent words can co-activate letter nodes (in Glenmore) or bi-gram (i.e., letter pair) nodes (in OB1-Reader) to support the concurrent activation of multiple words. Both models are thus consistent with the hypothesis that *lexical processing* (defined here as the activation of letter/bigram and word nodes) encompasses multiple words. However, like the IA model, both models posit that mutually inhibitory connections among word nodes to ensure that, at any given point in time, one and only one word is *identified*. Thus, although both models allow concurrent lexical processing, these models—like the word-identification models listed in Table 1—are restricted to the serial identification of words.

One way to sidestep the serial-identification restriction is to assume that individual words are represented within multiple (redundant) lexicons, as in Figure 2. Here, the simultaneous processing of two words (e.g., "at" and "the") might occur by mapping the visual forms of each word onto their respective word nodes within independent lexicons. However, although this solution affords the accurate identification of word pairs, it raises more questions than it answers. For example, given that the eyes move along a line of text during reading, one question is: How are the individual words aligned to the different lexicons so that each word only activates nodes within one lexicon? Furthermore, what happens to a given lexicon when the readers' eyes move to another position? And similarly, how is the structure of the lexicons learned? The finding that common words are identified more efficiently suggests that the quality of a word's lexical representation reflects the frequency with which the word has been encounter in text. If this account of the word-frequency effect is correct, then how (if one posits multiple lexicons) would the lexical representation(s) of a word be adjusted with each new encounter with that word?

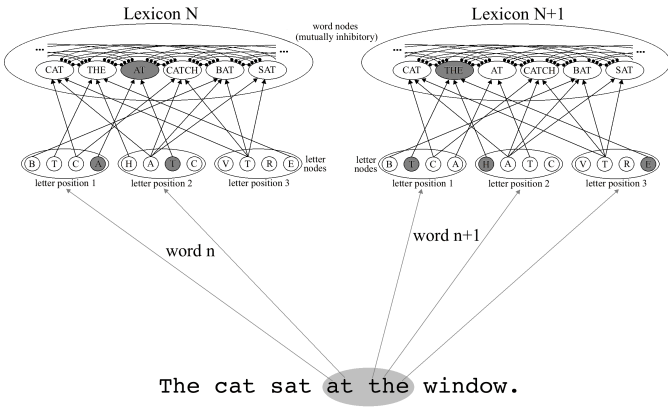

Figure 2: Parallel identification of words represented in multiple (redundant) lexicons.

Turning now to the remaining word-identification models in Table 1, one might ask whether they might accommodate the parallel identification of words. The *Triangle model* (Seidenberg & McClelland, 1989) and its variants (e.g., Plaut, McClelland, Seidenberg, & Patterson, 1996) learn to map patterns of features representing a word's spelling onto patterns of features representing a word's pronunciation. The model learns these mapping across hundreds of learning trials in which the model is given both the correct input and output patterns so that the connection weights between the two can be gradually adjusted. Because connectionist models of this ilk are capable of learning such complex mappings, it is reasonable to predict that, with sufficient training, the models might learn the spelling-to-pronunciation mappings for pairs of words (or even word triples). However, because the models require extensive training to learn how to accurately identify single words, one might also predict that the task of simultaneously identifying two or three words would both dramatically increase the number of training trials and—perhaps more significantly—require one to assume that such training accurately reflects what children experience when they learn how to read. Because language is inherently productive in nature, the extensive training required to train a connectionist model to identify word pairs would seem unreasonable given that words most often appear in novel combinations.

The remaining model, *Bayesian Reader* (Norris, 2006), is unique among the models listed in that it provides a task-level description of word identification, rather than algorithmic- or implementation-level descriptions. The model is thus agnostic about precisely how words are identified, and thus the question of how many words might be concurrently identified. However, the model is in many ways similar to the Multiple-Trace memory model (Ans et al., 1998) in that the task of identifying a word broadly entails the sampling of perceptual input generated by a word for the purpose of mapping that input onto a unique point in representational space. The latter model does this, however, by encoding individual word experiences as discrete memory traces that can then be used to generate a composite pattern representing the word. This approach is also similar to the Triangle model in that the model learns to generate phonological output from orthographic input, but can learn these mappings very rapidly, often with only a single encounter with a given word. The question, then, is whether or not the assumptions of this instance-based model are sufficient to support the concurrent identification of two or more words. The next section of this article answers this question by using a simplified variant of the Multiple-Trace memory model (Reichle, 2020) to examine the conditions under which pairs of commonly co-occurring words (e.g., "… in the …") might be accurately identified.

Über-Reader

The simulations reported below were completed using the word-identification core of the Über-Reader model of reading (Reichle, 2020). This core is based on principles of the Multiple-Trace memory model of word-identification and its precursor, the *MINERVA 2* model of episodic memory (Hintzman, 1984). In the model, experiences with words are encoded as discrete memory traces. In the formalism of the model, these memory traces are vectors of elements representing the presence (= 1) or absence (= 0) of specific orthographic (letter), phonological (phoneme), semantic, and syntactic features. For example, an encounter with the word "cat" would likely result in the encoding of a memory trace with the features corresponding to the letters "c," "a," and "t" in positions 1-3 being set equal to 1 and features corresponding to other letters being set equal to 0. The information in these traces can be accessed via a 'resonance' process in which a probe (in working memory) is used to activate the individual traces to the degree that their contents resemble the contents of the probe. The sum of the activation that is generated by the memory traces in response to a probe is called the 'echo intensity' and reflects the familiar-

ity of the probe and can thus be used to simulate recognition. The features of the activated trace can also be combined to generate a composite pattern of features called the 'echo content,' which can be used to simulate recall. Only those assumptions related to recall are provided below.

In the context of identifying words, a word's orthographic features are used as a probe to recall its other lexical features. Memory traces containing those orthographic features become active to the degree that a trace is *similar* to the probe, as described by Equation 1, where i indexes the memory traces, j indexes the N features, and N_r is the number of non-zero features in either the probe or trace. Because features take on values of 1 or 0, the probe-trace similarity can range from 0 to 1, with the former indicating complete dissimilarity and the latter representing perfect similarity.

$$(1)\ similarity_i = \left(\sum_{j=1}^{N} probe_j \bullet trace_{i,j}\right) \Big/ N_r$$

Trace activation is then determined using Equation 2, where the parameter δ (=17) enhances the signal-to-noise ratio by allowing those traces that are highly similar to the probe to become disproportionately active.

$$(2)\ activation = similarity^{\delta}$$

The signal-to-noise ratio can also be enhanced by delimiting those traces that become active to those that exceed some threshold of similarity to the probe, $\theta_{\text{similarity}}$ (= 0.9; see Dougherty, Gettys, & Ogden, 1999).

The echo content, of value of each recalled feature j, $content_j$, is determined using Equation 3, where M is an index of the number of memory traces and ω_i is a weight assigned to each trace as a function of its frequency of occurrence (Balota et al., 2007), as described by Equation 4. This weighting is used instead of encoding multiple traces per word for computational convenience (see Reichle, 2020).

$$(3)\ content_j = \sum_{i=1}^{M}\left(activation_i \bullet trace_{i,j} \bullet \omega_i\right)$$

$$(4)\ \omega_i = frequency_i \Big/ \max\left[freqeuncy_{i \in M}\right]$$

The echo content generated by Equations 3 and 4 is then normalized using Equation 5, so that the resulting values of the echo content span the range [0, 1].

$$(5)\ content_j = content_j \Big/ \max\left[\left|content_{j \in N}\right|\right]$$

The different lexical features of the echo content can then be scored for accuracy. For example, to score the accuracy of a generated spelling, the most active orthographic feature in each letter position must exceed some threshold, and be the most active feature in the correct letter position. The accuracy of a generated pronunciation is scored similarly, but using phonological features. The accuracy of a word's meaning is calculated as the proportion of correctly recalled semantic features, and a word's part of speech is scored by calculating the correlation, r, between the pattern of syntactic features returned in the normalized echo content and the patterns representing each of the seven possible syntactic categories and then selecting the best match.

Finally, eye-movement models explain visual-acuity constraints on reading (i.e., visual input is more precise in the center of vision and decreases with increasing eccentricity; Schotter, Angele, & Rayner, 2012). For example, serial models include parameters for eccentricity that affect the rate of lexical processing in addition to serial shifts in attention, and parallel models use eccentricity as a parameter that decreases processing efficiency of simultaneously processed words. The 'front end' of Über-Reader likewise provides visual input about letters and their positions using principles of the *Overlap model* (Gomez, Ratcliff, & Perea, 2008). By this account, evidence for a given letter in position x (i.e., the strengths of the features in an orthographic probe) is a function of the letter's true position, μ, as given by:

$$(6)\ f(x,\mu,\sigma) = \left(1/\sigma\sqrt{2\pi}\right)e^{-(x-\mu)^2/2\sigma^2}$$

where the variability is determined by the value of σ, which itself is determined by the absolute difference (in character spaces) between the true position of a letter and the fixation location and two free parameters, β_1 (=0.05) and β_2 (= 0.05):

$$(7)\ \sigma = \beta_1 + \beta_2\left|fixation - letter\right|$$

Simulations

The first set of simulations examined the model's accuracy recalling four types of lexical information (orthographic, phonological, semantic, and syntax) from single words (*Single*) and pairs of words (*Pair*) that were presented at four different fixation locations: (*i*) the center of the stimuli (*Center*); (*ii*) the first letter of the stimuli (*1*); (*iii*) three character spaces left of the stimuli (*-3*); and (*iv*) seven character spaces left of the stimuli (*-7*). Figure4A shows the mean recall accuracy for 16 extremely high-frequency (M = 6,598,697; SD = 5,956,309) 1-4 letter words (e.g., "a," "the", "that") and 16 word pairs derived from these words (e.g., "it is," "in the"). Figure4B shows the mean recall accuracy for 20 low-frequency (M = 10,204; SD = 17,223) 4-10 letter words (e.g., "ants," "hurricane," "parakeet," etc.) and 10 word pairs derived from these words (e.g., "carpenter ants," "sick parakeet"). These words and word pairs were taken from sentences used by Schilling, Rayner, and Chumbley (1998) because future simulations using Über-Reader will examine how the identification of word pairs influences the patterns of eye movements that are generated by the model. The word pairs were represented in the model's lexicon as discrete memory traces using the lowest possible

frequency weighting (i.e., $frequency_i$ = 1; see Equation 4) for the high-frequency word-pair traces, and using the joint probability to estimate the weightings for low-frequency word-pair traces (i.e., $frequency_i$ = 1-5). The model's performance recalling the high-frequency word pairs is thus a conservative test because those pairs would be expected to be represented by more traces if one were to use the joint probability of the two words occurring in written text to estimate their frequencies of occurrence. Finally, all of the simulations were completed using 100 statistical subjects per condition.

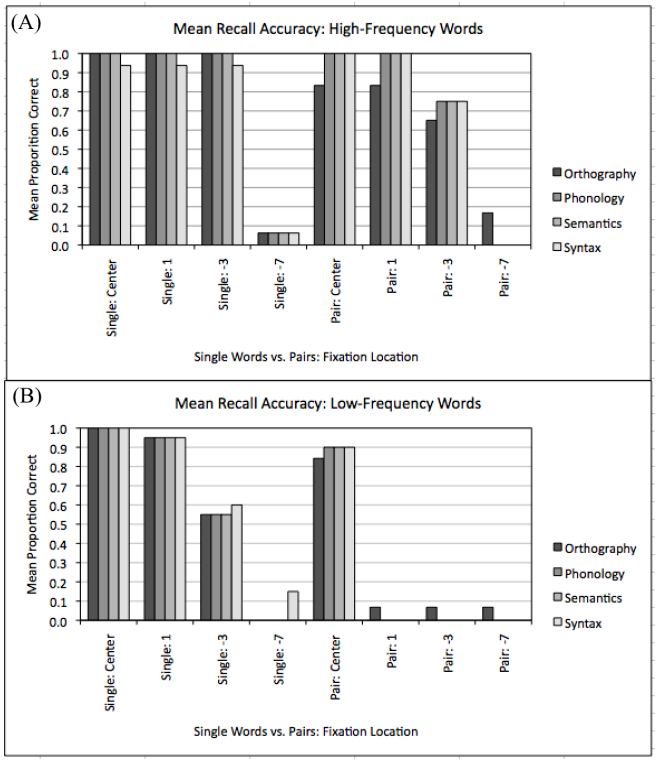

Figure 3: Mean recall accuracy of lexical information corresponding to (A) high- and (B) low-frequency single words vs. word pairs from four fixation locations.

As Figure 3A shows, the model accurately recalled both single words and word pairs if those items were high frequency. However, recall accuracy was slightly reduced for word pairs from fixations three spaces to the left of the first word, suggesting that visual-acuity limitations were slightly more disruptive to the identification of word pairs that single words. Recall of both single words and word pairs was markedly reduced from the distant fixation location, consistent with evidence that visual-acuity limitations delimit word-identification accuracy (Bouma, 1973). Finally, as Figure 3B shows, although recall of low-frequency single words was similar to recall of their high-frequency counterparts (Panel B), the recall of the low-frequency word pairs was significantly reduced at all viewing locations except fixations on the centers of the word pairs.

The second set of simulations (Figure 4) were partial replications of those shown in Figure 3, but using only the low-frequency words and word pairs (shown in Figure 3B) and introducing two manipulations to better understand why the recall of low-frequency word pairs was at a disadvantage in the first set of simulations. The first manipulation entailed reducing the distortion in letter position information by reducing the values of β_1 (= 0.01) and β_2 (= 0.01). As Panel A shows, this markedly improved the recall of the word pairs, effectively allowing them to be recalled with the same level of accuracy as the same words displayed in isolation.

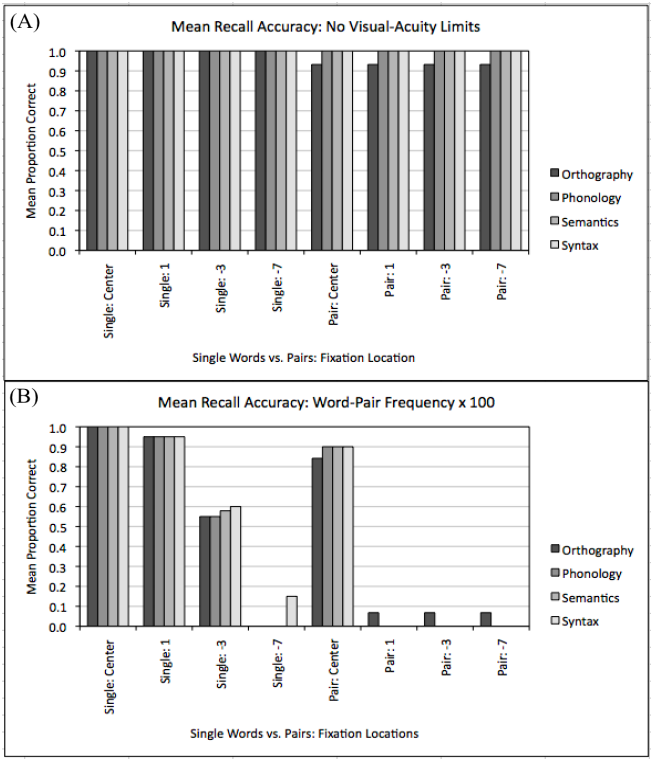

Figure 4: Mean recall accuracy of lexical information of low-frequency single words vs. word pairs from four fixation locations. The two panels show recall: (A) without visual-acuity limits; and (B) using larger frequency values.

The second manipulation increased the frequency weightings of the word-pair traces by setting the value of $frequency_i$ for each equal to the larger frequency value of its two constituent words. (For example, $frequency_i$ for "sick parakeet" was set equal to 22,109 because the frequencies of "sick" and "parakeet" are respectively 22,109 and 178.) As Figure 4B shows, despite the dramatic nature of this second manipulation, word-pair recall accuracy did not improve.

Discussion

Our simulations suggest that the current framing of the serial-vs.-parallel debate about attention allocation in reading is too simplistic. As shown above, an instance-based model of word identification can accurately identify pairs of words if those words are short and high frequency. The condition of being short in length reflects the constraints imposed by limited visual acuity, as suggested by the fact that low-

frequency word pairs could be accurately identified if visual-acuity limitations were removed (Figure 4A). The condition of occurring frequently reflects the necessity of having robust, easily accessible word representations, as suggested by the simulation results in Figure 4B. However, because most English words are longer than three or four letters and language is highly productive, it is unlikely that most word pairs are encountered often enough to be represented in memory. The parallel identification of two or more words would thus likely be limited to sequences like "in the," as well as perhaps idioms or commonly used phrases.

Although the review and simulations reported above have focused on word identification, the discussion can be extended to sentence processing and discourse representation because our understanding of these topics also inform the debate about attention allocation during reading.

Sentence-processing models. Table 1 also lists several influential models of sentence processing—models that are specifically designed to explain how readers construct representations of constituents, phrases, and sentences from individual words (see Reichle, 2020). These models share the assumption that larger representational units are constructed in a staged, incremental manner, using the syntactic category and meaning of each new word that is identified in conjunction with implicit or explicit 'rules' to generate the meaning of a given sentence. The critical part of this shared assumption for the present discussion is the fact that the lexical information is delivered in an *incremental* manner—one that presupposes and depends upon the words being identified one at a time, in their correct order within the sentence. The latter condition is necessary to construct an accurate sentence representation because word order often conveys meaning (e.g., topic focus) even in languages that allow free word order (e.g., German).

Discourse-representation models. Table 1 lists several influential models of discourse representation (see Reichle, 2020). These models are designed to explain how readers construct large units of meaning, deriving from individual sentences. These models share the assumption that the meanings of individual phrases and/or sentences are converted into some type of high-level (e.g., propositional) representation of the meaning of a text, devoid of specific sentential details (e.g., word-order information), and that the meanings of several phrases and/or sentences are concurrently maintained in working memory, subject to its capacity limitations. This latter assumption is critical to the present discussion because it indicates that, at the level of discourse representation, there is significant parallelism, with the meanings of multiple phrases and/or sentences being maintained in working memory over intervals of time so that those meanings can be encoded into long-term memory.

References

Ans, B., Carbonnel, S., & Valdois, S. (1998). A connectionist multiple-trace memory model for polysyllabic word reading. *Psychological Review, 105*, 687-723.

Anderson, J. R. (1996). ACT: A simple theory of complex cognition. *American Psychologist*, *51*, 355-365.

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., . . . Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods, 39*, 445-459.

Bouma, H. (1973). Visual interference in the parafoveal recognition of initial and final letters of words. *Vision research, 13,* 767-782.

Coltheart, M., Rastle, K., Perry, C. Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*, 204-256.

Davis, C. J. (2010). The spatial coding model of visual word identification. *Psychological Review*, *117*, 713-758.

Dougherty, M. R. P., Gettys, C. F., & Ogden, E. E. (1999). MINERVA-DM: A memory processes model for judgments of likelihood. *Psychological Review, 106*, 180-209.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*, 179-211.

Engbert, R., Nuthmann, A., Richter, E., & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review*, *112*, 777-813.

Fletcher, C. R. & Bloom, C. P. (1988). Causal reasoning in the comprehension of simple narrative texts. *Journal of Memory and Language*, *27*, 235-244.

Frank, S. L., Koppen, M., Noordman, L. G. M., & Vonk, W. (2003). Modeling knowledge-based inferences in story comprehension. *Cognitive Science*, *27*, 875-910.

Frazier, L. & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, *14*, 178-210.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, *68*, 1-76.

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

Golden, R. M. & Rumelhart, D. E. (1993). A parallel distributed processing model of story comprehension and recall. *Discourse Processes*, *16*, 203-237.

Goldman, S. R. & Varma, S. (1995). CAPping the construction-integration model of discourse representation. In C. Weaver, S. Mannes, & C. Fletcher (Eds.), *Discourse comprehension: Essays in honor of Walter Kintsch* (pp. 337-358). Hillsdale, NJ, USA: Erlbaum.

Gomez, P., Ratcliff, R., & Perea, M. (2008). The overlap model: A model of letter position coding. *Psychological Review, 115*, 577-600.

Hintzman, D. L. (1984). MINERVA 2: A simulation model of human memory. *Behavior Research Methods, Instruments, & Computers, 16*, 96-101.

Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science*, *20*, 137-194.

Just, M. A. & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*, 122-149.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, *95*, 163-182.

Langston, M. C. & Trabasso, T. (1999). Modeling causal integration and availability of information during comprehension of narrative texts. In H. van Oostendrop & S. R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 29-69). Mahwah, NJ, USA: Erlbaum.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*, 1126-1177.

Lewis, R. L. & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, *29*, 375-419.

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.

McDonald, S. A., Carpenter, R. H. S., & Shillcock, R. C. (2005). An anatomically constrained, stochastic model of eye movement control in reading. *Psychological Review*, *112*, 814-840.

Myers, J. L. & O'Brien, E. O. (1998). Accessing the discourse representation during reading. *Discourse Processes*, *26*, 131-157.

Norris, D. (2006). The Bayesian Reader: Explaining word recognition as an optimal Bayesian decision process. *Psychological Review*, *113*, 327-357.

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.

Perry, C., Ziegler, J. C., & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: The CDP+ model of reading aloud. *Psychological Review*, *114*, 273-315.

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.

Reichle, E. D. (2020). Computational models of reading: A handbook. *Oxford University Press.* Manuscript in press.

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye-movement control in reading. *Psychological Review*, *105*, 125-157.

Reilly, R. (1993). A connectionist framework for modeling eye-movement control in reading. In G d'Ydewalle & J. Van Rensbergen (Eds.), *Perception and Cognition: Advances in eye movement research* (pp. 193-212). Amsterdam, Holland: Elsevier.

Reilly, R. & Radach, R. (2003). Foundations of an interactive activation model of ye movement control in reading. In J. Hyönä, R. Radach & H. Deubel (Eds.), *The mind's eyes: Cognitive and applied aspects of oculomotor research* (pp. 429-456). Oxford, England: Elsevier.

Salvucci, D. D. (2001). An integrated model of eye movements and visual encoding. *Cognitive Systems Research*, *1*, 201-220.

Schilling, H. E. H., Rayner, K., & Chumbley, J. I. (1998). Comparing naming, lexical decision, and eye fixation times: Word frequency effects and individual differences. *Memory and Cognition, 26*, 1270-1281.

Schotter, E.R, Angele, B., & Rayner, K. (2012). Parafoveal processing in reading. *Attention, Perception & Psychophysics, 74,* 5-35.

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

Snell, J., van Leipsig, S., Grainger, J., & Meeter, M. (2018). OB1-Reader: A model of word recognition and eye movements in text reading. *Psychological Review*, *125*, 969-984.

Spivey, M. J. & Tanenhaus, M. K. (1998). Syntactic ambiguity resolution in discourse: Modeling the effects of referential context and lexical frequency. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *24*, 1521-1543.

St. John, M. F. & McClelland, J. L. (1990). Learning and applying contextual constraints in sentence comprehension. *Artificial Intelligence*, *46*, 217-257.

Tabor, W., Juliano, C., & Tanenhaus, M. K. (1997). Parsing in a dynamical system: An attractor-based account of the interaction of lexical and structural constraints in sentence processing. *Language and Cognitive Processes*, *12*, 211-271.

Taft, M. (1991). *Reading and the mental lexicon*. East Sussex, UK: Erbaum.

van den Broek, P., Risden, K., Fletcher, C. R., & Thurlow, R. (1996). A "landscape" view of reading: Fluctuating patterns of activation and the construction of a memory representation. In B. K. Britton & A. C. Graesser (Eds.), *Models of understanding text* (pp. 165-187). Mahwah, NJ, USA: Erlbaum.

Van Dyke, J. A. & Lewis, R. L. (2003). Distinguishing effects of structure and decay on attachment and repair: A cue-based parsing account of recovery from misanalyzed ambiguities. *Journal of Memory and Language*, *49*, 285-316.

Van Rijn, H. & Anderson, J. R. (2003). *Modeling lexical decision as ordinary retrieval*. Paper presented at the Fifth International Conference on Cognitive Modeling, Bamberg, Germany.

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

Zorzi, M., Houghton, G., & Butterworth, B. (1998). Two routes or one in reading aloud? A connectionist dual-process model. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 1131-1161.