Journal of Experimental Psychology: General
2011, Vol. 140, No. 3, 407–433

© 2011 American Psychological Association
0096-3445/11/$12.00 DOI: 10.1037/a0023517

Word Recognition and Syntactic Attachment in Reading: Evidence for a Staged Architecture

Adrian Staub
University of Massachusetts Amherst

In 3 experiments, the author examined how readers' eye movements are influenced by joint manipulations of a word's frequency and the syntactic fit of the word in its context. In the critical conditions of the first 2 experiments, a high- or low-frequency verb was used to disambiguate a garden-path sentence, while in the last experiment, a high- or low-frequency verb constituted a phrase structure violation. The frequency manipulation always influenced the early eye movement measures of first-fixation duration and gaze duration. The context manipulation had a delayed effect in Experiment 1, influencing only the probability of a regressive eye movement from later in the sentence. However, the context manipulation influenced the same early eye movement measures as the frequency effect in Experiments 2 and 3, though there was no statistical interaction between the effects of these variables. The context manipulation also influenced the probability of a regressive eye movement from the verb, though the frequency manipulation did not. These results are shown to confirm predictions emerging from the serial, staged architecture for lexical and integrative processing of the E–Z Reader 10 model of eye movement control in reading (Reichle, Warren, & McConnell, 2009). It is argued, more generally, that the results provide an important constraint on how the relationship between visual word recognition and syntactic attachment is treated in processing models.

Keywords: sentence processing, visual word recognition, eye movements in reading

In the course of normal language comprehension, words are encountered in sentence and discourse context; a listener or reader must both recognize each input word and integrate the word into the syntactic and semantic context in which it appears. Thus, a basic question regarding the architecture of the language comprehension system is about how these two processes are organized temporally and functionally.

The variable that exerts perhaps the most reliable and robust effect on the difficulty of the former of these processes, in both the visual and auditory modalities, is word frequency. Low-frequency words elicit slow response times (RTs), compared with high-frequency words, in isolated visual word recognition tasks such as lexical decision and naming (e.g., Andrews & Heathcote, 2001; Balota & Chumbley, 1984, 1985); they elicit longer eye fixations in normal reading (Inhoff & Rayner, 1986; Rayner & Duffy, 1986; Staub, White, Drieghe, Hollway, & Rayner, 2010); they elicit a larger N400 component in the event-related potential (ERP) paradigm (Allen, Badecker, & Osterhout, 2003; Van Petten & Kutas, 1990); and in the "visual world" paradigm, eye movements to mentioned objects are slower when the object has a low-frequency name (e.g., Dahan, Magnuson, & Tanenhaus, 2001). Thus, accounting for the word frequency effect is a benchmark for both visual (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Harm & Seidenberg, 2004; Norris, 2006) and auditory (Marslen-Wilson, 1987; McClelland & Elman, 1986; Norris & McQueen, 2008) word recognition models; see Norris (2006) and Norris and McQueen (2008) for models in which the frequency effect plays an especially prominent role.

One simple model of how the processes of word recognition and integration of a word into its context are related would hold that the former process precedes the latter (see, e.g., Friederici, 2002; Lewis & Vasishth, 2005, for examples). There are at least two considerations in favor of a staged model of this sort. First, evidence suggests that word recognition is "autonomous" (Forster, 1979; Tanenhaus & Lucas, 1987) in the sense that the initial output of the word recognition system does not depend on syntactic or semantic context (Rayner & Frazier, 1989; Seidenberg, Tanenhaus, Leiman, & Bienkowski, 1982; Swinney, 1979; Tanenhaus, Leiman, & Seidenberg, 1979; though cf. Dahan & Tanenhaus, 2004; Folk & Morris, 2003). Second, it is reasonable to assume that if the process of integrating a word into its context does not await a fair amount of lexical information, such as the major syntactic category (e.g., noun or verb) and principal semantic

This article was published Online First May 23, 2011.

Thanks to Chuck Clifton, Roger Levy, Erik Reichle, and members of the UMass Eyetracking Lab for helpful discussion. Thanks also to Mark Toczylowski, Holly Severance, Caressa DuBois, Patty Samuel, Sally Clouse, and Amorn Jantara for data collection assistance. Thanks also to audiences at the following venues, where portions of this work were presented: the 2009 City University of New York (CUNY) Conference on Human Sentence Processing, Davis, CA; the 2009 European Conference on Eye Movements, Southampton, England; the Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany; the University of Maryland Linguistics Department; the University at Buffalo Cognitive Science Colloquium; and the First International Psycholinguistics Congress of Associação Nacional de Pós-Graduação e Pesquisa em Letras e Linguística (ANPOLL) [National Association of Graduate Studies and Research in Literature and Linguistics], Rio de Janeiro, Brazil.

Correspondence concerning this article should be addressed to Adrian Staub, 430 Tobin Hall, Department of Psychology, University of Massachusetts, Amherst, MA 01003. E-mail: astaub@psych.umass.edu

features (e.g., animate or inanimate) of the input word, this process will be hopelessly error prone.

Experimental methods that deliver temporally fine-grained data would provide one kind of evidence in favor of such a model if the effects of putatively lexical variables like word frequency could be shown to reliably precede the effects of the context in which a word is embedded. However, this is not the case in any existing paradigm. For example, syntactic manipulations of the context of a word have, in many experiments, been shown to affect the duration of readers' first eye fixation on the word (see Clifton, Staub, & Rayner, 2007, for a review), as does the frequency variable. The plausibility of a word in context can also affect first-fixation duration (Staub, Rayner, Pollatsek, Hyönä, & Majewski, 2007; Warren & McConnell, 2007; Warren, McConnell, & Rayner, 2008). In the ERP paradigm, syntactic manipulations often affect the amplitude of the left anterior negativity (LAN) component (see Kutas & Federmeier, 2007, for a review), which peaks at approximately the same latency as the frequency-sensitive N400 component; plausibility in context affects the amplitude of the N400 component itself (Kutas & Hillyard, 1980). In some studies, syntactic manipulations have affected the even earlier ELAN component (e.g., Hahne & Friederici, 2002; Neville, Nicol, Barss, Forster, & Garrett, 1991).

However, the fact that temporal sequencing of frequency effects and effects of a word's "fit" in context has not been reliably witnessed is not, by itself, evidence against the staged model described earlier. It is possible that the temporal grain of the methods that have been employed to address this issue is not sufficiently fine to reveal "hidden" stages operating within, for example, the duration of a reader's first eye fixation on a word.

The E-Z Reader 10 Model

In fact, the latest iteration of the E-Z Reader model of eye movement control in reading (Pollatsek, Reichle, & Rayner, 2006; Reichle, Rayner, & Pollatsek, 2003), E-Z Reader 10 (Reichle, Warren, & McConnell, 2009), assumes exactly this: It implements a simple staged architecture like the one described previously, in which lexical processing precedes integration of a word with its context, while at the same time accounting for the fact that both frequency and contextual fit are able to influence the duration of the first eye fixation on a word. As the predictions of this model provide a framework for the experiments to be presented here, it is necessary to describe certain aspects of the E-Z Reader architecture in some detail and how this basic architecture has been modified in E-Z Reader 10.

Previous versions of E-Z Reader have assumed that it is solely the progress of lexical processing that determines when the eyes move forward off of a word; historically, the model has restricted its focus to forward eye movements. The model assumes two separate stages of lexical processing, L_1 and L_2. In early versions of the model, these were interpreted as specific stages of lexical access, while in more recent versions, they have been defined solely in terms of their eye movement consequences. The durations of both lexical stages are affected principally by two variables, which operate additively: word frequency, and the cloze probability (i.e., predictability) of the word. The model assumes that, when the eyes begin fixating word n, it is word n that is being lexically processed. When the L_1 stage of processing word n is complete, the eye movement control system begins planning a saccade to the next word, $n + 1$. At the same time, the L_2 stage of processing word n begins, and when this stage is complete, attention is shifted to word $n + 1$. Because saccade planning and execution will generally take longer than the L_2 stage, there is usually some period during which attention has shifted to word $n + 1$ and lexical processing of word $n + 1$ is taking place, but the eyes have not yet moved on from word n. Eventually, the eyes do move, and the same operations are repeated. (How the model accounts for word skipping, as well as many other details, will not be discussed here.)

E-Z Reader 10 (Reichle et al., 2009) is motivated by the recognition that higher level variables such as plausibility and syntactic processing difficulty do have effects in the eye movement record, both on reading times and on the rate of regressive saccades. E-Z Reader 10 accounts for such higher level effects by adding a processing stage (integration stage, I) for each word that integrates the word into its syntactic and semantic context. Reichle et al. (2009) remarked that the I stage "is assumed to reflect all of the postlexical processing necessary to integrate word n into the higher level representation that readers construct online—for example, linking word n into a syntactic structure, generating a context-appropriate semantic representation, and incorporating its meaning into a discourse model" (pp. 5–6). The I stage begins for word n when the L_2 stage for word n is complete and runs concurrently with the process of shifting attention to word $n + 1$ and then with lexical processing of word $n + 1$. Recall that the processing of word $n + 1$ begins parafoveally (i.e., before the eyes have actually moved to word $n + 1$) but generally completes after the eyes have moved.

E-Z Reader 10 assumes that the I stage usually runs in the background, having no measurable effects on eye movements. However, there are two ways in which the operations of the I stage can have overt consequences in the eye movement record. The first is when integration of a word lags sufficiently far behind lexical processing of the next word. When the I stage for word n has not completed by the time the L_1 and L_2 stages for word $n + 1$ are complete, both attention and the eyes are directed back toward the source of difficulty: word n. Note that the model does not make a serious attempt to account for the actual targeting of regressive eye movements; Reichle et al. (2009) remarked, specifically, that "the model is clearly not sufficient to provide any detailed account of long-distance regressions" (p. 17). The probability of this event is determined primarily by a model parameter reflecting the duration of the I stage, $t(\mathrm{I})$. As $t(\mathrm{I})$ is increased, it becomes increasingly likely that L_2 for word $n + 1$ will finish before the I stage for word n.

The second circumstance in which the integration process has overt consequences is in the case of outright *integration failure*, in which the I stage terminates in a failed, rather than successful, integration. It will often be the case that lexical processing of word $n + 1$ is still in progress when integration of word n fails. Sometimes this will be while the eyes are still fixating word n and the reader is processing word $n + 1$ parafoveally, in which case the integration failure may cancel a forward saccade. The relevant model parameter is p_F, the probability of failure. If the eyes have already moved to word $n + 1$ before an integration failure, or if the failure occurs while the eyes are still fixating word n but too late for a forward saccade to be canceled, which are likely states of

affairs when t(I) is also high, failure again directs the eyes back toward the source of difficulty.

Reichle et al. (2009) demonstrated in a series of simulations that both first-fixation duration and gaze duration (the sum of all fixations before leaving a word) for word n and the rate of regressive saccades from word n and from word $n + 1$ can be affected by increases in the time required for integration, or t(I), of word n and the probability of rapid integration failure (p_F). When the value of the t(I) parameter is high and the value of p_F remains low, effects of integration difficulty will generally not appear until the eyes have moved on to word $n + 1$. However, when t(I) is fairly low, first-fixation duration and gaze duration (the sum of all eye fixations on the first inspection of the word) on word n increase monotonically with values of the p_F parameter: "As p_F increases, so does the probability that the resulting integration failure will cancel whatever labile saccade would have otherwise moved the eyes forward, resulting in a pause (increasing both the first-fixation and gaze durations) and/or a refixation (increasing the gaze duration)" (p. 10). It is important that the model be able to account for each of these patterns. As noted earlier, there are many experiments in which the difficulty of syntactic or semantic integration of a word is manifested in the earliest eye movement measures on the word itself, but there are also numerous cases in which effects of the difficulty of integrating a word appear only downstream of the word. See Clifton et al. (2007) for a review of this variability.

For present purposes, what is most relevant about E-Z Reader 10 is the assumption that integration processes for word n are indeed *postlexical*, beginning after lexical processing (L_1 and L_2) for word n is complete. The goal of the experiments presented here was to test a set of specific predictions arising from this serial, staged architecture, relating to how the effects of word frequency and syntactic attachment difficulty are each manifested in the eye movement record. The predictions are as follows:

First, effects of word frequency should appear in reading time measures on a word whose frequency is manipulated, regardless of syntactic context. First-fixation duration and gaze duration on a word are determined largely by the durations of the L_1 and L_2 stages, which are a function of frequency, and any effect of the integration parameters is not manifested until after the L_2 stage is complete. Second, while the difficulty of integrating a word into its context can affect both reading times on the critical word and the probability of making a regressive saccade from this word, word frequency should affect only reading times, not the rate of regressions. Summarizing these first two predictions, E-Z Reader 10 predicts that regardless of the level of integration difficulty, word frequency affects reading times but does not affect the probability of a regression from the word in question.

The third prediction is that it is indeed possible for the effects of word frequency and integration difficulty to show strict sequencing in the eye movement record. Though the model provides a mechanism by which the integration difficulty of word n can influence the movement of the eyes before they leave word n, it also allows for a pattern in which the frequency of word n has an effect on the earliest eye movement measures on word n itself, while the difficulty of integrating word n shows up only downstream of this word.

The fourth prediction is that when the difficulty of integrating word n does show up in early measures on word n itself, this should not modulate the size of the frequency effect. As just described, E-Z Reader 10 implements a mechanism in which factors such as syntactic attachment difficulty or implausibility of word n may influence the movement of the eyes before the reader has moved on from this word but in which these factors actually exert their influence during or after lexical processing of word $n + 1$, that is, after lexical processing of word n is complete. E-Z Reader 10 accounts for the fact that integration difficulty can influence the duration of the first eye fixation on a word, while preserving serial ordering between lexical processing and contextual integration.

E-Z Reader 10 Simulations

These predictions from the E-Z Reader 10 model are demonstrated through a set of simulations that were conducted with the Java version of E-Z Reader 10 made available by Erik Reichle at http://www.pitt.edu/~reichle/ezreader.html. Consider the Sentences 1a and 1b:

1a. The professor saw that the students walked across the quad.

b. The professor saw that the students ambled across the quad.

These sentences (which were used in Experiment 1 of the present study) differ in the frequency of the verb in the embedded clause; the Francis and Kučera (1982) frequencies (which are used for E-Z Reader simulation) of *walked* and *ambled* are 159 and 1, respectively. The first simulation simply demonstrates how this frequency difference is predicted by E-Z Reader to affect eye movements, assuming no particular integration difficulty of these words, that is, with the integration parameters of I(n) and $p_F(n)$ set at the their default values of 25 ms and .01, respectively. The words were assumed to have 0 cloze probability. Each sentence was read by 10,000 simulated subjects. The leftmost bars in the panels in Figure 1 illustrate that both first-fixation duration and gaze duration were longer on *ambled* than on *walked*, while there was no effect of the frequency difference on the probability of a regression from the word. The leftmost bars in Figure 2 illustrate that frequency actually had very little effect on first-fixation duration and gaze duration on the next word (*across*) and had no effect on the probability of regressing from that word.

Consider now the processing effects of removing the word *that*, which turns the Sentences 1a and 1b into garden-path sentences. Much previous research has shown that in the absence of the optional complementizer *that*, a reader is likely to initially interpret the noun phrase preceding the critical verb (*the students*) as the direct object of the preceding matrix verb (e.g., Ferreira & Henderson, 1990; Trueswell, Tanenhaus, & Kello, 1993). Thus, *walked* or *ambled* is likely to be initially difficult to integrate into the syntactic analysis that the reader has maintained to that point; several experiments have found evidence of subsequent difficulty in the eye movement record (see Clifton et al., 2007, for a review).

Within the E-Z Reader 10 framework, this integration difficulty could be modeled by increasing one or both of the I(n) or $p_F(n)$ parameters—that is, by increasing the duration of the integration stage, by increasing the probability that integration will fail out-

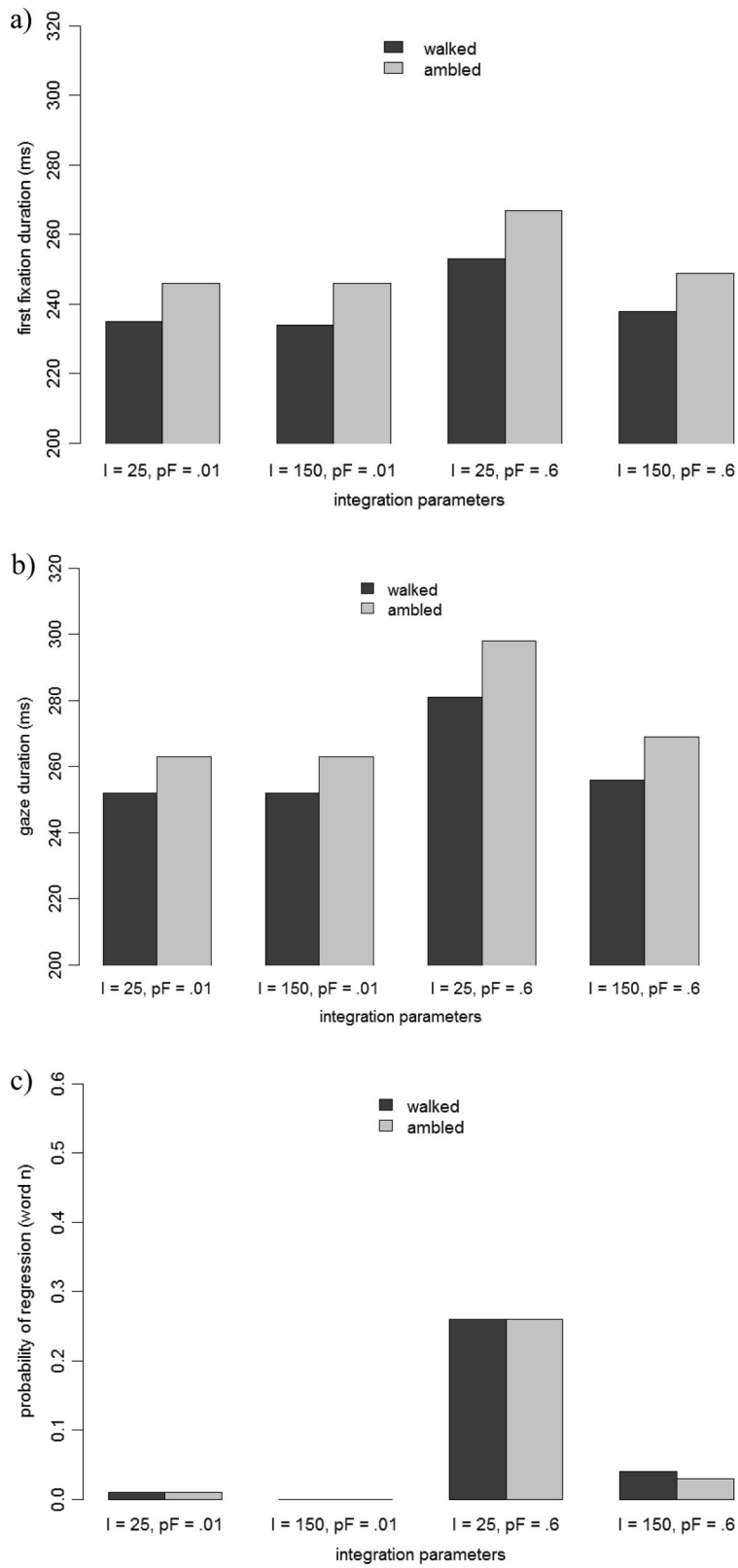

Figure 1. Simulated mean (a) first-fixation duration, (b) gaze duration, and (c) regression probabilities for the words *walked* and *ambled* across four settings of the integration parameters in E-Z Reader 10. I = integration; p_F, the probability of failure.

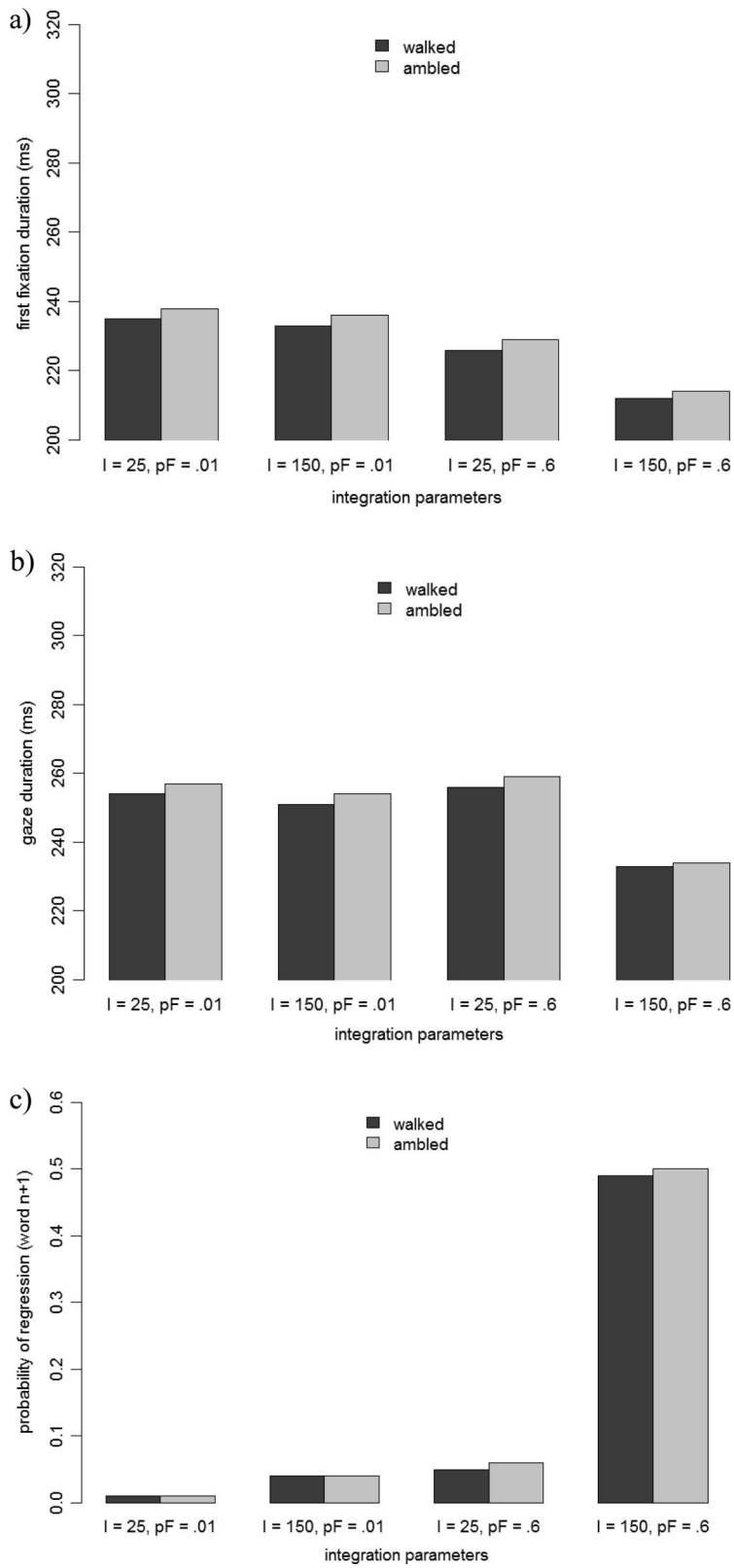

Figure 2. Simulated mean (a) first-fixation duration, (b) gaze duration, and (c) regression probabilities for the word *across*, following *walked* or *ambled*, across four settings of the integration parameters in E-Z Reader 10. I = integration; p_F, the probability of failure.

right, or both. The second set of bars in each panel shows the consequences of increasing I(n) to 150 ms (near the high end of the range explored by Reichle et al., 2009), while keeping $p_F(n)$ constant. It is evident that this change had essentially no effect on either first-fixation duration or gaze duration on the critical word, or on the probability of regressing from that word. This change also had no noticeable effect on reading times on the next word; its only effect was a small increase in the probability of regressing from that word. Note, however, that the same pattern of frequency effects was observed with these settings as with the default parameter settings.

The third set of bars in each panel shows the consequences of increasing $p_F(n)$ to .6, while returning I(n) to 25 ms. Now reading times were substantially longer on the critical verb, and the probability of regressing from the verb was dramatically increased. On the next word, reading times were not markedly different from the baseline condition, but the rate of regressions was slightly inflated. Again, the same frequency effects were in evidence.

Finally, the fourth set of bars shows the effects of increases in both integration parameters. An interesting pattern results, in which reading times on the critical verb returned to their baseline levels, and regressions from the verb were not very common. Reading times on the next word were actually very short, but regressions from this word were extremely likely. Once again, the same frequency effects appeared.

These simulations make clear, on the one hand, that there are complex interactions between the two integration parameters of E-Z Reader 10. For present purposes, however, the central point is that these simulations demonstrate the four predictions previously enumerated:(a) Frequency always affects reading times on the critical word; (b) integration difficulty, but not frequency, can influence the regression rate from this word; (c) integration difficulty may show up on the critical word itself but may also show up only downstream of the critical word; and (d) frequency and integration difficulty never have interactive effects on reading times on the critical word. In the present article, these predictions were tested by means of three eye movement experiments manipulating a target word's frequency and also manipulating whether the target word was easily integrated into its syntactic context. In Experiments 1 and 2, this latter manipulation was accomplished using garden-path sentences and, in Experiment 3, using an outright syntactic violation.

With respect to the specific prediction of noninteractivity between effects of frequency and integration difficulty on reading times, it is important to note that it is not the case that *only* a serial, staged process can generate a strictly additive pattern in temporal dependent measures such as fixation durations. It is well known that under some conditions, parallel or cascaded models can also generate additivity in reaction time (RT) measures (McClelland, 1979; Roberts & Sternberg, 1993; Townsend & Nozawa, 1995). Thus, support for a serial, staged architecture would arise from a combination of additivity in RT measures and confirmation of the other predictions previously enumerated, most notably the dissociations between lexical and syntactic effects in their timing and in the measures in which they appear.

Two other patterns emerge from these simulations that, while not the main focus of the present work, are worth noting in order to anticipate aspects of the experimental data to be presented later. The first is that the frequency manipulation resulted in only a very small spillover effect (Kliegl, Nuthmann, & Engbert, 2006; Rayner & Duffy, 1986) on the subsequent word; though reading times on *across* were slightly slower after *ambled* than after *walked*, this effect was very slight indeed, ranging from 1 ms to 5 ms. Second, note that while increasing the difficulty of integration did enhance the probability of a regression from *across*, it never increased reading times on this word. Indeed, the opposite pattern is in evidence: Both first-fixation duration and gaze duration were clearly shortest on *across* when $p_F(n) = .6$ and I(n) = 150. This is due to the fact that many inspections of this word will be rapidly terminated by a regressive eye movement, triggered by failed integration of the previous word.

Alternative Approaches

It should be noted that different theoretical assumptions do generate predictions that are clearly in opposition to those emerging from E-Z Reader 10. For example, Caplan and Waters (1999) have suggested that a single working memory resource (which, they proposed, is not involved in other verbal memory tasks) may underlie the various rapid, obligatory interpretive processes involved in language comprehension. This view predicts that "the effects of increasing the working memory demands of different interpretive operations are interactive" (Caplan & Waters, 1999, p. 93), and they proposed testing this idea by looking for a superadditive interaction (i.e., an interaction in which the joint manipulation of two variables produces a larger effect than the sum of their separate effects) "of variables that affect lexical access, such as frequency, and of variables that affect sentence interpretation, such as syntactic complexity" (p. 93). In other words, if the process of word recognition overlaps in time with, and shares a limited processing resource with, processes such as syntactic attachment, then word frequency effects should be larger when syntactic attachment is difficult.

Still another perspective is suggested by the recent turn toward probabilistic models of language processing. Jurafsky (1996) noted the traditional partitioning of language comprehension research into work on lexical access and syntactic disambiguation and remarked that this partitioning is based on the assumption that "the lexical and syntactic levels . . . are functionally distinct within the human language processor" (p. 138). Jurafsky proposed, instead, that "a single probabilistic mechanism underlies the access and disambiguation of linguistic knowledge at every level" (p. 186). Thus, while Caplan and Waters (1999) merely proposed that the processes of lexical access and syntactic attachment may not operate in a serial, staged manner, Jurafsky (1996) made the more radical suggestion that it may not be necessary to assume that these are even distinct cognitive processes.

Most recently, this idea has been embodied in the Surprisal model (Demberg & Keller, 2008; Hale, 2001; Levy, 2008). A detailed exposition of Surprisal is well beyond the scope of this article, but the central idea is as follows. Surprisal posits that the processing difficulty of a word is a function of how likely the word is to occur in context; the Surprisal value of a word reflects the difference between the probability of the string that constitutes the sentence up to, but not including, the word in question, and the probability of the string including the word in question; see Hale (2001) and Levy (2008) for different, but mathematically equivalent, formalizations of this idea. As this value increases, processing

difficulty monotonically increases. While this difference will be positive in essentially all cases, as a new input word is never perfectly predictable, processing will be especially difficult when the input string through word $n - 1$ is relatively probable, but the string through word n is much less probable. This may be because word n requires the comprehender to employ a phrase structure rule that is very unlikely given the preceding context or because word n itself is very low in frequency. Levy (2008) wrote that the model "links processing difficulty to the conditional probability of each word in its context. This encompasses lexical probabilities, so a rare word as a syntactically likely continuation may well be more surprising than a common word as a syntactically unlikely continuation" (p. 1165). While the model may remain relatively agnostic about precisely how lexical frequency and syntactic surprise interact (i.e., additively or not), the important point for present purposes is that lexical frequency and syntactic context both have an effect only by virtue of their relationship to conditional probability; it is conditional probability alone that determines processing difficulty. (The reader is referred to section 2.3 of Levy, 2008, which focuses on the notion of Surprisal as a *causal bottleneck*.) Thus, there should not be differences in the time course of frequency and syntactic context effects or in the behavioral measures in which they appear.[1]

Finally, it is important to note the results of a recent study by Tily, Federenko, and Gibson (2009) that appear to contradict some of the predictions enumerated earlier. In a phrase-by-phrase self-paced reading experiment, Tily et al. manipulated whether a sentence contained a subject cleft structure (Sentences 2a and 2b) or an object cleft structure (Sentences 2c and 2d) and whether the verb in the cleft construction was relatively high in frequency (Sentences 2a and 2c) or relatively low in frequency (Sentences 2b and 2d).

2a. It was Vivian who *lectured Terrence* for always being late. (669 ms)

b. It was Vivian who *chided Terrence* for always being late. (849 ms)

c. It was Vivian who *Terrence lectured* for always being late. (787 ms)

d. It was Vivian who *Terrence chided* for always being late. (864 ms)

The critical result from this experiment was an underadditive interaction in reading times on the two-word cleft region (italicized in the example; mean reading times for this region are shown to the right of each sentence). While cleft type and verb frequency both affected reading times as predicted, with longer times for object clefts and for low-frequency verbs, the effect of cleft type was large when the verb was high in frequency (118 ms) but very small when the verb was low in frequency (15 ms). On the next region of the sentence, there was a trend toward a reversal of this pattern. Tily et al. interpreted the critical interaction as suggesting a substantially delayed effect of syntactic integration difficulty when lexical processing is difficult. As shown, E-Z Reader 10 simulations do not predict this pattern for eye movements in reading; the timing with which syntactic integration effects appear is not affected by word frequency.

On the other hand, it is possible to interpret the Tily et al. interaction by focusing on the verb frequency effect, which was very large in the subject cleft conditions (180 ms) but smaller in the object cleft conditions (77 ms). This suggests that lexical processing may be relatively shallow, or even incomplete, in a context in which the word is difficult to integrate; perhaps the reason there was a smaller frequency effect in the object cleft conditions is that readers did not complete lexical access of the low-frequency verb (*chided*) once the initial stages of lexical processing, perhaps up through access of the word's syntactic category, indicated a difficult syntactic or semantic integration. Again, E-Z Reader 10 simulations do not predict this, as the frequency effect is the same size, on all measures, regardless of the degree (or kind) of integration difficulty. Thus, on either interpretation of the Tily et al. pattern, a replication of this interaction in eyetracking would be clearly inconsistent with E-Z Reader 10's predictions.

Thus, the predictions that emerge from the serial, staged architecture for lexical and integrative processing assumed by E-Z Reader 10 can be contrasted with predictions arising from other processing architectures and with an existing set of empirical results (Tily et al., 2009).

Overview of Experiments

In these experiments, subjects read normally while their eye movements were monitored. Each experimental sentence contained a critical past tense verb with the regular English *–ed* inflection. This verb could be either high frequency (e.g., *walked*) or low frequency (e.g., *ambled*); the properties of the critical verbs, which were the same across all three experiments, are described in detail later. The syntactic context in which the critical verb appeared was also manipulated. In Experiments 1 and 2, this verb either was difficult to attach into the phrase marker, by virtue of being presented at a point at which it resolved an ambiguous garden-path sentence toward its normally dispreferred analysis, or was relatively easy to attach, due to prior disambiguation of the sentence. In Experiment 1, the critical verb resolved the direct object/sentence complement ambiguity as in the simulation example earlier. In Experiment 2, the verb resolved the subordinate clause object/main clause subject ambiguity (e.g., Pickering & Traxler, 1998; Staub, 2007). Previous work (Sturt, Pickering, & Crocker, 1999) has shown that the latter of these garden-path types is substantially more difficult to resolve than is the former. In Experiment 3, by contrast, the critical verb was, in two of the conditions, not merely unexpected but actually ungrammatical at the point at which it appeared. Thus, across the three experiments, the difficulty of syntactic attachment of the critical word (in the difficult conditions) became progressively more pronounced: In Experiment 1, attaching the critical word required resolving a

[1] Roger Levy (personal communication, January 31, 2011) has remarked that, "[I]n the strictest interpretation of Surprisal, lexical frequency is an epiphenomenon as far as sentence processing is concerned: The Surprisal of a word will *on average* look something like its frequency, and if one has a poor understanding of how context affects the probability distribution over next words, lexical frequency is probably a decent rough guess."

relatively easy garden path; in Experiment 2, attaching the critical word required resolving a more difficult garden path; and in Experiment 3, attaching this word into a connected phrase marker was ultimately impossible.

Experiment 1

Method

Subjects. Thirty-six students at the University of Massachusetts Amherst participated in the experiment in exchange for course credit. Eight subjects had to be excluded from data analysis due to poor calibration or excessive track loss; these subjects were replaced. All subjects were native speakers of English, had normal or corrected-to-normal vision, and reported no reading or other language-related disorders.

Materials. Forty-eight critical verbs were selected for inclusion as target words in Experiment 1; the same 48 verbs were employed in Experiments 2 and 3. These verbs consisted of 24 pairs, with each pair consisting of one high-frequency (HF) and one low-frequency (LF) verb. The verbs in each pair were matched on length in characters. Length ranged from 5 to 10 characters, $M = 7.46$, $SD = 1.22$. Frequencies were obtained from the 400-million-word Hyperspace Analogue to Language (HAL) corpus (Burgess & Livesay, 1998), available in the English Lexicon Project database (Balota et al., 2002). Word forms as opposed to stem frequencies were used in item selection; the correlation between word forms and stem frequencies across the 48 verbs was .90. The mean corpus frequency of the HF verbs was 15,024, while the mean corpus frequency of the LF verbs was 1,159; thus, on average, the HF verbs were about 13 times as frequent as the LF verbs. Mean log(frequency) for the HF verbs was 8.91, $SD = 1.35$; mean log(frequency) for the LF verbs was 5.61, $SD = 1.70$; $p < .001$.

Each of the verb pairs was embedded in an item set like the one shown in Sentence 3:

3a. The professor saw that the students walked across the quad. (HF, unambiguous)

b. The professor saw that the students ambled across the quad. (LF, unambiguous)

c. The professor saw the students walked across the quad. (HF, ambiguous)

d. The professor saw the students ambled across the quad. (LF, ambiguous)

To ensure that the target word was not predictable in context, experimenters presented the sentences up to the critical word (with the word *that* included) in a cloze norming study. Nine subjects provided completions for each sentence. In not a single instance was the target verb elicited; thus, it is safe to conclude that these verbs were not predictable.

The 24 item sets, which are provided in the Appendix, were arranged in four lists, so that each subject would read six items in each condition, and each item would be seen in one of its versions by each subject. The 24 experimental items were intermixed with 96 unrelated filler items, none of which involved the direct object/sentence complement ambiguity employed in the experimental items. The items were presented in an individually randomized order to each subject after eight practice sentences. Simple two-choice comprehension questions were presented after approximately one fourth of items (including fillers) to ensure that subjects were reading for comprehension. Subjects averaged 88% correct on these questions.

Procedure. Subjects were tested individually, and eye movements were recorded using an EyeLink 1000 (SR Research, Toronto, ON, Canada) eyetracker, interfaced with a PC computer. The sampling rate was 1000 Hz. Stimuli were displayed on a CRT monitor (Iiyama Corp., Tokyo, Japan). Subjects were seated 55 cm from the computer screen. The angular resolution of the eyetracker is 10–30 min of arc (substantially less than one character). Viewing was binocular, but only the movement of the right eye was recorded. All critical sentences were displayed on a single line. Sentences were presented in 12-point Monaco font. Before the experiment began, each subject was instructed to read for comprehension in a normal manner. A calibration procedure was then performed, and recalibration was carried out between trials as needed. The subject triggered the onset of each sentence by fixating a box on the left edge of the computer screen. The experiment lasted approximately 45 min. The experiment was implemented with the EyeTrack software, and initial stages of data analysis were carried out with EyeDoctor and EyeDry software (http://www.psych.umass.edu/eyelab/software/).

Results

Prior to analysis, trials on which there was a track loss or any other data collection error (5.6% of trials) were deleted. In addition, eye fixations less than 80 ms in duration and within one character of the previous or subsequent fixation were incorporated into this neighboring fixation. Remaining fixations shorter than 80 ms (less than 1.7% of fixations) or longer than 1,000 ms (three fixations in total; none on the critical verb) were deleted.

For the purpose of analysis of the eye movement data, the critical sentences were divided into four regions as illustrated in Sentence 4:

4. The professor saw (that) the students| walked| across| the quad.

Analysis of eye movement data focused on the second region, the *critical* region, the third region, the *spillover* region, and the *final* region.

It should be noted that though E-Z Reader's predictions are defined over individual words, it was not possible to define a useful one-word spillover region for many items. Because the word following the critical word was, in general, extremely short (*across* was one of the longest; in many items, this word was an article; see Drieghe, Rayner, & Pollatsek, 2005, for data on word skipping and word length), the overall skipping rate for this word was 51% in Experiment 1 (49%–53% across conditions), and 60% in Experiment 2 (57%–63% across conditions). Thus, the paucity of data precluded meaningful statistical analyses of reading times, as many subjects, and many items, had no observations in one or more conditions. Thus, the spillover region that was used here consisted of two words, except for items in which the word

immediately following the critical word was five or more letters in length. Data from material downstream of the critical word were directly relevant to only one of the critical predictions made by E-Z Reader, namely, the prediction that integration difficulty may show up only downstream; for testing the prediction in this general form, it is not critical whether this difficulty shows up on word $n + 1$ or word $n + 2$.

For each region, the following reading time measures were computed. *First-fixation duration* is the duration of the reader's first eye fixation on a region for those trials on which the region was fixated on the reader's first pass through the sentence. *Gaze duration* is the sum of all first-pass fixations on a word before the eyes leave it for the first time, either to the left or to the right; this measure is known as *first-pass time* when the region consists of multiple words. *Go-past time* is the sum of all fixations beginning with the first fixation on a region (again, including only those trials on which the reader made a first-pass fixation) until the reader leaves it to the right, including any time spent to the left of the region as a result of regressive eye movements and any time spent rereading the region before moving on to the rest of the sentence. For the critical word only, two other reading time measures were computed: the duration of the last fixation before fixating the word, which was examined in order to determine whether there was any evidence of a so-called parafoveal-on-foveal effect (i.e., an effect of a manipulation of a word that is manifested before the eyes have fixated the word directly; Kennedy & Pynte, 2005; cf. Drieghe, Rayner, & Pollatsek, 2008), and *single-fixation duration*, which is identical to first-fixation duration but is restricted to trials in which the reader made exactly one fixation on the word and then moved on to the right. See Staub and Rayner (2007) for discussion of these and other eye movement measures. To assess the statistical reliability of the effects on reading time of frequency, ambiguity, and their interaction, analyses of variance (ANOVAs) with subjects (F_1) and items (F_2) as random factors were conducted for each measure.

An additional analysis, for each region, determined whether the reader's first pass through the region ended with a regressive saccade to an earlier region of the sentence, rather than a forward saccade. For the critical word, the proportion of trials on which the word was skipped on first-pass reading is also reported. The regressions and skipping effects were analyzed using mixed-effects logistic regression, with (centered) ambiguity and verb frequency and their interaction as fixed effects, and with random intercepts for subjects and items, as well as random slopes for subjects and items associated with each of the fixed effects and their interaction. (See Jaeger, 2008, for extended discussion of the advantages of this approach over ANOVA for categorical dependent measures). Analyses were carried out with R, an open source programming language and environment for statistical computing (R Development Core Team, 2007), and in particular with the lme4 package for linear mixed effects models (Bates, 2005; Bates & Sarkar, 2007).

The critical data patterns are illustrated in Figures 3 and 4. Figure 3 illustrates, for each condition, the reading time means for first-fixation duration, single-fixation duration, and gaze duration on the critical word, and Figure 4 illustrates the regression rate from the critical word, the spillover region, and final region. The means on the full set of measures are presented in Tables 1, 2, and 3. The results of the statistical analyses are presented in Tables 4 and 5.

The results for the critical word will be discussed first. There were no significant effects of the experimental manipulations on the prior fixation measure or on the probability of skipping this word. However, all reading time measures on this word were influenced by frequency, with significantly longer times for LF words. There were no significant effects of the ambiguity manipulation on these measures, and no interaction effects. The probability of a regression out of the critical word was not influenced by either manipulation or by their interaction. In sum, frequency affected reading times on the critical word but not the regression rate, while neither ambiguity nor the interaction of the two factors significantly affected any measure on this word.

On the spillover region, frequency did not significantly affect either first-fixation duration or first-pass time, and while this factor had a significant effect on go-past time on the spillover region in the items analysis, with longer times in the LF conditions, this

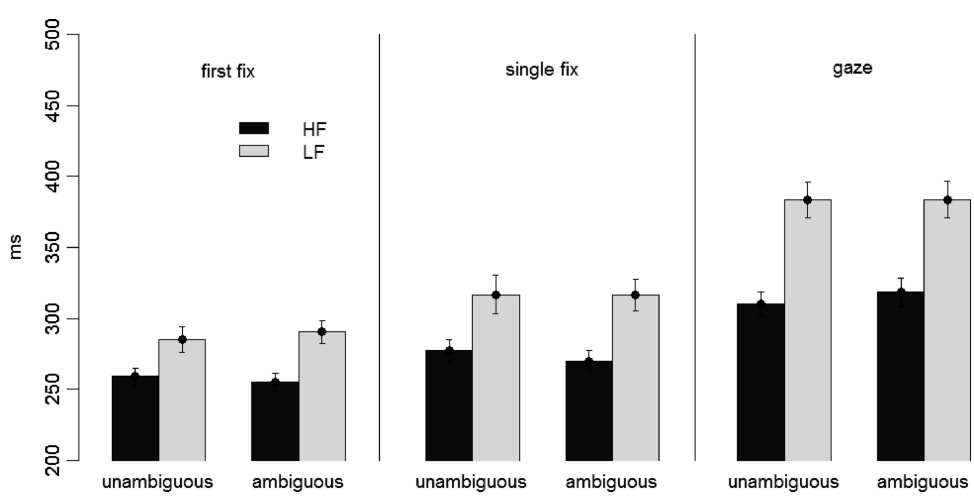

Figure 3. Experiment 1 mean reading times, by condition, on the critical word. Error bars represent standard error of the mean. HF = high frequency; LF = low frequency.

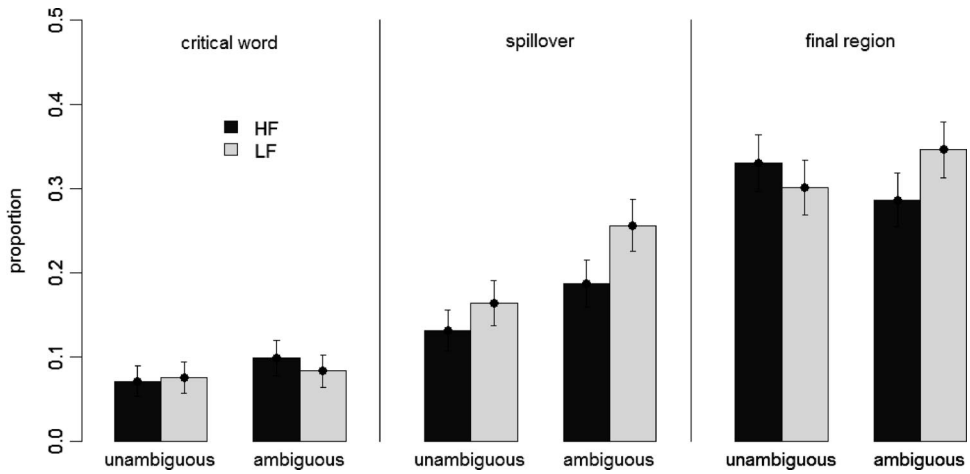

Figure 4. Experiment 1 regression proportions, by condition, on the critical word, spillover region, and final region. Error bars represent standard error of the proportion. HF = high frequency; LF = low frequency.

effect was only marginal (p = .09) in the subjects analysis. While there was a numerical trend toward more regressions from the spillover region following LF verbs, this was not significant (p = .16). The ambiguity manipulation also did not affect first-fixation duration or first-pass time, but regressions from the spillover region were significantly more likely in the ambiguous condition, and go-past times were significantly longer in this condition. There were no significant or marginal interaction effects. On the final region, there were no significant effects of frequency, except for go-past time in the subjects analysis; however, this effect did not approach significance in the items analysis (p = .40). Ambiguity had a significant effect on first-fixation duration on this region in the subjects analysis, though this was only marginal (p = .06) in the items analysis. There were no other significant effects of ambiguity, and there were no significant interaction effects. In short, the only fully reliable downstream effects were main effects of ambiguity on the regression rate from the spillover region and on go-past time on this region.

Discussion

The results from this experiment are easily summarized. The frequency of the critical verb influenced reading times on the verb itself, but it did not influence the probability of making a regressive eye movement and had no significant effects downstream from the verb itself. On the other hand, the difficulty of integrating this verb into the phrase marker did not affect any eye movement measure on the verb itself; its influence appeared downstream, once the eyes had moved past the verb to the right. There were more regressive eye movements from the spillover region, and consequently longer go-past times, when the critical verb resolved a garden path. There were no interaction effects.

These results suggest several conclusions. First, when frequency effects and garden-path effects are both triggered by the same word, these effects can indeed show strict temporal sequencing in the eye movement record. This conclusion has been implicit in the literature, as frequency effects essentially always appear immediately, while garden-path effects are sometimes delayed; however, this may be the first time that this pattern has been demonstrated in a single experiment. As discussed earlier, within the E-Z Reader 10 framework, an effect of syntactic attachment difficulty that appears only downstream of the critical word arises at several locations in parameter space (see also Reichle et al., 2009, Figure 4). Second, the difficulty of lexical access and the difficulty of syntactic attachment may indeed play functionally distinct roles

Table 1
Means for Reading Time Measures and Proportions for Regressions and Skipping Measures by Condition on Critical Verbs in Experiment 1

	Fixation						
Sentence/frequency	Prior	First	Single	Gaze	Go-past	p(skipping)	p(regression)
Unambiguous							
HF	207	259	277	310	353	.02	.07
LF	210	285	317	383	432	.01	.08
Ambiguous							
HF	215	255	270	318	374	.01	.10
LF	214	290	317	384	432	.02	.08

Note. Means are given in milliseconds. HF = high frequency; LH = low frequency; p = proportion.

Table 2
Means for Reading Time Measures and Proportions for Regressions by Condition on Spillover Region in Experiment 1

Sentence/frequency	First-fixation	First-pass	Go-past	p(regression)
Unambiguous				
HF	241	381	450	.13
LF	248	380	506	.16
Ambiguous				
HF	253	386	536	.19
LF	248	394	594	.26

Note. Means are given in milliseconds. HF = high frequency; LH = low frequency.

with respect to eye movement control, as in this experiment they showed qualitatively distinct effects, with the former influencing first-pass reading time and the latter influencing the probability of a regression. As illustrated in Figures 1–2, there are multiple locations in E-Z Reader parameter space where integration difficulty will have its primary effect on the regression rate.

Though the results obtained in Experiment 1 are not unexpected given patterns in the previous literature, the results together with this previous literature provide a challenge for the probabilistic models briefly discussed earlier. The findings that behavioral indices of the difficulty of lexical and syntactic processing of the same word can be strictly temporally sequenced and evident in distinct behavioral measures suggest that there is at least a partial functional distinction between the cognitive processes involved in lexical access and those involved in syntactic attachment (cf. Jurafsky, 1996).

Before presenting Experiment 2, two additional aspects of the results of Experiment 1 should be noted. First, it may be seen as surprising that the effect of the frequency of the critical verb did not spill over onto the subsequent region, as in many experiments fixations are lengthened on the word following an LF word (Kliegl et al., 2006). This pattern is not universal, however (e.g., Drieghe et al., 2008), and as shown earlier, the E-Z Reader 10 simulations actually predicted spillover effects of only a few milliseconds. Second, it may also be seen as surprising that the garden-path manipulation resulted in an increased rate of regression from the spillover region but had little effect on the first-fixation or first-pass measures on this region. Again, this unintuitive result was actually predicted by the E-Z Reader 10 simulations.

Table 3
Means for Reading Time Measures and Proportions for Regressions by Condition on Final Region in Experiment 1

Sentence/frequency	First-fixation	First-pass	Go-past	p(regression)
Unambiguous				
HF	229	625	1003	.33
LF	229	595	921	.30
Ambiguous				
HF	246	637	1081	.29
LF	250	608	1007	.34

Note. Means are given in milliseconds. HF = high frequency; LH = low frequency; p = proportion.

Table 4
Results of by-Subject (F_1) and by-Item (F_2) Analyses of Variance for Reading Time Measures in Experiment 1

Variable	$F_1(1, 35)$	p	$F_2(1, 23)$	p
Critical word				
Prior fixation				
Frequency	0.05	.83	0.26	.62
Ambiguity	0.49	.49	1.22	.28
Frequency × Ambiguity	0.20	.66	0.18	.67
First fixation				
Frequency	14.24	<.001	20.87	<.001
Ambiguity	0.00	.99	0.00	.96
Frequency × Ambiguity	0.20	.65	0.65	.43
Single fixation				
Frequency			19.05	<.001
Ambiguity			0.34	.57
Frequency × Ambiguity			0.03	.88
Gaze duration				
Frequency	29.91	<.001	40.81	<.001
Ambiguity	0.03	.85	0.46	.50
Frequency × Ambiguity	0.89	.35	0.09	.76
Go-past time				
Frequency	17.06	<.001	9.85	<.01
Ambiguity	0.52	.48	0.85	.37
Frequency × Ambiguity	0.98	.33	0.39	.54
Spillover region				
First fixation				
Frequency	0.22	.64	0.02	.89
Ambiguity	0.99	.33	0.36	.55
Frequency × Ambiguity	1.03	.32	1.19	.29
First pass time				
Frequency	0.02	.89	0.02	.88
Ambiguity	0.61	.44	0.22	.64
Frequency × Ambiguity	0.01	.93	0.23	.64
Go-past time				
Frequency	3.13	.09	7.64	.01
Ambiguity	10.84	<.01	18.89	<.001
Frequency × Ambiguity	0.00	.96	0.07	.79
Final region				
First fixation				
Frequency	0.41	.53	0.00	.98
Ambiguity	5.02	.03	4.07	.06
Frequency × Ambiguity	0.07	.79	0.00	.96
First pass time				
Frequency	1.88	.18	0.71	.67
Ambiguity	0.08	.78	0.19	.41
Frequency × Ambiguity	0.00	.99	0.00	.98
Go-past time				
Frequency	4.50	.04	0.72	.40
Ambiguity	3.30	.08	1.33	.26
Frequency × Ambiguity	0.00	.96	0.20	.96

Note. For single fixation, only F_2 is shown as some subjects had no single fixations in one or more cells.

It may be noted that the possibility of a still another syntactic analysis, a "small clause" analysis (e.g., "The professor saw the students walk"), may have played a role in the comprehension of the ambiguous sentences. In principle, subjects could have adopted such an analysis (though for only some of the items, specifically those with a main verb of perception such as *see*, *hear*, or *notice*), overlooking the past tense morphology on the verb, rather than analyzing the sentence as intended (i.e., as containing a full embedded clause in the past tense). However, the fact that effects of the garden-path manipulation did emerge on the immediately succeeding spillover region suggests that readers were not simply

Table 5
Mixed-Effects Logistic Regression Model Estimates of Effects of Frequency and Ambiguity and Their Interaction on Skipping of Critical Word and Regressions From Critical Word, Spillover Region, and Final Region in Experiment 1

Variable	Estimate (log odds)	*SE*	*z*	*p*
Critical word: Skipping				
Frequency	1.66	22.91	0.07	.94
Ambiguity	−3.57	23.54	−0.15	.88
Frequency × Ambiguity	−2.61	46.33	0.06	.96
Critical word: Regressions				
Frequency	−0.25	0.36	−0.70	.48
Ambiguity	0.47	0.30	01.57	.12
Frequency × Ambiguity	0.09	0.61	0.16	.88
Spillover: Regressions				
Frequency	0.39	0.28	01.40	.16
Ambiguity	0.67	0.23	02.94	<.01
Frequency × Ambiguity	0.19	0.58	0.33	.74
Final region: Regressions				
Frequency	0.18	0.27	0.68	.50
Ambiguity	0.00	0.19	0.01	.99
Frequency × Ambiguity	0.50	0.46	01.09	.27

ignoring the past tense morphology on the embedded verb and were forced to re-analyze on at least some occasions.

The goal in conducting Experiment 2 was to generate an earlier garden-path effect in order to test the predictions of E-Z Reader 10 that pertain to cases in which frequency and integration effects appear on the same word (i.e., that only integration difficulty will increase the regression rate, and that frequency and integration effects on reading times will be strictly additive). In Experiment 2, the same verbs were used to disambiguate the subordinate clause object/main clause subject ambiguity. Previous research has shown that it is more difficult to revise an analysis on which an ambiguous noun phrase is the subordinate clause object to one on which it is the main clause subject than it is to complete the revision that was required in Experiment 1 (Sturt et al., 1999). More important for present purposes, previous eyetracking experiments in which this ambiguity was employed have shown effects as early as the first fixation on the disambiguating material (e.g., Staub, 2007).

Experiment 2

Method

Participants. Thirty-six students at the University of Massachusetts Amherst participated in the experiment in exchange for course credit. Two subjects had to be excluded from data analysis due to poor calibration or excessive track loss; these subjects were replaced. All subjects were native speakers of English, had normal or corrected-to-normal vision, and reported no reading or other language-related disorders.

Materials. The same 24 verb pairs used in Experiment 1 were used to form item quadruplets as in Sentence 5:

5a. While the professor lectured, the students walked across the quad. (HF, unambiguous)

b. While the professor lectured, the students ambled across the quad. (LF, unambiguous)

c. While the professor lectured the students walked across the quad. (HF, ambiguous)

d. While the professor lectured the students ambled across the quad. (LF, ambiguous)

The critical disambiguation in the Sentence 5a and Sentence 5b versions of each item was achieved by the inclusion of a comma at the clause boundary. As for Experiment 1, cloze norms were obtained for each of the unambiguous contexts, with nine participants providing continuations for each item. There was only a single instance in which a subject provided a form of the target verb as the next word of the sentence.

The 24 item sets were arranged in four lists, so that each subject would read six items in each condition, and each item would be seen in one of its versions by each subject. The 24 experimental items were intermixed with 100 unrelated filler items, none of which involved the subordinate clause object/main clause subject ambiguity. The items were presented in an individually randomized order to each subject after eight practice sentences. Two-choice comprehension questions were presented after approximately one third of items. Subjects averaged 83% correct on these questions; note that this figure is influenced by the fact that some filler items, associated with unrelated experiments, were designed to be confusing or ambiguous.

Procedure. The procedure was identical to Experiment 1, with the exception that Courier New font was used, due to requirements of an unrelated filler experiment.

Results

Track loss resulted in deletion of 3.9% of trials. Less than 1.1% of fixations fell below the 80-ms cutoff, while a total of seven fixations, none of which fell on the critical verb, exceeded the 1,000-ms cutoff.

For the purpose of analysis of the eye movement data, the critical sentences were again divided into four regions, as illustrated in Sentence 6:

6. While the professor lectured(,) the students| walked| across| the quad.

The same eye movement measures were computed as in Experiment 1, and the same statistical models were applied. The means for each measure are shown in Tables 6, 7, and 8, and the results of the statistical analyses are shown in Tables 9 and 10. The reading time means for the critical word are illustrated in Figure 5, and the regression proportions for each region are shown in Figure 6.

For the fixation prior to fixation on the critical word, there was no effect of frequency. There was a marginal effect of ambiguity in both subject ($p = .09$) and item ($p = .06$) analyses, with longer times in the ambiguous conditions. There was a significant interaction in the items analysis but not in the subjects analysis ($p = .14$). The rate of skipping of the critical word was not significantly affected by either factor, but it was affected by their interaction; there was more skipping of LF verbs when the sentence was unambiguous, and more skipping of HF verbs when the sentence was ambiguous.

Table 6
Means for Reading Time Measures and Proportions for Regressions and Skipping Measures by Condition on Critical Verbs in Experiment 2

	Fixation						
Sentence/frequency	Prior	First	Single	Gaze	Go-past	p(skipping)	p(regression)
Unambiguous							
HF	224	242	251	287	315	.02	.05
LF	214	271	289	353	399	.05	.09
Ambiguous							
HF	222	285	296	345	495	.07	.19
LF	233	304	322	392	558	.03	.16

Note. Means are given in milliseconds. HF = high frequency; LH = low frequency; p = proportion.

All reading time measures on the critical word were significantly affected by both factors, with longer times in the LF conditions and in the ambiguous conditions. However, there were no significant or marginal interaction effects on the reading time measures. The probability of regression from the critical word was significantly affected by ambiguity, as regressions were more likely in the ambiguous conditions. There was a marginal effect of frequency ($p = .08$) on the regression rate, but note that this effect is not in the direction that would be expected if lexical processing difficulty can trigger a regression: The sign of the parameter estimate in the logistic regression model indicates that regressions were less likely from LF verbs.

On the spillover region, there were no significant or marginal effects of frequency on the reading time measures, and ambiguity affected only go-past time. The interaction effects did not approach significance. However, both factors influenced the regression rate, and there was a significant interaction between them. This interaction is analyzed in more detail later. On the final region, the only significant reading time effect was, once again, an ambiguity effect on go-past time. Again, there were significant effects of both factors on the regression rate and a significant interaction.

As Figure 6 illustrates, the interaction on the spillover region was due to the fact that in the ambiguous conditions, the regression rate was high following both HF and LF verbs, while in the unambiguous conditions, the regression rate was higher for LF verbs than for HF verbs. On the final region, there is a very surprising pattern, with more regressions occurring in sentences with HF verbs than in sentences with LF verbs and with this reverse frequency effect being somewhat larger for the unambiguous sentences. It is interesting that neither frequency nor its interaction with ambiguity affected the go-past measure; only the garden-path manipulation had a significant effect on go-past time in both regions.

To further investigate these regression patterns, an additional mixed-effects logistic regression analysis combined the spillover and final regions into one region. For this combined region, the pattern was somewhat similar to the pattern for the spillover region alone, though less extreme: The probability of a regression was .30 in both of the ambiguous conditions, .23 in the LF unambiguous condition, and .15 in the HF unambiguous condition. This pattern resulted in a significant effect of ambiguity ($p < .001$), a nonsignificant frequency effect ($p = .13$), and a significant interaction ($p < .01$).

Discussion

Experiment 2 successfully elicited an early garden-path effect, with evidence of syntactic processing difficulty showing up both in the duration of the very first fixation on the disambiguating word and in later reading time measures. The garden-path manipulation also affected the rate of regressive eye movements from the critical word. As in Experiment 1, all reading time measures on the critical word were inflated for LF verbs, but the probability of a regression was not higher for these verbs; indeed, it was marginally lower. As predicted by E-Z Reader, the frequency and garden-path effects on reading times did not interact; the size of the frequency effect was not modulated by the presence of a garden-path effect. The frequency effect tended to be somewhat smaller for the ambiguous

Table 7
Means for Reading Time Measures and Proportions for Regressions by Condition on Spillover Region in Experiment 2

Sentence/frequency	First-fixation	First-pass	Go-past	p(regression)
Unambiguous				
HF	239	323	356	.06
LF	247	322	425	.19
Ambiguous				
HF	246	336	569	.21
LF	259	331	585	.22

Note. Means are given in milliseconds. HF = high frequency; LH = low frequency; p = proportion.

Table 8
Means for Reading Time Measures and Proportions for Regressions by Condition on Final Region in Experiment 2

Sentence/frequency	First-fixation	First-pass	Go-past	p(regression)
Unambiguous				
HF	240	631	867	.24
LF	235	655	822	.16
Ambiguous				
HF	242	618	1023	.27
LF	253	651	995	.23

Note. Means are given in milliseconds. HF = high frequency; LH = low frequency; p = proportion.

sentences than for the unambiguous ones (for first-fixation duration, 19 ms vs. 29 ms; for single-fixation duration, 26 ms vs. 38 ms; for gaze duration, 47 ms vs. 66 ms), but this interaction never approached statistical significance.

In sum, the present experiment confirmed the additive patterns on the critical verb that E-Z Reader 10 predicted would arise when both frequency effects and integration effects appeared in the same measures. The results also confirmed that when frequency and integration difficulty both affect reading times on a word, only integration difficulty significantly affects the probability of a regressive eye movement. In addition, the results confirmed some less central predictions: Spillover effects of word frequency on reading times on the subsequent region were very small (indeed, nonsignificant), and downstream effects of integration difficulty took the form of an increase in regressions, rather than first-fixation or first-pass time effects.

There are a few minor puzzles in the data, however. First, though the critical word was rarely skipped in general, there was an interaction effect on this probability. This is would not be expected in any existing account, so any interpretation should await an assessment of the reliability of this effect. Second, the frequency and interaction effects on the rate of regressions from the spillover and final regions were not predicted by E-Z Reader 10 (see Figure 2) and were somewhat puzzling in their own right. The analysis with the combined region revealed a pattern that was clearly more interpretable than the patterns on the original spillover and final regions, but there was still an interaction effect, with verb frequency affecting the probability of a downstream regression when the sentence was unambiguous. Given that there was also a nonsignificant trend toward a frequency effect on regressions from the spillover region in Experiment 1, it does appear that E-Z Reader 10 may be missing a real, though relatively small, effect of word frequency on the probability of a regression from subsequent material. Whether the regressions effect is indeed reliable and, if so, how it should be handled within the E-Z Reader framework will be left as questions for future research.

However, additional analyses of the postcritical regions in Experiment 1 and Experiment 2 supported the contention that overall, frequency has essentially no effect on processing difficulty beyond the critical word, and that frequency and integration difficulty do not have interactive effects on processing difficulty beyond the critical word. When the two downstream regions in Experiment 1 were combined into a single region, mean go-past times (arguably the most complete measure of processing difficulty) for this region were as follows: HF unambiguous, 1,410 ms; LF unambiguous,

Table 9
Results of by-Subject (F_1) and by-Item (F_2) Analyses of Variance for Reading Time Measures in Experiment 2

Variable	$F_1(1, 35)$	p	$F_2(1, 23)$	p
Critical word				
Prior fixation				
Frequency	0.02	.90	0.04	.84
Ambiguity	2.95	.09	3.76	.06
Frequency × Ambition	2.22	.14	4.66	<.05
First fixation				
Frequency	21.99	<.001	8.88	<.01
Ambiguity	18.39	<.001	24.11	<.001
Frequency × Ambition	0.08	.78	1.48	.24
Single fixation				
Frequency			8.27	<.01
Ambiguity			10.04	<.01
Frequency × Ambition			2.09	.16
Gaze duration				
Frequency	25.63	<.001	18.30	<.001
Ambiguity	17.27	<.001	19.44	<.001
Frequency × Ambition	0.83	.37	1.34	.26
Go-past time				
Frequency	7.78	<.01	4.21	.052
Ambiguity	14.62	<.001	33.90	<.001
Frequency × Ambition	0.16	.69	0.14	.72
Spillover region				
First fixation				
Frequency	1.40	.25	2.65	.12
Ambiguity	1.47	.23	1.05	.32
Frequency × Ambition	0.22	.64	0.00	.97
First pass time				
Frequency	0.10	.75	0.00	.96
Ambiguity	0.43	.52	1.13	.30
Frequency × Ambition	0.01	.94	0.19	.67
Go-past time				
Frequency	1.00	.32	0.95	.34
Ambiguity	19.24	<.001	16.22	<.001
Frequency × Ambition	1.69	.20	0.51	.48
Final region				
First fixation				
Frequency	0.13	.72	0.32	.58
Ambiguity	2.32	.14	2.68	.12
Frequency × Ambition	0.94	.34	1.54	.23
First pass time				
Frequency	0.90	.35	1.22	.28
Ambiguity	0.11	.74	0.26	.61
Frequency × Ambition	0.06	.82	0.02	.89
Go-past time				
Frequency	0.68	.41	0.17	.68
Ambiguity	11.82	<.01	5.37	<.05
Frequency × Ambition	0.15	.70	0.00	.96

Note. For single fixation, only F_2 is shown, as some subjects had no single fixations in one or more cells.

Table 10
Mixed-Effects Logistic Regression Model Estimates of Effects of Frequency and Ambiguity and Their Interaction on Skipping of Critical Word and Regressions From Critical Word, Spillover Region, and Final Region in Experiment 2

Variable	Estimate (log odds)	*SE*	*z*	*p*
Critical word: Skipping				
Frequency	.02	.45	0.06	.96
Ambiguity	.39	.46	0.85	.40
Frequency × Ambiguity	−2.19	.90	−2.43	<.05
Critical word: Regressions				
Frequency	−.77	.43	−1.77	.08
Ambiguity	2.04	.44	4.67	<.001
Frequency × Ambiguity	.64	.75	0.86	.39
spillover: regressions				
Frequency	.56	.27	2.12	<.05
Ambiguity	.86	.24	3.64	<.001
Frequency × Ambiguity	−1.38	.53	−2.58	<.01
Final region: Regressions				
Frequency	−.89	.28	−3.20	<.01
Ambiguity	.73	.27	2.72	<.01
Frequency × Ambiguity	1.03	.47	2.20	<.05

Note. The model for the skipping analyses did not include random slopes, as this larger model failed to converge.

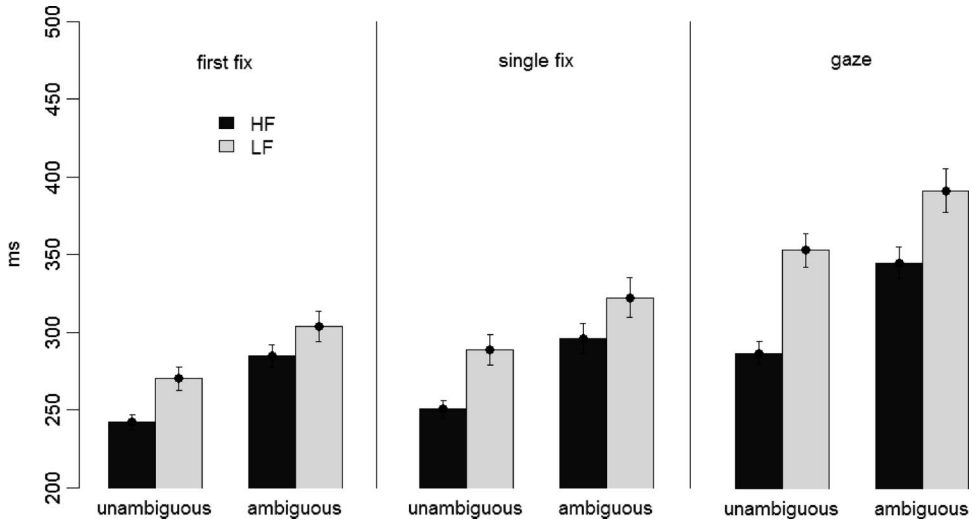

Figure 5. Experiment 2 mean reading times, by condition, on the critical word. Error bars represent standard error of the mean. HF = high frequency; LF = low frequency.

1,389 ms; HF ambiguous, 1,571 ms; and LF ambiguous, 1,560 ms. For the combined downstream region in Experiment 2, go-past times were as follows: HF unambiguous, 1,156 ms; LF unambiguous, 1,169 ms; HF ambiguous, 1,486 ms; and LF ambiguous, 1,489 ms. In both experiments, there was a very large effect of ambiguity (ps $< .01$), no hint of a frequency effect (ps $> .5$), and no hint of an interaction effect (ps $> .5$).

Turning back to the main results, two obvious questions emerge. The first arises from the fact that the lack of an interaction effect on reading times on the critical word is essentially a null effect; the absence of such an effect involves a failure to find a significant interaction between the effects of frequency and ambiguity. Thus, it is important to consider whether the experiment simply lacked the power to reveal a small, but theoretically important, underadditive interaction, as observed in self-paced reading by Tily et al. (2009). One possibility is that the underadditive trend would become more pronounced if syntactic integration difficulty was made even more extreme; perhaps there is some degree of integrative processing difficulty that could effectively terminate lexical processing, resulting in a small or nonexistent frequency effect. Experiment 3 allowed for an evaluation of this possibility.

The second question may be viewed as orthogonal to the main question of interest in the present article: Why there was such a pronounced difference in the timing of the garden-path effects in Experiment 1 and Experiment 2? Experiment 1 showed strict sequencing of frequency and garden-path effects, while in Experiment 2, the two effects appeared in the same early measures on the critical verb. As noted earlier, the garden path used in Experiment 2 has been shown to be substantially more difficult to resolve; this difference has been attributed to a difference in the

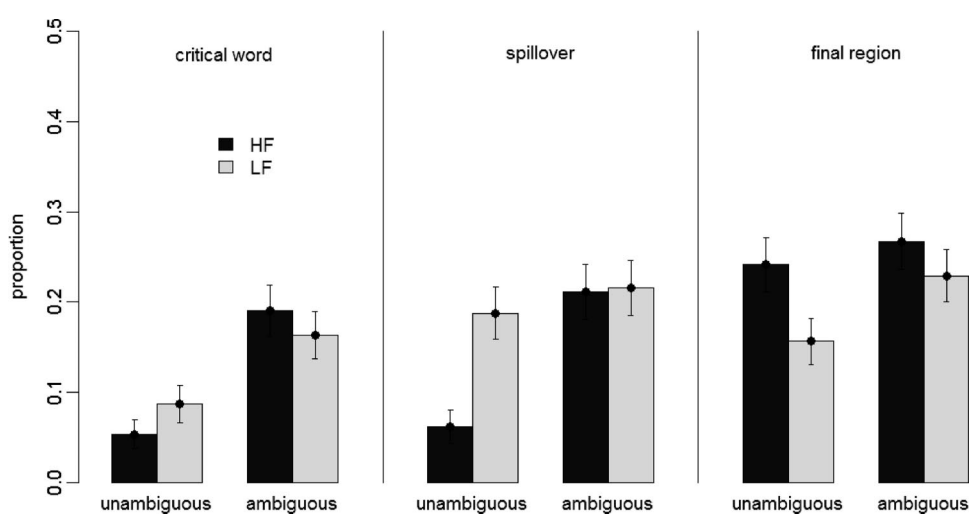

Figure 6. Experiment 2 regression proportions, by condition, on the critical word, spillover region, and final region. Error bars represent standard error of the proportion. HF = high frequency; LF = low frequency.

type of structural revision that is required in the two cases (Sturt et al., 1999) or to the fact that a prosodic revision is necessary in the second case but not in the first (Bader, 1998). Comparison of the overall amount of garden-path disruption in the two experiments shows that the difficulty difference emerged here as well. However, it is important to recognize that a difference in reanalysis *difficulty* does not, by itself, predict a difference in reanalysis *timing*; the speed with which the syntactic processor detects a problem and the amount of difficulty that eventually follows may or may not be directly related. In other words, the fact that the garden path in Experiment 2 caused more overall disruption does not explain why this disruption appeared earlier in the eye movement record.

It is worth noting that, in a review of a large number of eye movement studies of syntactic ambiguity resolution, Clifton et al. (2007) were not able to identify specific factors governing the timing with which garden-path effects first appeared. One possibility (Staub, 2009) is that garden-path effects are likely to appear earlier if, at the point of disambiguation, the parser is maintaining an incomplete syntactic dependency, as in the structure used in Experiment 2. At the point of encountering the critical verb in the ambiguous sentences used in Experiment 2 (e.g., after reading "While the professor lectured the students"), the reader may be maintaining an expectation for the main clause subject, which the arrival of a verb rapidly disconfirms. By contrast, at the point of encountering the critical verb in the ambiguous sentences used in Experiment 1 (e.g., after reading "The professor saw the students"), the reader is unlikely to be maintaining any specific syntactic expectation, as the sentence may continue in a wide variety of ways. This explanation is speculative, however, and E-Z Reader 10 has yet to incorporate any explicit account of the factors that affect each of the two integration parameters and that would therefore govern the time course with which integration difficulty appears in the eye movement record.

Experiment 3 was designed to assess the reliability of the critical patterns observed in Experiment 2, in which frequency and integration difficulty had additive effects on reading times on the critical word, and only integration difficulty influenced the rate of regressions from this word. Experiment 3 also attempted to answer the first question raised earlier: Would the underadditive trend observed in Experiment 2 reach significance if syntactic integration is made still more difficult? Unlike in Experiments 1 and 2, in which integration difficulty could ultimately be resolved, in the critical conditions of Experiment 3 the verb was, in fact, an ungrammatical continuation of the sentence. The central prediction arising from E-Z Reader 10 is that the frequency of the verb should have its usual effect on reading times, even in this case. Within the E-Z Reader framework, this situation would presumably be modeled as giving rise to many rapid integration failures (i.e., an inflated p_F parameter). As discussed earlier, however, such failures cannot arise until after L_1 and L_2 have been completed, by which point frequency is done having its effect.

Experiment 3

Method

Participants. Thirty-six students at the University of Massachusetts Amherst participated in the experiment in exchange for course credit. Eight subjects had to be excluded from data analysis due to poor calibration or excessive track loss; these subjects were replaced. All subjects were native speakers of English, had normal or corrected-to-normal vision, and reported no reading or other language-related disorders.

Materials. The same verb pairs used in Experiments 1 and 2 were embedded in item quadruplets like Sentence 7:

7a. The professor saw the students that walked across the quad. (HF, grammatical)

b. The professor saw the students that ambled across the quad. (LF, grammatical)

c. The professor saw the students over walked across the quad. (HF, ungrammatical)

d. The professor saw the students over ambled across the quad. (LF, ungrammatical)

In the grammatical versions (Sentences 7a and 7b), the critical verb always appeared in a subject-extracted relative clause. The ungrammatical versions (Sentences 7c and 7d) were identical except that the relative pronoun *that* or *who* was replaced with a three- or four-letter preposition: *over*, *near*, *for*, or *off*.

This specific syntactic manipulation satisfied two criteria. First, the sentence had to become ungrammatical only at the critical verb itself, and not before; the preposition *over* in Sentences 7c and 7d could be followed by a grammatical continuation (e.g., "The professor saw the students over by the administration building," or "The professor saw the students over winter break"). Second, eye movement patterns prior to the critical verb had to differ as little as possible, so that differences in eye movement patterns on the critical verb between the grammatical and ungrammatical conditions could be attributed to the grammatical manipulation itself, rather than to irrelevant properties of the preceding context.

The 24 item sets were arranged in four lists, so that each subject would read six items in each condition, and each item would be seen in one of its versions by each subject. The 24 experimental items were intermixed with 120 unrelated filler items. The items were presented in an individually randomized order to each subject after eight practice sentences. Two-choice comprehension questions were presented after approximately one fourth of items. Subjects averaged 90% correct on these questions.

Procedure. The procedure was identical to Experiment 1, except that 11-point Monaco font was used (again to satisfy requirements of an unrelated filler experiment).

Results

Due to an error in programming of the experimental script, two of the 24 item sets had to be excluded from all analyses. Track loss resulted in deletion of 7.8% of trials. Less than 2.6% of fixations fell below the 80-ms cutoff, while a total of seven fixations, none of which fell on the critical verb, exceeded the 1,000-ms cutoff.

For the purpose of analysis of the eye movement data, the critical sentences were divided into four regions, as illustrated in Sentence 8:

8. The professor saw the students| that| walked| across the quad."

Unlike in Experiments 1 and 2, the word prior to the critical verb was treated as a distinct region, as this word differed between the grammatical and ungrammatical conditions; this will be referred to as the *precritical* word. In addition, the material following the critical verb was treated as a single region for the purpose of analysis for two reasons. The first was practical: In several of the items, the criteria used to define region boundaries in the previous experiments would have resulted in only a single downstream region. The second was more theoretically motivated, as it is not clear what can be learned from detailed analysis of the eye movement record after the reader encounters an ungrammaticality. Once a genuine ungrammaticality has been registered, it is not at all clear what the eyes "should" do. While the reader may continue to progress through the sentence, he or she may also abandon any attempt to comprehend it in a normal fashion. This may be one reason that there are relatively few eye movement experiments involving responses to outright ungrammaticality (exceptions include Braze, Shankweiler, Ni, & Palumbo, 2002; and Ni, Fodor, Crain, & Shankweiler, 1998) and implausibility (exceptions include Rayner, Warren, Juhasz, & Liversedge, 2004; Warren & McConnell, 2007; and Warren et al., 2008), while such manipulations have played a very central role in the ERP literature (Kutas & Federmeier, 2007).)

The same eye movement measures were computed as in Experiments 1 and 2, and the same statistical models were applied. The means for each measure are shown in Tables 11, 12, and 13, and the statistical analyses are reported in Tables 14 and 15. The reading time means on the critical verb are shown in Figure 7, and the regression data are shown in Figure 8.

The data patterns on the precritical word will be described first. Note that because of the very high rate of skipping of the precritical word (over 40% of trials in all conditions), numerous subjects had no observations in one or more cells of the design, so only by-items ANOVAs could be conducted for this word. The frequency manipulation had no effect on any measure on the precritical word, as expected from the fact that the sentences were identical through this point across the levels of the frequency variable. The grammaticality manipulation did not affect the rate of skipping of the precritical word but had marginal effects on first-fixation duration ($p = .09$) and single-fixation duration ($p = .07$), with longer times when this word was a preposition (and was therefore followed by an ungrammatical verb) than when it was a relative pronoun. The effect on gaze duration did not approach significance ($p = .28$). However, there was a significant effect on go-past time due to a significant effect on the probability of a regression, with more regressions in the ungrammatical conditions, though the regression rate was low in general (.14 in the ungrammatical conditions, .07 in the grammatical conditions). There were no interaction effects on the precritical word on any measure. The interpretation of these data is discussed in detail in the following.

Turning to the critical word, frequency had a significant effect on skipping, which was more likely in the HF conditions. The critical word was also more likely to be skipped in the grammatical conditions than in the ungrammatical ones. On the critical word itself, the effect of frequency was significant in the subjects analysis for both first-fixation duration ($p < .05$) and gaze duration ($p < .01$); the by-subjects analyses could not be conducted for single-fixation duration due to missing data. The frequency effect was not fully significant in the items analyses ($p = .12$ for first-fixation duration, $p = .06$ for single-fixation duration, and $p = .15$ for gaze duration), but note that the power of the items analyses was slightly reduced by the loss of the two excluded items mentioned earlier, as well as by the higher rate of word skipping in this experiment, which resulted in the loss of 16% of trials in the LF ungrammatical condition. First-fixation duration, gaze duration, and go-past time were all significantly affected by the grammaticality manipulation. Frequency did not significantly affect the regression rate from the critical word and did not affect go-past time, but the grammaticality manipulation did affect both of these measures, with more regressions and longer go-past times when the word was ungrammatical. There were no interaction effects, either significant or near significance, on the critical word.

Frequency had no effect on any measure on the spillover region. However, as in the previous experiments, the grammaticality manipulation did affect the regression rate from the spillover region. Unlike in the previous experiments, the grammaticality manipulation also affected first-fixation duration and first-pass time on this material. On the spillover region, there were no interaction effects on reading time that were significant in both the subjects and items analyses, though there was a significant interaction effect on regressions: When the sentence was grammatical, readers were more likely to regress following a LF verb, but when the sentence was ungrammatical, the pattern was reversed. However, as already noted, inferences from reading patterns on material following an outright syntactic violation should be drawn only with caution.

Table 11
Means for Reading Time Measures and Proportions for Regressions and Skipping Measures for Precritical Word (Relative Pronoun or Preposition) in Experiment 3

Construction/frequency	First-fixation	Single-fixation	Gaze	Go-past	p(skipping)	p(regression)
Grammatical						
HF	222	223	237	252	.47	.06
LF	224	224	239	266	.40	.07
Ungrammatical						
HF	239	246	249	323	.43	.16
LF	239	239	252	303	.40	.12

Note. Means are given in milliseconds. HF = high frequency; LH = low frequency; p = proportion.

Table 12
Means For Reading Time Measures and Proportions for Regressions and Skipping Measures, by Condition, on Critical Verb in Experiment 3

	Fixation						
Construction/frequency	Prior	First	Single	Gaze	Go-past	p(skipping)	p(regression)
Grammatical							
HF	210	233	236	266	308	.16	.13
LF	209	243	259	306	379	.08	.17
Ungrammatical							
HF	222	243	246	322	531	.06	.29
LF	232	261	280	340	550	.03	.34

Note. Means are given in milliseconds. HF = high frequency; LH = low frequency; p = proportion.

Discussion

An important initial question is how to interpret the grammaticality effect on regressions from the precritical word. It should be kept in mind that these results are based on limited data and no by-subjects analyses. Assuming that these results are reliable, there are two possibilities. The first is that the preposition itself triggered integrative processing difficulty, which might be explained in a probabilistic parsing framework (e.g., Levy, 2008) by the reasonable supposition that a preposition is a relatively low-probability sentence continuation at this point, compared with a relative pronoun. The second is that the effects that appear on the preposition are actually due to parafoveal processing of the next truly ungrammatical word.

The former conclusion would compromise the interpretation of results on the critical word, as it would suggest that in the ungrammatical conditions, integration difficulty actually began with a word prior to the critical word itself. Fortunately, there is substantial support for the latter conclusion. Several studies have reported that the amount of parafoveal processing of the word to the right of fixation depends on the processing load associated with the currently fixated word (Henderson & Ferreira, 1990; White, Rayner, & Liversedge, 2005); in this experiment, unlike in Experiments 1 and 2, the word prior to the critical word was a short, HF function word, which would induce minimal foveal processing load. In addition, as discussed earlier, there were both frequency and grammaticality effects on the rate of skipping the critical word in this experiment, which did not appear in either of the previous experiments. A range of empirical results (Drieghe et al., 2005) have suggested that a word is skipped when it has received relatively complete parafoveal processing in time for the saccade program that would have targeted the word to be cancelled and replaced with a new saccade to the following word. Thus, there are both logical and empirical reasons for attributing the effects on the precritical word to parafoveal processing of the critical word, rather than to integrative difficulty associated with processing the precritical word itself. Note that Drieghe et al. (2008) have proposed that parafoveal-on-foveal effects (i.e., an effect of the word to the right of fixation that appears in the eye movement record before the eyes have actually left the fixated word) may be likely to arise in cases in which the saccadic targeting system initially targeted the parafoveal word but undershot its target. This would be especially likely when as in the present experiment, the foveal word is one that is very frequently skipped. Drieghe et al. (2008) also demonstrated that such effects are most likely to arise on fixations that are within a few character positions of the parafoveal word. In this experiment, it was necessarily the case that any fixation on the precritical word was close to the critical word because the precritical word was so short. In sum, the present experiment provided the ideal conditions for both maximal parafoveal processing of the critical word and for such processing to be manifested on the critical word itself.

Table 13
Means for Reading Time Measures and Proportions for Regressions by Condition on Spillover Region in Experiment 3

Construction/frequency	First-fixation	First-pass	Go-past	p(regression)
Grammatical				
HF	241	810	1209	.40
LF	241	765	1269	.46
Ungrammatical				
HF	248	675	1660	.59
LF	271	730	1444	.52

Note. Means are given in milliseconds. HF = high frequency; LH = low frequency; p = proportion.

As predicted, both the frequency manipulation and the grammaticality manipulation influenced reading times on the critical word (though the former effect was only marginal in the items analyses, perhaps due to reduced power and a relatively high skipping rate). There was no sign of an interaction between these effects. In addition, the grammaticality manipulation, but not the frequency manipulation, influenced the rate of regressions from the critical word.

The consistent trend toward underadditivity in Experiment 2 across eye movement measures raised the question of whether a more compelling underadditive interaction would emerge if the degree of syntactic integration difficulty was increased still further. But in Experiment 3, the earliest eye movement measures showed a trend toward superadditivity, not underadditivity: The frequency effect was numerically larger in the ungrammatical conditions than in the grammatical conditions in first-fixation duration (18 ms vs. 10 ms) and single-fixation duration (34 ms vs. 23 ms). Thus, the opposing trends in the two experiments suggest that the failure to find a significant interaction in either experiment is not due to a lack of statistical power. In gaze duration, however, an underaddi-

Table 14
Results of by-Subject (F_1) and by-Item (F_2) Analyses of Variance for Reading Time Measures in Experiment 3

Variable	$F_1(1, 35)$	p	$F_2(1, 21)$	p
Precritical word				
First fixation				
Frequency			0.09	.77
Grammaticality			3.14	.09
Frequency × Grammaticality			0.02	.87
Single fixation				
Frequency			0.21	.65
Grammaticality			3.67	.07
Frequency × Grammaticality			0.08	.78
Gaze duration				
Frequency			0.00	.95
Grammaticality			1.22	.28
Frequency × Grammaticality			0.12	.73
Go-past time				
Frequency			0.01	.91
Grammaticality			8.02	<.01
Frequency × Grammaticality			0.52	.48
Critical word				
First fixation				
Frequency	5.15	<.05	2.65	.12
Grammaticality	4.04	.052	5.76	<.05
Frequency × Grammaticality	0.42	.52	0.74	.40
Single fixation				
Frequency			4.01	.058
Grammaticality			3.04	.10
Frequency × Grammaticality			1.28	.27
Gaze duration				
Frequency	8.17	<.01	2.18	.15
Grammaticality	20.16	<.001	13.44	<.01
Frequency × Grammaticality	0.34	.57	0.03	.86
Go-past time				
Frequency	2.20	.15	2.48	.13
Grammaticality	37.34	<.001	33.12	<.001
Frequency × Grammaticality	1.59	.22	0.00	.96
Spillover region				
First fixation				
Frequency	3.67	.06	0.98	.33
Grammaticality	4.05	.052	7.76	<.05
Frequency × Grammaticality	2.53	.12	4.38	<.05
First-pass time				
Frequency	0.07	.80	0.00	.95
Grammaticality	4.14	<.05	0.56	<.05
Frequency × Grammaticality	1.88	.18	2.06	.17
Go-past time				
Frequency	2.17	.15	1.81	.19
Grammaticality	24.97	<.001	14.24	<.01
Frequency × Grammaticality	4.44	<.05	3.22	.09

Note. For single fixation only, F_2 is shown, as some subjects had no single fixations in one or more cells.

tive trend appeared once again, with a smaller frequency effect appearing in the ungrammatical conditions (18 ms vs. 40 ms). Thus, the combined results of Experiments 2 and 3 do raise the possibility that a design with greater power would detect an underadditive interaction in the gaze-duration measure, which reflects not only the durations of individual fixations but also the probability of a refixation on a word. The present results are consistent with the possibility that readers are relatively unlikely to refixate a word when the word is both low in frequency and difficult to integrate syntactically.

General Discussion

Taken together, the results of the three experiments reported here provide striking confirmation of the specific predictions made by E-Z Reader 10 regarding the relationship between the effects of word frequency and syntactic attachment difficulty. First, regardless of syntactic context, the frequency of a verb influenced the early eye movement measures of first-fixation duration, single-fixation duration, and gaze duration; this was the case both in the control conditions in which a verb was expected and in three different experimental conditions in which a verb was unexpected. Second, while syntactic attachment difficulty influenced the probability of a regression from the critical word in Experiments 2 and 3, the frequency manipulation never influenced regression probability. Third, in Experiment 1, strict sequencing of frequency and integration effects was demonstrated, with integration difficulty appearing only in the form of an increased regression probability from the spillover region. Finally, when the effect of integration difficulty did show up in the same reading time measures as the effect of word frequency, these effects did not interact statistically. In Experiment 2, the two manipulations resulted in a trend toward an underadditive interaction, while in Experiment 3, the trend was toward a superadditive interaction, at least in the earliest eye movement measures.

Clearly, of these four basic results, the failure to find a significant interaction supports the weakest inference. Though the conflicting trends in Experiments 2 and 3 in the first-fixation and single-fixation measures suggest that this lack of an interaction is not due to a lack of power, the consistently underadditive trend in

Table 15
Mixed-Effects Logistic Regression Model Estimates of Effects of Frequency and Grammaticality and Their Interaction in Experiment 3

Variable	Estimate (log odds)	*SE*	*z*	*p*
Precritical word				
Skipping				
Frequency	−0.25	0.17	−1.44	.15
Grammaticality	−0.09	0.20	−0.45	.66
Frequency × Grammaticality	0.21	0.33	0.63	.53
Regressions				
Frequency	0.15	0.52	0.29	.77
Grammaticality	1.19	0.44	2.71	<.01
Frequency × Grammaticality	−0.96	0.89	−1.08	.28
Critical word				
Skipping				
Frequency	−0.78	0.33	−2.38	<.05
Grammaticality	−1.01	0.33	−3.07	<.01
Frequency × Grammaticality	0.25	0.66	0.38	.70
Regressions				
Frequency	−0.27	0.28	−0.95	.34
Grammaticality	1.55	0.31	4.97	<.001
Frequency × Grammaticality	0.85	0.60	1.41	.16
Spillover region: regressions				
Regressions				
Frequency	0.05	0.20	0.27	.79
Grammaticality	0.58	0.21	2.69	<.01
Frequency × Grammaticality	−0.77	0.36	−2.17	<.05

Note. Model for the skipping analysis on the critical word did not include random slopes, as this larger model failed to converge.

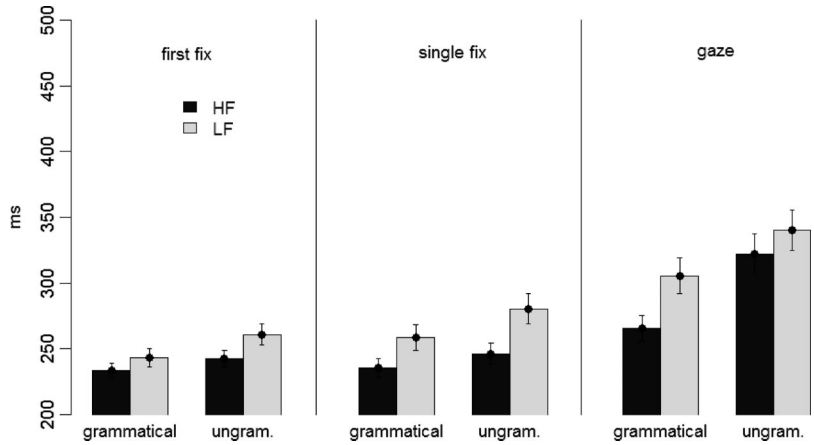

Figure 7. Experiment 3 mean reading times, by condition, on the critical word. Error bars represent standard error of the mean. HF = high frequency; LF = low frequency; umgram = ungrammatical.

the gaze duration measure suggests that a design with more power might reveal an interaction. However, a null effect in a previous study provides additional support for the conclusion that there is not a meaningful interaction between word frequency and syntactic attachment difficulty. Demberg and Keller (2008) investigated the role of various lexical, syntactic, and orthographic factors in predicting word reading times in a corpus of English eye movement data. They found that a significant proportion of the gaze-duration variance was explained by a purely syntactic version of the Surprisal metric, in which the surprise associated with an input word's part of speech—but not with its specific lexical identity—was taken into account. In addition, they found that word frequency was a significant predictor of gaze duration. However, model fit was not significantly improved by including a Syntactic Surprisal × Lexical Frequency interaction term, even though this study had the power to detect even very small effects (e.g., a significant 1.13-ms effect of the character position within a word on which the eyes first landed). Boston, Hale, Kliegl, Patil, and Vasishth (2008) conducted a similar investigation on a German eye movement corpus with similar main effects, but they did not report testing the relevant interaction term.

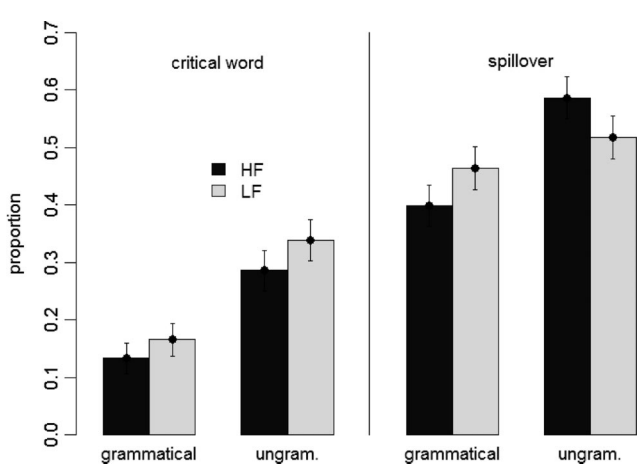

Figure 8. Experiment 3 regression proportions, by condition, on the critical word and spillover region. Error bars represent standard error of the proportion. HF = high frequency; LF = low frequency; umgram = ungrammatical.

In the Introduction, it was claimed that E-Z Reader 10 warrants special interest because it reconciles an attractive theoretical position with apparently conflicting data. The attractive theoretical position is that lexical processing precedes, in part or in its entirety, the process of integration of a word into its sentence context. The conflict comes from effects of contextual manipulations that appear as early as effects of lexical manipulations (e.g., in first-fixation duration). A series of simulations demonstrated a set of testable predictions that result from the specific manner in which E-Z Reader 10 accomplishes this reconciliation. The experiments presented here largely confirm these predictions, and in so doing, they lend support to the architectural features that generated them. In particular, they support the notions that (a) lexical and integrative processing are serially ordered, resulting in either strict sequencing of effects or, when effects of the two manipulations do show up in the same measures, in additivity and (b) they have functionally distinct roles, affecting different aspects of eye movement control, which result in effects on different measures.

The remainder of this discussion takes up three topics. The first is how to reconcile the present data with the self-paced reading results obtained by Tily et al. (2009). The second is whether other data on eye movement are entirely consistent with the claim that in normal reading, lexical processing precedes integration of a word into its context. Finally, broader implications for language processing theories will be discussed.

The Findings of Tily et al. (2009)

Tily et al. (2009) found that a syntactic integration effect was delayed when the word that required a difficult integration was low in frequency, and they interpreted this effect as suggesting that slowed lexical access can delay the point at which integration takes place. Thus, Tily et al. also endorsed the conclusion that lexical processing precedes integrative processing, and indeed, they also raised the question of whether this conclusion can be reconciled with Surprisal (Hale, 2001; Levy, 2008).

There were nonsignificant hints of underadditivity in reading time measures on the critical word in Experiment 2 and in some measures in Experiment 3; to this extent, the present data may be regarded as consistent with Tily et al. But in the present experiments, the point in the eye movement record at which integration difficulty appeared was not influenced by the frequency of the word that required a difficult integration. In Experiment 1, this difficulty showed up in the form of increased regressions from the spillover region, regardless of whether the critical word was high or low frequency. In Experiments 2 and 3, integration difficulty appeared before the eyes left the critical word, even when this word was low in frequency.

One possible explanation for this discrepancy would focus on the substantive differences between the Tily et al. experiment and the present experiments, such as the fact that the present experiments involved syntactic ambiguity or ungrammaticality, while Tily et al. (2009) compared object and subject clefts. But in addition, it is likely that differences between the self-paced reading and eyetracking paradigms played a role. First, the word frequency effect itself may not be fully reliable in self-paced reading, unlike both eyetracking and single word tasks such as lexical decision and naming. For example, Staub, Grant, Clifton, and Rayner (2009) included identical materials in an eyetracking experiment and a self-paced reading experiment and found a highly significant effect of the frequency of a critical word on first-fixation duration (p = .002) and gaze duration ($p = .001$) on this word, but no frequency effect on button-press latency (p = .44). Second, Staub (2010) has shown that the location at which processing difficulty first appears in the eye movement record during the processing of object relative clauses (which are closely related to the object cleft structures employed by Tily et al.) is much earlier than the point at which it appears in self-paced reading (Grodner & Gibson, 2005). Thus, comparing results from the two paradigms is far from straightforward. Clearly, it is important to determine whether the Tily et al. results can be replicated in normal reading or, by contrast, whether an eyetracking replication of the Tily et al. experiment would reveal consistent timing of integration effects across levels of the frequency variable.

Contextual Effects on Eye Movements: Cloze Probability and Ambiguity

Readers familiar with the eye movement literature may have noted that there are other reliable effects of a word's sentence context on reading times. This section addresses the question of whether these effects undermine the claim that there is an autonomous word recognition stage.

Cloze probability. In addition to frequency, a variable that is known to reliably affect fixation durations is cloze probability (e.g., Ehrlich & Rayner, 1981; Rayner, Ashby, Pollatsek, & Reichle, 2004), with a word receiving shorter fixations when it is more predictable in context, as measured by the probability that naïve respondents will provide the word in question as the most likely continuation of the sentence. As noted in the Introduction, E-Z Reader posits that cloze probability, together with frequency, influences the durations of the L_1 and L_2 stages. Thus, in E-Z Reader, cloze probability actually is treated as influencing the duration of lexical processing itself, rather than a postlexical integration stage. Note, however, that in E-Z Reader, the effects of frequency and predictability are constrained to be additive, based on the fact that studies in which these variables were factorially manipulated (Altarriba, Kroll, Sholl, & Rayner, 1996; Ashby, Rayner, & Clifton, 2005; Hand, Miellet, O'Donnell, & Sereno, 2010; Rayner, Ashby, et al., 2004) have failed to show significant interactions between them or even consistent trends. The SWIFT (Saccade-Generation With Inhibition by Foveal Targets) model (Engbert, Nuthmann, Richter, & Kliegl, 2005) also assumes additivity for the same reason. So while E-Z Reader does treat cloze probability as affecting the lexical processing stages of the model, the additivity of frequency and predictability effects is at least consistent with the claim that these two variables do not influence the same stage of processing.

A second way to account for these data would be to concede that predictability effects are, at least in part, effects on lexical access but also to claim that these effects on lexical access arise by way of intralexical priming and thus should not be regarded as threatening the theory that lexical processing is autonomous from sentence context. Morris (1994) has demonstrated that priming does reduce fixation durations when predictability is controlled, though predictability also has an independent effect. For an interesting discussion of the possible role of priming in predictability effects in the ERP domain, see Lau, Phillips, and Poeppel (2008).

Ambiguity. In many eye movement studies, investigators have examined how readers process lexically ambiguous words in context (e.g., Duffy et al., 1988; Rayner & Frazier, 1989; see Morris, 2006, for a review). There are two relevant patterns in this literature. First, a balanced ambiguous word (i.e., a word with two equally frequent meanings) receives longer inspections when both meanings of the word are consistent with the preceding context than when the preceding context strongly favors one of the meanings. Second, a biased ambiguous word (i.e., a word for which one meaning is substantially more frequent than the other) receives longer inspections when the preceding context favors the LF meaning than it would in a neutral context. For an entirely autonomous lexical processing stage to be maintained, these effects should actually arise during integration. For example, Rayner and Frazier (1989) have suggested that long reading times for balanced ambiguous words in neutral contexts arise from the requirement that the processor entertain two possible integrations of the word with its context. Some support for the position that ambiguity effects arise relatively late comes from the fact that in contrast to frequency effects, ambiguity effects have generally not appeared in the first-fixation duration measure. Other support comes from the cross-modal priming studies cited in the Introduction (e.g., Seidenberg et al., 1982), which suggest initial activation of all word meanings, regardless of context.

In sum, it is clear that cloze probability and the relationship between an ambiguous word and its context affect relatively early eye movement measures. It is not clear, however, that either of these findings requires that the assumption of an autonomous word recognition stage be abandoned.

Implications

The present results support the traditional psycholinguistic distinction between visual word recognition, on the one hand, and parsing and sentence interpretation, on the other (cf. Jurafsky,

1996). It has been known for some time that there are impressive correspondences between RTs in single-word recognition tasks such as lexical decision and naming and fixation durations in reading. For example, there are strong correlations across subjects in the size of the frequency effect shown in fixation durations, naming, and lexical decision (Schilling, Rayner, & Chumbley, 1998), and word frequency has similar effects on the shapes of RT distributions in single-word tasks and fixation duration distributions (Staub et al., 2010). The present results show that word frequency has its typical effect even when a word is encountered in a very odd sentence context, namely one in which the word is either quite difficult to attach into the phrase marker (Experiments 1 and 2) or impossible to attach (Experiment 3). Thus, to the extent that the frequency effect is a marker of lexical access processes, a fairly large body of data suggests that when individuals encounter a word in context, the initial process of lexical access is not much (if at all) different from the process of lexical access when they encounter the same word in isolation.

An issue that has not been addressed here is whether it is possible to generalize from the present findings to conclusions about processing in the auditory domain and thus to a modality-independent language-processing architecture. The answer is, clearly, that it is not possible to generalize in this way. Indeed, there is reason to suspect that the process of spoken word recognition may be substantially more sensitive to the context of a word than is the process of visual word recognition. Essentially all auditory word recognition models posit that because a spoken word is spread out in time, a comprehender entertains hypotheses about a word's identity on the basis of partial input. It is possible that relatively early in the process of spoken word recognition, lexical candidates that are consistent with the auditory input are nonetheless inhibited when they are inconsistent with context, so that recognition of such a word must overcome this inhibition. For example, Zwitserlood (1989) found that while a strongly biasing context did not prevent activation of a contextually inappropriate spoken word when this word was consistent with the initial sensory input (cf. Dahan & Tanenhaus, 2004), the context did reduce activation of this word at a slightly later point, still before the point at which the word was ruled out by later sensory input. Thus, if it had turned out that the spoken word was, in fact, the contextually inappropriate one, recognition of this word would presumably have been relatively difficult, due to the need to overcome the earlier contextual inhibition of its activation.

Finally, as has already been noted, qualitative differences in the timing and behavioral consequences of the effects of frequency and syntactic attachment difficulty are not predicted by models in which lexical access and syntactic attachment are collapsed into a single processing operation, such as the Surprisal model. It is important to recognize, however, that these findings are only problematic for the standard version of Surprisal (e.g., Hale, 2001; Levy, 2008) on which there is a single processing difficulty term that is a function of both lexical and syntactic surprise. For example, they are not problematic for a strictly part-of-speech-based version of Surprisal, as tested by both Demberg and Keller (2008) and Boston et al. (2008). On this version of Surprisal, word frequency can be treated as a distinct predictor of eye movement behavior, and indeed both of the aforementioned studies do exactly this. In principle, one could combine such a restricted version of Surprisal with the serial architecture of E-Z Reader 10, perhaps by treating the values of one or both of the two integration parameters for word *n* as a function of the word's part-of-speech-based Surprisal value. There may still be nontrivial difficulties with this enterprise, most notably arising from the fact that Surprisal is a continuous quantity, while E-Z Reader 10 posits that integrative processing has discrete, and only occasional, eye movement consequences. See Staub (2010) for discussion of this "linking" problem.

Conclusion

Three eye movement experiments tested predictions arising from the staged architecture for lexical and integrative processing of the E-Z Reader 10 model. These experiments showed that lexical frequency effects on early eye movement measures were not influenced by the syntactic context of a word. They also showed that while effects of syntactic attachment difficulty were sometimes delayed relative to effects of word frequency, when these two manipulations did influence the same eye movement measures, they did so additively. In addition, only syntactic attachment difficulty influenced the probability of a regressive eye movement from the critical word. These results confirm the predictions of E-Z Reader 10 and provide evidence against the notion of that word recognition and syntactic attachment can be collapsed into a single processing operation.

References

Allen, M., Badecker, W., & Osterhout, L. (2003). Morphological analysis in sentence processing: An ERP study. *Language and Cognitive Processes, 18,* 405–430. doi:10.1080/01690960244000054

Altarriba, J., Kroll, J. F., Sholl, A., & Rayner, K. (1996). The influence of lexical and conceptual constraints on reading mixed-language sentences: Evidence from eye fixations and naming times. *Memory & Cognition, 24,* 477–492. doi:10.3758/BF03200936

Andrews, S., & Heathcote, A. (2001). Distinguishing common and task-specific processes in word identification: A matter of some moment? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 514–544. doi:10.1037/0278-7393.27.2.514

Ashby, J., Rayner, K., & Clifton, C. Jr. (2005). Eye movements of highly skilled and average readers: Differential effects of frequency and predictability. *Quarterly Journal of Experimental Psychology: Section A. Human Experimental Psychology, 58,* 1065–1086.

Bader, M. (1998). Prosodic influences on reading syntactically ambiguous sentences. In J. D. Fodor & F. Ferreira (Eds.), *Reanalysis in sentence processing* (pp. 1–46). Dordrecht, the Netherlands: Kluwer.

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. doi:10.1037/0096-1523.10.3.340

Balota, D. A., & Chumbley, J. I. (1985). The locus of word-frequency effects in the pronunciation task: Access and/or production? *Journal of Memory and Language, 24,* 89–106. doi:10.1016/0749-596X(85)90017-8

Balota, D. A., Cortese, M. J., Hutchison, K. A., Neely, J. H., Nelson, D., Simpson, G. B., & Treiman, R. (2002). The English Lexicon Project: A web-based repository of descriptive and behavioral measures for 40,481 English words and nonwords. Retrieved from http://elexicon.wustl.edu/

Bates, D. M. (2005). Fitting linear mixed models in R: Using the lme4 package. *R News: The Newsletter of the R Project, 5*(1), 27–30.

Bates, D. M., & Sarkar, D. (2007). *lme4: Linear mixed effects models using S4 classes. R package version 0.998875–6.* Madison, WI: Authors.

Boston, M. F., Hale, J., Kliegl, R., Patil, U., & Vasishth, S. (2008). Parsing costs as predictors of reading difficulty: An evaluation using the Potsdam Sentence Corpus. *Journal of Eye Movement Research, 2,* 1–19.

Braze, D., Shankweiler, D., Ni, W., & Palumbo, L. C. (2002). Readers' eye movements distinguish anomalies of form and content. *Journal of Psycholinguistic Research, 31,* 25–44. doi:10.1023/A:1014324220455

Burgess, C., & Livesay, K. (1998). The effect of corpus size in predicting reaction time in a basic word recognition task: Moving on from Kučera and Francis. *Behavior Research Methods, Instruments, & Computers, 30,* 272–277. doi:10.3758/BF03200655

Caplan, D., & Waters, G. S. (1999). Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences, 22,* 77–94. doi:10.1017/S0140525X99001788

Clifton, C., Jr., Staub, A., & Rayner, K. (2007). Eye movements in reading words and sentences. In R. van Gompel (Ed.), *Eye movements: A window on mind and brain* (pp. 341–371). Amsterdam, the Netherlands: Elsevier.

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. doi:10.1037/0033-295X.108.1.204

Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology, 42,* 317–367. doi:10.1006/cogp.2001.0750

Dahan, D., & Tanenhaus, M. K. (2004). Continuous mapping from sound to meaning in spoken-language comprehension: Immediate effects of verb-based thematic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 498–513. doi:10.1037/0278-7393.30.2.498

Demberg, V., & Keller, F. (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition, 109,* 193–210. doi:10.1016/j.cognition.2008.07.008

Drieghe, D., Rayner, K., & Pollatsek, A. (2005). Eye movements and word skipping during reading revisited. *Journal of Experimental Psychology: Human Perception and Performance, 31,* 954–959. doi:10.1037/0096-1523.31.5.954

Drieghe, D., Rayner, K., & Pollatsek, A. (2008). Mislocated fixations can account for parafoveal-on-foveal effects in eye movements during reading. *Quarterly Journal of Experimental Psychology, 61,* 1239–1249. doi:10.1080/17470210701467953

Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language, 27,* 429–446. doi:10.1016/0749-596X(88)90066-6

Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior, 20,* 641–655. doi:10.1016/S0022-5371(81)90220-6

Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review, 112,* 777–813. doi:10.1037/0033-295X.112.4.777

Ferreira, F., & Henderson, J. M. (1990). Use of verb information in syntactic parsing: Evidence from eye movements and word-by-word self-paced reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 555–568. doi:10.1037/0278-7393.16.4.555

Folk, J. R., & Morris, R. K. (2003). Effects of syntactic category assignment on lexical ambiguity resolution in reading: An eye movement analysis. *Memory & Cognition, 31,* 87–99. doi:10.3758/BF03196085

Forster, K. I. (1979). Levels of processing and the structure of the language processor. In W. E. Cooper & E. C. Walker (Eds.), *Sentence processing* (pp. 27–85). Hillsdale, NJ: Erlbaum.

Francis, W. N., & Kučera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar.* Boston, MA: Houghton Mifflin.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences, 6,* 78–84. doi:10.1016/S1364-6613(00)01839-8

Grodner, D. J., & Gibson, E. A. F. (2005). Consequences of the serial nature of linguistic input for sentential complexity. *Cognitive Science: A Multidisciplinary Journal, 29,* 261–290. doi:10.1207/s15516709cog0000_7

Hahne, A., & Friederici, A. D. (2002). Differential task effects on semantic and syntactic processes as revealed by ERPs. *Cognitive Brain Research, 13,* 339–356. doi:10.1016/S0926-6410(01)00127-6

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model *Proceedings of NAACL, 2,* 159–166.

Hand, C. J., Miellet, S., O'Donnell, P. J., & Sereno, S. C. (2010). The frequency-predictability interaction in reading: It depends on where you're coming from. *Journal of Experimental Psychology: Human Perception and Performance, 36,* 1294–1313. doi:10.1037/a0020363

Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. *Psychological Review, 111,* 662–720. doi:10.1037/0033-295X.111.3.662

Henderson, J. M., & Ferreira, F. (1990). Effects of foveal processing difficulty on the perceptual span in reading: Implications for attention and eye movement control. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 417–429. doi:10.1037/0278-7393.16.3.417

Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception & Psychophysics, 40,* 431–439. doi:10.3758/BF03208203

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59,* 434–446. doi:10.1016/j.jml.2007.11.007

Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science: A Multidisciplinary Journal, 20,* 137–194. doi:10.1207/s15516709cog2002_1

Kennedy, A., & Pynte, J. (2005). Parafoveal-on-foveal effects in normal reading. *Vision Research, 45,* 153–168. doi:10.1016/j.visres.2004.07.037

Kliegl, R., Nuthmann, A., & Engbert, R. (2006). Tracking the mind during reading: The influence of past, present, and future words on fixation durations. *Journal of Experimental Psychology: General, 135,* 12–35. doi:10.1037/0096-3445.135.1.12

Kutas, M., & Federmeier, K. D. (2007). Event-related brain potential (ERP) studies of sentence processing. In G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp. 385–406). Oxford, UK: Oxford University Press.

Kutas, M., & Hillyard, S. A. (1980, January 11). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207,* 203–205. doi:10.1126/science.7350657

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (De)constructing the N400. *Nature Reviews Neuroscience, 9,* 920–933. doi:10.1038/nrn2532

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition, 106,* 1126–1177. doi:10.1016/j.cognition.2007.05.006

Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science: A Multidisciplinary Journal, 29,* 375–419. doi:10.1207/s15516709cog0000_25

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition, 25,* 71–102. doi:10.1016/0010-0277(87)90005-9

McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review, 86,* 287–330. doi:10.1037/0033-295X.86.4.287

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18,* 1–86. doi:10.1016/0010-0285(86)90015-0

Morris, R. K. (1994). Lexical and message-level sentence context effects on fixation times in reading. *Journal of Experimental Psychology:*

Learning, Memory, and Cognition, 20, 92–103. doi:10.1037/0278-7393.20.1.92

Morris, R. K. (2006). Lexical processing and sentence context effects. In M. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed., pp. 377–401). London, UK: Elsevier. doi:10.1016/B978-012369374-7/50011-0

Neville, H. J., Nicol, J. L., Barss, A., Forster, K. I., & Garrett, M. (1991). Syntactically based sentence processing classes: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience, 3,* 151–165. doi:10.1162/jocn.1991.3.2.151

Ni, W., Fodor, J. D., Crain, S., & Shankweiler, D. (1998). Anomaly detection: Eye movement patterns. *Journal of Psycholinguistic Research, 27,* 515–539. doi:10.1023/A:1024996828734

Norris, D. (2006). The Bayesian reader: Exploring word recognition as an optimal Bayesian decision process. *Psychological Review, 113,* 327–357. doi:10.1037/0033-295X.113.2.327

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review, 115,* 357–395. doi:10.1037/0033-295X.115.2.357

Pickering, M. J., & Traxler, M. J. (1998). Plausibility and recovery from garden paths: An eye-tracking study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 940–961. doi:10.1037/0278-7393.24.4.940

Pollatsek, A., Reichle, E. D., & Rayner, K. (2006). E-Z Reader: Testing the interface between cognition and eye movement control in reading. *Cognitive Psychology, 52,* 1–56. doi:10.1016/j.cogpsych.2005.06.001

Rayner, K., Ashby, J., Pollatsek, A., & Reichle, E. D. (2004). The effects of frequency and predictability on eye fixations in reading: Implications for the E-Z Reader model. *Journal of Experimental Psychology: Human Perception and Performance, 30,* 720–732. doi:10.1037/0096-1523.30.4.720

Rayner, K., & Duffy, S. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition, 14,* 191–201. doi:10.3758/BF03197692

Rayner, K., & Frazier, L. (1989). Selection mechanisms in reading lexically ambiguous words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 779–790. doi:10.1037/0278-7393.15.5.779

Rayner, K., Warren, T., Juhasz, B. J., & Liversedge, S. P. (2004). The effect of plausibility on eye movements in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 1290–1301. doi:10.1037/0278-7393.30.6.1290

R Development Core Team. (2007). *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria*. Available at http://www.R-project.org

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E-Z Reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences, 26,* 445–476. doi:10.1017/S0140525X03000104

Reichle, E. D., Warren, T., & McConnell, K. (2009). Using E-Z Reader to model the effects of higher-level language processing on eye movements during reading. *Psychonomic Bulletin & Review, 16,* 1–21. doi:10.3758/PBR.16.1.1

Roberts, S., & Sternberg, S. (1993). The meaning of additive reaction-time effects: Tests of three alternatives. In D. E. Meyer and S. Kornblum (Eds.), *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience.* (pp. 611–653). Cambridge, MA: MIT Press.

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 & Cognition, 26,* 1270–1281. doi:10.3758/BF03201199

Seidenberg, M. S., Tanenhaus, M., Leiman, J., & Bienkowski, M. (1982). Automatic access of the meanings of ambiguous words in context: Some limitations of knowledge-based processing. *Cognitive Psychology, 14,* 489–537. doi:10.1016/0010-0285(82)90017-2

Staub, A. (2007). The parser doesn't ignore intransitivity, after all. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 550–569. doi:10.1037/0278-7393.33.3.550

Staub, A. (2009, March). The timing of garden path effects on eye movements: Structural and lexical factors. Paper presented at the 22nd City University of New York (CUNY) Conference on Human Sentence Processing, Davis, CA.

Staub, A. (2010). Eye movements and processing difficulty in object relative clauses. *Cognition, 116,* 71–86. doi:10.1016/j.cognition.2010.04.002

Staub, A., Grant, M., Clifton, C., Jr., & Rayner, K. (2009). Phonological typicality does not influence fixation durations in normal reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 806–814. doi:10.1037/a0015123

Staub, A., & Rayner, K. (2007). Eye movements and on-line comprehension processes. In G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp. 327–342). Oxford, UK: Oxford University Press.

Staub, A., Rayner, K., Pollatsek, A., Hyönä, J., & Majewski, H. (2007). The time course of plausibility effects on eye movements in reading: Evidence from noun-noun compounds. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 1162–1169. doi:10.1037/0278-7393.33.6.1162

Staub, A., White, S. J., Drieghe, D., Hollway, E. C., & Rayner, K. (2010). Distributional effects of word frequency on eye fixation durations. *Journal of Experimental Psychology: Human Perception and Performance, 36,* 1280–1293, doi:10.1037/a0016896

Sturt, P., Pickering, M. J., & Crocker, M. W. (1999). Structural change and reanalysis difficulty in language comprehension. *Journal of Memory and Language, 40,* 136–150. doi:10.1006/jmla.1998.2606

Swinney, D. (1979). Lexical access during sentence comprehension: (Re) consideration of context effects. *Journal of Verbal Learning and Verbal Behavior, 18,* 645–659. doi:10.1016/S0022-5371(79)90355-4

Tanenhaus, M. K., Leiman, J. M., & Seidenberg, M. S. (1979). Evidence for multiple stages in the processing of ambiguous words in syntactic contexts. *Journal of Verbal Learning and Verbal Behavior, 18,* 427–440. doi:10.1016/S0022-5371(79)90237-8

Tanenhaus, M. K., & Lucas, M. M. (1987). Context effects in lexical processing. *Cognition, 25,* 213–234. doi:10.1016/0010-0277(87)90010-2

Tily, H., Federenko, E., & Gibson, E. (2010). The time-course of lexical and structural processes in sentence comprehension. *Quarterly Journal of Experimental Psychology, 63,* 910–927. doi:10.1080/17470210903114866

Townsend, J. T., & Nozawa, G. (1995). Spatio-temporal properties of elementary perception: An investigation of parallel, serial, and coactive theories. *Journal of Mathematical Psychology, 39,* 321–359. doi:10.1006/jmps.1995.1033

Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 528–553. doi:10.1037/0278-7393.19.3.528

Van Petten, C., & Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Memory & Cognition, 18,* 380–393. doi:10.3758/BF03197127

Warren, T., & McConnell, K. (2007). Investigating effects of selectional restriction violations and plausibility violation severity on eye-movements in reading. *Psychonomic Bulletin & Review, 14,* 770–775. doi:10.3758/BF03196835

Warren, T., McConnell, K., & Rayner, K. (2008). Effects of context on eye

movements when reading about possible and impossible events. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 1001–1010. doi:10.1037/0278-7393.34.4.1001

White, S. J., Rayner, K., & Liversedge, S. P. (2005). Eye movements and the modulation of parafoveal processing by foveal processing difficulty: A reexamination. *Psychonomic Bulletin & Review, 12,* 891–896. doi:10.3758/BF03196782

Zwitserlood, P. (1989). The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition, 32,* 25–64. doi:10.1016/0010-0277(89)90013-9

Appendix

Experimental Items Used in Experiment 1

Four versions of each of the 24 items were constructed by varying the verb (the high-frequency verb is shown first) and including or omitting the complementizer *that*:

1. The airport manager heard that the employees hurried/hustled across the open field.
2. The audience saw that the chef poured/sifted the flour onto the counter.
3. The bailiff remembered that the judge noticed/berated the defense attorneys.
4. The boss heard that the manager remembered/suppressed some inconvenient facts.
5. The chairman remembered that the mathematician created/derived a solution to the well-known problem.
6. The crowd remembered that the team defeated/trounced their greatest rivals.
7. The customers saw that the farmer lifted/roused the chickens from their coop.
8. The dancers knew that their footwork ignored/flouted some basic principles.
9. The directions explained that the table top screwed/adhered directly to the base.
10. The economists heard that the speaker predicted/recounted a good year for the industry.
11. The executive announced that the report considered/vindicated Ms. Reynolds from accounting.
12. The farmer knew that the tractor piled/sowed the seeds in long rows.
13. The father claimed that the children bothered/pestered him about the trip to the beach.
14. The general knew that the soldiers climbed/hoisted the big rock that blocked their path.
15. The jockey forgot that the saddle rubbed/chafed the horse's skin.
16. The kids heard that the bus driver wondered/inquired about the location of a hotel.
17. The neighbor observed that the couple purchased/renovated the old Victorian house.
18. The parents noticed that the young girl changed/omitted a critical part of the story.
19. The pilot noticed that the ground crew remained/loitered on the runway for a long time.
20. The professor saw that the students walked/ambled across the quad.
21. The reporter announced that the firemen attacked/drenched the house with high-powered hoses.
22. The speaker believed that the organization proposed/bestowed an annual prize.
23. The teacher suspected that the class skimmed/perused the reading for the week.
24. The zookeepers noticed that the monkeys charged/twisted the bars of their cage.

Experimental Items Used in Experiment 2

Four versions of each of the 24 items were created by varying the verb (the high-frequency verb is shown first) and including or omitting the comma between clauses.

1. While the professor lectured, the students walked/ambled across the quad.
2. As the crowd cheered, the team defeated/trounced their greatest rivals.

(*Appendix continues*)

3. When the jockey jumped on, the saddle rubbed/chafed the horse's skin.

4. As soon as the father woke up, the children bothered/pestered him about the trip to the beach.

5. While the neighbor observed, the couple purchased/renovated the old Victorian house.

6. While the audience watched, the chef poured/sifted the flour onto the counter.

7. As the farmer drove, the tractor piled/sowed the seeds in long rows.

8. When the chairman asked, the mathematician created/derived a solution to the well-known problem.

9. Before the weather interrupted, the soldiers climbed/hoisted the big rock that blocked their path.

10. After the speakers finished, the organization proposed/bestowed an annual prize.

11. Because the pilot delayed, the ground crew remained/loitered on the runway for a long time.

12. After the witnesses left, the judge noticed/berated the defense attorneys.

13. Though the dancers practiced, their footwork ignored/flouted some basic principles.

14. Before the teacher started, the class skimmed/perused the reading for the week.

15. When the customers approached, the farmer lifted/roused the chickens from their coop.

16. As the airport shut down, the employees hurried/hustled across the open field.

17. Because her parents were watching, the young girl changed/omitted a critical part of the story.

18. Because the boss was visiting, the manager remembered/suppressed some inconvenient facts.

19. Though the rumors had begun, the report considered/vindicated Ms. Reynolds from accounting.

20. Though the economists already knew, the speaker predicted/recounted a good year for the industry.

21. Though the directions did not explain, the table top screwed/adhered directly to the base.

22. Though the kids didn't hear, the bus driver wondered/inquired about the location of a hotel.

23. As the zookeepers stood by, the monkeys charged/twisted the bars of their cage.

24. When the trucks stopped, the firemen attacked/drenched the house with high-powered hoses.

Experimental Items Used in Experiment 3

Four versions of each of the 22 items (as noted in the text, two items were deleted due to a programming error) were created by varying the verb (the high-frequency verb is shown first) and by substituting a preposition for the relative pronoun.

1. The professor saw the students that/over walked/ambled across the quad.

2. We celebrated with the team that/over defeated/trounced our rival.

3. We rented the bowling shoes that/over rubbed/chafed our ankles.

4. While they were at the park, they saw the children that/over bothered/pestered them yesterday.

5. At the supermarket, we ran into the people that/over purchased/renovated our house last year.

6. While in the recycling center, we looked at the machines that/over poured/sifted through our garbage.

7. We watched the workers that/over piled/sowed the seeds onto the fields.

8. The TV show with the mountaineers who/for climbed/hoisted the rocks was a huge success.

9. The professor gave extra credit to the students that/near created/derived the best formulas.

10. The principal praised the teacher that/near noticed/berated the foreign exchange student.

11. The parents scolded the children who/for ignored/flouted the house rules.

12. A lot of people did better than the ones who/off skimmed/perused the material.

13. Many people stared at the farmers who/for lifted/roused the animals from their cages.

14. Many harsh glances were directed at the people who/off hurried/hustled through the crowd rudely.

(*Appendix continues*)

15. The jury was not favorable toward the defendant who/for changed/omitted part of his story.

16. The psychiatrist worked with a patient who/off remembered/suppressed most of her childhood family vacations.

17. He learned a lot from the lawyer who/for considered/vindicated the politician.

18. Due to its style, nobody took the news story that/off predicted/recounted the disaster seriously.

19. Everybody marveled at the boy who/off screwed/adhered all the pieces together in under a minute.

20. The instructor taught an optional class for the students who/off wondered/inquired about the details of the experiment.

21. It was of concern to the tourists that the monkey who/for charged/twisted the bars seemed rather angry.

22. The parents sat and talked about the children who/off attacked/drenched each other with Super Soakers.

Received July 26, 2010
Revision received February 11, 2011
Accepted February 14, 2011 ■