

Running head: IN SEARCH OF ON-LINE LOCALITY EFFECTS

In Search of On-line Locality Effects  
in Sentence Comprehension

Brian Bartek

University of Michigan, Ann Arbor

Richard L. Lewis

University of Michigan, Ann Arbor

Shravan Vasishth

University of Potsdam, Potsdam

Mason R. Smith

University of Michigan, Ann Arbor

Brian Bartek

Department of Psychology

University of Michigan

Ann Arbor, Michigan, USA 48109

E-mail: bartek@umich.edu

### Abstract

Many comprehension theories assert that increasing the distance between elements (e.g., a verb and an NP argument) participating in a linguistic relation increases the difficulty of establishing that relation during on-line comprehension. Such *locality effects* are expected to inflate reading times, and are thought to reveal properties and limitations of the short-term memory system that supports comprehension. Despite their theoretical importance and putative ubiquity, however, existing evidence for on-line locality effects is quite narrow linguistically and methodologically: it is restricted almost exclusively to self-paced reading (SPR) of complex structures involving a particular class of movement relations in which the observed effects may arise from other factors. We present four experiments (two self-paced reading and two-eyetracking experiments) demonstrating locality effects in establishing subject-verb dependencies in simple materials that are read quickly and easily. These locality effects are observable in the earliest possible eye-movement measures. The combined results of all four experiments also support a specific hypothesis concerning the source of long reading times in prior SPR measures of locality effects: they are primarily the result of recovery from parsing *failures* (perhaps triggered by short-term memory retrieval failures), rather than measures of successful initial parsing operations that depend on locality-modulated memory processes.

## In Search of On-line Locality Effects in Sentence Comprehension

One important goal of psycholinguistic research is to understand the memory processes that support the rapid comprehension of linguistic input, with its many temporally nonlocal relations. Both spoken and written comprehension require the comprehender to incrementally bring new input into contact with partial representations created on the basis of input that occurred earlier.

This functional requirement for memory in the short term is easily seen in the nature of intrasentential linguistic relations such as those in (1) below, in which representations initially created upon reading or hearing *the manager* must be accessed at *quit* in order to establish the relationship between subject and verb:

- (1) a. The manager unexpectedly quit her job yesterday.  
 b. The manager who the supervisor admired unexpectedly quit her job yesterday.

The nature of the constraints and capacities of these memory processes has long been the topic of empirical and theoretical work in sentence processing, and current theories continue to advance a number of hypotheses about specific properties of this system, such as decay and similarity-based interference (Lewis & Vasishth, 2005; Gibson, 1998; Just & Carpenter, 1992; Gordon, Hendrick, Johnson, & Lee, 2006; Lewis, Vasishth, & Van Dyke, 2006).

One of the most straightforward and theoretically influential empirical generalizations to emerge from this work is that the *locality* of linguistic relations, such as the subject-verb relation in (1) above, is a primary determinant of the speed and efficacy of the short-term memory processes in parsing (Chomsky, 1965; Just & Carpenter, 1992; Gibson, 1998). More specifically, increasing the distance over which these relations must

be computed degrades the underlying memory processes in some way. For example, the implication of this view for (1) is that the subject-verb relation in (1b) is more difficult to compute than the same relation in (1a).

This theoretical view has been expressed most transparently in *Dependency Locality Theory* (DLT) (Gibson, 1998, 2000), which uses as a measure of locality the number of new linguistic referents interposed between a dependent and its head. DLT claims that the degree of locality should be reflected in a continuous and monotonic way in on-line reading time measures, thus yielding testable empirical predictions. In this paper we refer to the general class of such effects on reading times as *locality effects*, without intending to associate them exclusively with the details of DLT or other specific parsing models or metrics. Locality effects are important and relevant to a very broad range of extant memory and parsing theories (see Lewis et al. (2006) for a summary)—even those which do not have mechanisms in place to directly produce them.

Our aims for this paper are threefold. First, we briefly advance and defend the claim that current empirical evidence for on-line locality effects is narrow both linguistically and methodologically, and perhaps surprisingly difficult to find under the assumption that locality is a ubiquitous factor in short-term memory processes. More specifically, we raise the possibility that locality effects may be evident only in relatively complex structures whose difficulty may be traceable to independent factors. If this is the case, it has major implications for how these phenomena bear on theory development.

Given the key role that locality effects play in shaping current parsing theory, we believe that it is important to significantly broaden its base of empirical support, and this relates to our second and third aims. Our second aim is to extend locality investigations to include eye tracking measures, which we will show has a number of specific advantages over self-paced reading (SPR) for investigating locality effects. Furthermore, we adopt an approach of running identical experiments in both paradigms. This structure will

facilitate future efforts to develop detailed theories of the link between the underlying short-term memory processes and the control of eye-movements and button-presses (and therefore the relationship between SPR and eye tracking as empirical measures).

Our third aim is to demonstrate (possibly more subtle) locality effects using linguistic material that is, overall, significantly easier to process than materials that form the basis of existing locality demonstrations, thus providing stronger evidence for the claim that locality exerts pervasive and continuous effects on sentence processing. To foreshadow some of the key results, our new experiments succeed in this demonstration, and the eye tracking evidence provides new insight into separable processes driving the self-paced reading effects.

The remainder of this paper is structured as follows. We first provide our assessment of the current evidence for locality effects, and discuss its potential theoretical implications. We then describe the design and results from four new experiments, which consist of two pairs of SPR and eye tracking experiments. Finally, we discuss the theoretical and methodological implications of the results in the General Discussion.

### **Assessing current empirical evidence for locality effects**

The existing empirical evidence for locality effects is surprisingly mixed. Locality effects have been found in studies of English sentences, but *antilocality* effects—faster processing in longer-distance dependency integration—have been found in head-final languages including German, Hindi and Japanese (e.g., Konieczny, 2000; Vasishth & Lewis, 2006; Vasishth, 2002; Nakatani & Gibson, 2004), as well as English (Levy, 2008, reporting an unpublished experiment by Jaeger and colleagues). Although antilocality effects place important constraints on psycholinguistic parsing theory—and it is important to assess theories of locality effects in their context—it remains possible that independent factors give rise to both kinds of effects, and they need not be mutually incompatible. Our

concern in this paper is to develop a better understanding of the nature and extent of *positive* locality effects. Through this effort, we will better understand how the parser's efficacy depends upon the salience of memory representations. In other work we have outlined a theoretical model that provides an integrated explanation of both locality and antilocality (Vasishth & Lewis, 2006; Lewis et al., 2006; Lewis & Vasishth, 2005).

Table 1 provides an overview of the existing experimental evidence for locality effects (which is mostly restricted to English—a cross-linguistic gap that we do not fill in this paper). We now consider the theoretical implications of the present narrow range of structure and methodology.

*Existing on-line locality effects are restricted to points of extraction*

Locality effects have been observed in both ambiguous and relatively unambiguous structures. In ambiguous structures, locality plays a role in both resolving ambiguities (Kimball, 1973; Frazier & Fodor, 1978; Grodner, Gibson, & Tunstall, 2002; Gibson, Pearlmutter, Canseco-Gonzales, & Hickock, 1996; Pearlmutter & Gibson, 2001; Gibson, Pearlmutter, & Torrens, n.d.; Altmann, Nice, Garnham, & Henstra, 1998) and in affecting the difficulty of garden path reanalysis (Pritchett, 1992; Gibson, 1991; Van Dyke & Lewis, 2003). While these results have yielded useful constraints on parsing theory (Lewis & Vasishth, 2005), our present aim is to understand and find evidence for on-line locality effects in globally unambiguous structures.

In the studies summarized in Table 1, evidence for on-line locality effects in *unambiguous* structures is restricted to English, and to points of extraction—more specifically, to relations conventionally analyzed as  $\bar{A}$ -movement (of an argument) from an argument position to a non-argument position (Mahajan, 1990). It has been posited in Grodner and Gibson (2005)(p. 284) and elsewhere (?, ?) that  $\bar{A}$ -movement may be an important condition for the occurrence of locality effects.

Given this restricted evidential base, there are two plausible accounts for the locality effects that have been obtained experimentally. Locality effects may be a direct result of the degradation of memory representations between initial activation and subsequent retrieval for integration into a dependency, which would imply ubiquity of the effects. Alternatively, locality effects could reflect a source of difficulty unique to movement operations. Although most theories of working memory in sentence processing do not distinguish the computational demands of movement and non-movement relations, there is a line of work that does make such a distinction, starting with the *Hold Hypothesis* in the ATN model of Wanner and Maratsos (1978), and continuing in the Grodzinsky (2000) theory of function in Broca's area. But as we explain in the next section, prior experiments that could have determined if locality effects generalize beyond movement have yielded ambiguous results.

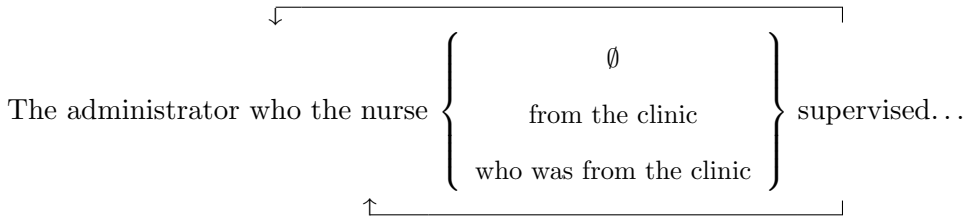
The nature of the existing evidence can be understood by considering three of the experimental conditions in Grodner and Gibson (2005) Experiment 2 (underlining is used here to indicate the word at which the locality effects are predicted to be observed). Note that, in these sentences,  $\bar{A}$  movement occurs when the object is moved from its base position (by the embedded verb) to the beginning of the sentence.

(2) *Embedded verb conditions from Grodner and Gibson (2005) Experiment 2*

- a. The administrator who the nurse supervised scolded the medic while ...
- b. The administrator who the nurse from the clinic supervised scolded the medic while...
- c. The administrator who the nurse who was from the clinic supervised scolded the medic while...

In all three structures in (2a), the region of interest is the embedded verb *supervised*, and the locality manipulation involves increasing the distance from the embedded verb to its subject (*the nurse*) and its extracted object (*the administrator*). In (2a), no material intervenes between the embedded verb and the subject; in (2b), a three word prepositional phrase (PP) intervenes; and in (2c), a five word relative clause (RC) intervenes. The structure of this design is shown schematically in (3). The top arrow denotes the relation between the verb and the relative pronoun *who* that mediates the object extraction, and the bottom arrow denotes the subject relation. The  $\emptyset$  symbol denotes the null string (nothing interposed).

(3) *Structure of the embedded verb conditions from Grodner and Gibson (2005)*



The assumption (as expressed, e.g., in DLT) is that the computation of these dependency relations happens immediately at *supervised* by accessing short-term memory representations associated with the relativizing pronoun and the subject<sup>1</sup>, and that this computation takes longer as the input items that triggered the target representations become more distant. Thus, the straightforward prediction is that reading times at *supervised* should increase monotonically in the three conditions (nothing interposed, PP interposed, and RC interposed). This prediction is consistent with what Grodner and Gibson (2005) found in their Experiment 2 using self-paced reading, with the sharpest increase in reading times observed for the RC condition (we discuss the empirical results in more detail below).

This manipulation has several attractive features that lead us to adopt it for the new experiments reported in this paper—in particular, the specific verbs in the critical



region and the head nouns of the target subject and object noun phrases are kept constant while changing the locality of the relations.

*Alternative explanations of existing on-line locality effects*

There are several possible reasons for the increased reading times at *supervised* in (3) above:

- Completing the subject-verb dependency (*nurse* and *supervised*) may be more difficult due to increased distance between the subject and verb.
- Completing the subject-verb dependency (*nurse* and *supervised*) may be more difficult due to increased similarity-based interference between the subject and object. Both the subject and the direct object of the critical verb were noun phrases referring to humans. This could cause interference effects in addition to any locality effects due to embedding, because both noun phrases overlap in features that by hypothesis are used as retrieval cues by the critical verb to establish the subject relation (such as *+animate* or *+human*).
- Completing the dependency between the extracted object and the verb may be difficult because of similarity-based interference from the additional intervening embedded relative clause. In particular, this could be due to (retroactive) retrieval interference contributed by the second relative pronoun (Lewis & Vasishth, 2005).
- Completing the dependency between the extracted object and the verb may be difficult for other reasons related to the object's movement from its argument position (after the critical verb) to a non-argument position in the relative
  - There could be difficulty arising from computing two non-local relations *simultaneously* (see the two relations depicted in (3)).
  - Verb reading times may increase when the surrounding syntactic context is rare.

The Production-Distribution-Comprehension (PDC) account, in particular, predicts

slower comprehension of the embedded verb in an object-relative sentence because encountering the verb in this context amounts to a violation of expectation. In this case, the animacy of the extracted object makes the subsequent embedded object-relative even less probable, leading to even longer reading times that would resemble locality effects.<sup>2</sup>

- Any of the previously listed sources of difficulty may interact with each other, or with another factor that contributes to the complexity of  $\bar{A}$  movement.

- There could be differential amounts of spillover on the verb from preceding regions in the three conditions (Mitchell, 1984; R. Rayner & Duffy, 1986).

How do we empirically distinguish among these possibilities? The concern about interference from the embedded relative can be avoided by focusing on the contrast between the PP condition and the no-interposed-material condition; we discuss this further below. In order to distinguish between the locality effects due to the subject and extracted object relations, we can compare the effects in (3) above to three other conditions in Grodner and Gibson (2005):

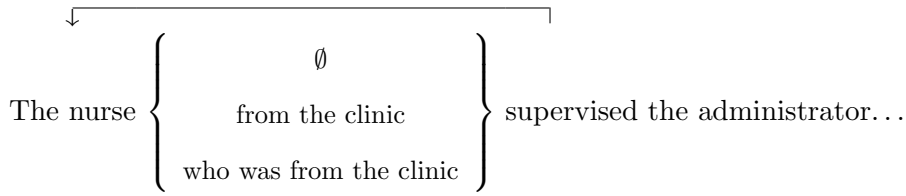
(4) *Matrix verb conditions from Grodner and Gibson (2005) Experiment 2*

- a. The nurse supervised the administrator while ...
- b. The nurse from the clinic supervised the administrator while ...
- c. The nurse who was from the clinic supervised the administrator while ...

These three conditions test for locality effects at a matrix verb from which no arguments have been extracted; the only linguistic relation affected by locality is the subject relation.

The structure of the main verb conditions is shown schematically in (5):

(5) *Structure of the matrix verb conditions from Grodner and Gibson (2005) Experiment 2*



If a locality effect is observed at *supervised* in (5), this would provide evidence that dependencies that are not the result of  $\bar{A}$ -movement relations are also subject to locality effects.<sup>3</sup> The presence of locality effects in both kinds of structures would mean that it is computationally costly to resolve both simple subject-verb dependencies and extracted object-verb dependencies.

Locality effects in (5) would also pose a challenge to experience-based accounts like PDC. All three matrix conditions have an animate referent in a typical subject position, are frequently produced by speakers, and should therefore (under PDC) also be easily comprehended (Gennari & MacDonald, 2009). PDC does not make a clear prediction that locality effects should occur in the matrix conditions.

Figure 1 (upper left) shows the readings times observed by Grodner and Gibson (2005) at the critical verb. (This figure also contains the reading times for the four experiments in this paper, but the reader should focus for now on the upper-left graph). We can now ask whether these extant results help to distinguish among the possibilities outlined above.

Unfortunately, they do not. The locality contrasts within the matrix verb condition were not reported in Grodner and Gibson (2005), but do not appear to be reliable. The contrast between the PP and no-interposition conditions in the embedded structures also was not reported, and also appears not to be reliable. Finally, spillover was not taken into account in the Grodner and Gibson (2005) study.

In short, it is quite possible that the locality effects are driven entirely by independent sources of difficulty resulting from embedding the verb and from

center-embedding the relative clauses.

Despite the ambiguity attending the Grodner and Gibson (2005) results, we believe that the structure of their Experiment 2 is still a promising way, in principle, to tease apart these issues, and we adopt its structure for the four experiments presented here. But before moving on to the new experiments, we consider briefly the implications of the narrow methodological base for investigating locality effects.

*A concern about the existing self-paced reading evidence for locality*

Self-paced reading has the virtue of yielding a simple measure that is often sensitive to the fluctuating processing demands of incremental comprehension. But because each word (or phrase) disappears as soon as the reader presses a button, the stakes of each button press are high relative to moving the eyes forward in reading. If the reader encounters difficulty that would best be resolved by regressing to an earlier part of the sentence, for instance to find a particular argument, he or she has no recourse in self-paced reading but to try to remember or mentally rehearse what came before. Eye-movements could potentially leave an interpretable record of such recovery processes, but SPR cannot—except perhaps in significantly inflated reading times.

This difference between SPR and eye-tracking turns out to be crucial for interpreting SPR reading time data such as that in Grodner and Gibson (2005). In order to appreciate the potentially problematic nature of the existing data, and the potential gains from pursuing eye-tracking studies, it is important to understand what problems SPR raises for evaluating locality effects.

As we pointed out above, the existing evidence for on-line locality effects in reading comes from self-paced reading of extraction structures such as (2), which are relatively complex and cause noticeable processing difficulty. The locality results observed by Grodner and Gibson (2005) are marked by an increase in reading times for the most

difficult condition (the doubly embedded relative clause, (2c)). It is therefore possible that these effects reflect recovery from failed argument-verb integration, triggered by factors related to the embedding of relative clauses (e.g., increased interference), rather than locality. In other words, the observed 125–150ms increase in reading time may not be due to longer integration or memory retrieval processes affected by locality, but primarily recovery processes—perhaps covert rehearsal—triggered by interference-induced retrieval or integration *failures*. To anticipate one of the key findings reported in this paper: the combined results of our experiments provide strong support for this interpretation of existing SPR locality effects.

### Overview of the empirical strategy and four experiments

We now provide a brief overview of our empirical strategy and describe how it is realized in the four new experiments that follow. The overall goal is to determine if it is possible to observe locality effects that are not subject to the critiques above. Ideally, this means observing locality effects at points of computing relations that do not involve  $\bar{A}$  movement, and observing locality effects under conditions of relatively easy processing, using methods that make it possible to more clearly separate locality-modulated processes of integration or retrieval from processes of recovery from initial parsing failures.

We pursue four empirical devices to achieve these goals:

1. We adopt the six-condition structure of Grodner and Gibson (2005), outlined above in (2) and (4), which in principal has the potential to reveal locality effects in the main clause conditions at points that do not involve extraction.
2. We run eye-tracking as well as SPR versions of each experiment. The specific aims are to (a) provide potentially more sensitive measures of locality effects in easy, non-extraction structures; (b) distinguish between locality effects on early first-pass and late recovery processes in the eye-movement record; and (c) provide a better understanding

of the nature of locality effects observed in SPR by providing evidence bearing on the specific hypothesis above concerning the role of parsing failure and recovery in SPR.

4. We adopt a new set of stimuli based on these structures but with content words drawn from a list of relatively short (three to six letter), high frequency words. The specific aims are to (a) increase the overall ease of processing and therefore provide an additional test of the hypothesis that locality effects might only be evident in the presence of other sources of processing difficulty; (b) decrease item-dependent variance related to the length and frequency of content words; and (c) increase the proportion of single fixations in the eye-movement record which might provide the best opportunity to reveal early locality effects.

4. In the new set of stimuli, we also attempt to rule out alternative explanations for locality effects that rely on the animacy of the object (*administrator*) in the critical dependency. In the new materials, this role is always filled by an inanimate noun (e.g., *sports*). This change reduces the predicted impact of structural rarity on comprehension in the embedded-verb conditions (containing an object-relative clause). Using inanimate objects will also reduce the effect of similarity-based retrieval interference at the critical verb. As described above, in the original Grodner and Gibson (2005) materials, both the subject and extracted object were noun phrases referring to humans. Thus in addition to increasing locality, the embedding manipulation also potentially increases similarity-based interference. In the new materials, we expect to see a smaller effect of embedding that is independent of similarity-based interference.

The four experiments thus cross materials (original Grodner & Gibson stimuli and new stimuli) with method (SPR and eyetracking). Experiment 1 is SPR with the original Grodner & Gibson materials (a replication of their Experiment 2), Experiment 2 is eyetracking with the original materials, Experiment 3 is SPR with the new materials, and Experiment 4 is eye-tracking with the new materials. For simplicity of presentation and

analysis, we perform complete analyses on each experiment separately, but report a small number of key comparisons that test materials effects directly between Experiments 1 and 2, and 3 and 4.

### Experiment 1: Replication of Grodner & Gibson (2005) Exp.

#### 2

##### *Method*

A self-paced reading replication of Grodner and Gibson's (2005) Experiment 2 was run.

##### *Participants*

Forty-nine University of Michigan undergraduates participated for payment or for partial course credit. All participants were native English speakers with normal or corrected-normal vision, and were naive to the purpose of the experiment.

##### *Stimuli*

Participants in experiment 1 read thirty experimental sentences taken from Grodner and Gibson (2005) Experiment 2. Six versions of each item were used, as originally shown in (2) and (4), and repeated in Table 2 with condition labels.

For every item, the *matrix/unmodified* condition was a declarative sentence containing a transitive verb with human NP arguments. In the *matrix/PP-modified* condition the subject was modified with a prepositional phrase. In the *matrix/RC-modified* condition, a subject-modifying relative clause was made by placing the words *who was* were placed at the beginning of the PP. In these three conditions, the object never undergoes movement.

The remaining three conditions were created by applying the same series of modifications (unmodified, PP-modified, RC-modified) to an adaptation of the core

sentence. In all three conditions the object NP became the subject of the matrix clause (through  $\bar{A}$ -movement), and the rest of the sentence became an RC modifying that subject. A clausal connective always followed the matrix object.

Thirty experimental sentences were created and assigned to lists with a Latin square design. Forty-eight fillers and sixty-four sentences from unrelated experiments completed each list. Experimental trials never appeared consecutively, and no verbs or arguments were re-used.

We report reading times at the first verb (e.g., *supervised*), which always occupied the same underlined position as in the examples in Table 2. This was where the dependency initiated by the first argument (*nurse* in the first three conditions or *administrator* in the last three conditions) was resolved. In the first three conditions this verb was in the matrix clause, so we called these the *matrix verb conditions*. In the last three conditions, the same verb was in an embedded clause, so we called these the *embedded verb conditions*. In all conditions, the verb integrated with the same arguments across the sentence.

Crucially, the critical verb also never occurred at a point of disambiguation. It is possible that, in some semantic contexts, a reader could parse the critical verb (*supervised*) as the beginning of a reduced relative clause rather than as the verb taking *nurse* as its subject; but this should not affect the usefulness of the present manipulation because (a) Disambiguation would occur downstream of the region where locality effects would be observed. (b) It is unlikely that a reduced relative reading of the critical verb would be more preferred following a subject-modifying relative clause than after a subject-modifying PP or a bare subject — in part because, as experience-based accounts suggest (Gennari & MacDonald, 2009), rarely produced structures are less expected and take longer to parse.



*Procedure*

Participants were seated with their eyes approximately twenty inches in front of a 17-inch Apple LCD display. After reading instructions, they read twenty practice sentences in the moving-window SPR paradigm, each followed by a comprehension question. Participants then began experimental trials.

In the moving-window paradigm, a series of dashes appeared wherever a word would appear for the current sentence. Participants pressed the spacebar to reveal the first word. Subsequent space bar presses revealed the next word while replacing the prior word with dashes. Some sentences were long enough to require a second line of text, but in all cases the line break occurred after the critical verb.

Pressing the space bar after the final word of a sentence removed the sentence from the screen and displayed a comprehension question. Participants responded *yes* to the question by pressing *f* on the keyboard or *no* by pressing *j*. If they answered correctly, “correct!” was displayed briefly; “incorrect” was briefly displayed if they answered incorrectly. Each press of the space bar during sentence presentation was used as an RT measure approximately reflecting processing time for the text that had just been displayed.

*Statistical techniques used in the analysis*

Data analysis was carried out using linear mixed models (LMMs) (Bates & Sarkar, 2007) available as the package `lme4` in the R programming environment (R Development Core Team, 2006). In the analyses, participants and items were treated as random intercepts (sometimes referred to as random effects) and the contrasts (discussed below) as the fixed factors (or fixed effects). The effect of each contrast was derived by computing 95% highest posterior density (HPD) intervals for the coefficient estimates. Compared to conventionally used confidence intervals, the HPD interval is easier to interpret since it demarcates a range within which the population coefficient is expected to lie; this is how

the 95% CI is usually (incorrectly) interpreted. For details on how the HPD intervals are computed, see Gelman and Hill (2007).

Following Grodner and Gibson (2005), analyses included all reading times within three standard deviations of the condition-mean reading time. (Less than 1% of the data were affected by this procedure.) Reaction time data from the critical verb in every experiment were log-transformed to correct for the typical positively skewed distributions observed with reaction times. Log-transforming reaction times yields an approximately normal distribution of data points, which allows for more reliable and precise statistical models of the data.

Two sets of five orthogonal contrasts across the six conditions were run in separate iterations of a linear mixed model that included both subject and item as crossed random factors. The key theoretical contrasts of interest in these sets are specified in Table 3. Contrasts were normalized to make the contrast coefficients in our models directly interpretable as estimated mean differences between the two groups represented by the contrast<sup>4</sup>. We refer to the difference between the means of the three matrix conditions and the three embedded conditions as the *embedding* effect (the first contrast in Table 3). We refer to the difference between the local (no modification) condition and the mean of the non-local conditions (the PP and RC modifications) as the *locality* effect, and specify two such effects, one for the matrix conditions (the second contrast in Table 3) and one for the embedded conditions (the fourth contrast in Table 3). The difference between these two locality effects is the *locality by embedding* interaction (the sixth contrast specified in Table 3). Similarly, we specify contrasts testing the difference between the two kinds of non-locality (PP and RC modification), separately for the matrix and embedded conditions (the third and fifth contrasts in Table 3). The difference between these two modification contrasts is the *modification type by embedding* interaction (the last contrast in Table 3).

*Spillover*

Although the critical verb was identical across conditions, the immediately preceding region was different in the unmodified (local) vs. modified (non-local) conditions, so spillover is a possible contributing factor to estimates of the two locality contrasts. We adapted the statistical control for spillover used by Vasishth and Lewis (2006) as follows. In the analysis of data from self-paced reading experiments (1 and 3), reading time from the prior region, as well as the length and frequency of the word in the prior region, were included in the model.

*Results*

Two items were removed because they were improperly designed. One item was ungrammatical because it was missing the matrix verb in the object-extracted sentences. The other item contained an intransitive verb in the critical position. This left twenty-eight experimental items. The design errors that affected the excluded items here were unique to this replication, and did not, to our knowledge, affect the Grodner and Gibson original study.

*Question accuracy*

Participants answered 74% of all trials correctly. Participants who answered fewer than 70% of the comprehension questions correctly were removed from analysis. Ten participants were excluded by this procedure, leaving thirty-nine participants' data to be analyzed.

*Word length and frequency*

The critical verb region does not vary from condition to condition, but we can potentially obtain tighter estimates of the contrast coefficients by explicitly modeling the effect of word length and frequency. The results reported for this experiment, and for

Experiments 2–4, are from mixed effects models that include length and frequency as covariates.

*Overview of the results figures*

Before describing the results of Experiment 1, we provide here an overview of Figures 1, 2 and 3, which systematically depict the results of all the experiments in this paper (as well as Grodner and Gibson (2005) Experiment 2).<sup>5</sup>

Reading times in milliseconds at the critical verb are presented in Figure 1. Each separate panel in this figure depicts the reading times (and standard errors) across the six conditions. The three panels in the top row display SPR results (Experiments 1 and 3 and Grodner and Gibson (2005) Experiment 2) alongside the Total Fixation Times from the eyetracking experiments (Experiments 2 and 4). Data obtained from the original Grodner and Gibson (2005) materials (Experiments 1 and 2) are depicted with black lines; data obtained from the new materials (Experiments 3 and 4) are depicted with grey lines. As we describe in more detail below, the second row of panels in Figure 1 depicts the early eye-tracking measures, and the last row depicts the late measures. The scale on the y-axis is always consistent across a row in the figure, but note that the early eye-tracking measures are plotted on a different scale.

Rather than report the details of the statistical analyses in-line in the text, we summarize the results of the tests graphically by plotting the mixed effect models' point estimates of the contrasts as well as the surrounding 95% HPD intervals. The *locality* and *modification* contrasts within the matrix and embedded conditions (described above) are plotted in Figure 2. The *embedding* effect and its two associated interactions are plotted separately in Figure 3. The layout of both Figures 2 and 3 corresponds to the reading time panels in Figure 1.

The coefficient estimates depicted in Figures 2 and 3 are contrasts on the

log-transformed reading times (normalized as described above) and so may be directly interpreted as differences on the log scale, or as multiplicative effects on the original untransformed scale. As in Figure 1, effects obtained with the original Grodner and Gibson (2005) materials are plotted in black lines, and effects obtained with the new materials are plotted in grey lines. The HPD intervals that include zero (and therefore fail to reach conventional levels of significance) are plotted as dotted lines; intervals corresponding to conventionally significant effects are plotted as solid lines.

### *Results*

Analyses were conducted first using all trials, then again excluding trials on which the comprehension question was answered incorrectly. Because none of the analyses were affected by excluding incorrect trials, we report analyses that include all trials.

*Locality effects (see middle panel, top row of Figure 2).* There was an effect of locality in the embedded verb conditions but not in the matrix verb conditions; i.e. the non-local conditions (where the critical verb and its subject were separated by a PP or RC) were read more slowly than the local condition (where the subject and critical verb were adjacent, or local), but this effect was only reliable in the embedded conditions. In the embedded conditions, critical verbs in the RC condition were read more slowly than critical verbs in the PP condition, but this was not true in the matrix conditions.

*Embedding effect and interactions (see middle panel, top row of Figure 3).* Reading times at the critical verb were reliably slower overall in the embedded verb conditions than in the matrix verb conditions. The *locality* effect was larger in the embedded verb conditions, and the difference between PP and RC modification was also larger in the embedded conditions; though these differences were only marginally reliable.

*Discussion*

Experiment 1 replicated the basic pattern observed in Grodner and Gibson (2005). There was a locality effect in the embedded verb conditions but not in the matrix verb conditions. There was also a marginal interaction between the locality of the subject-verb dependency and the complexity introduced by embedding the verb.

These results therefore are also ambiguous concerning the nature of locality effects in the ways detailed above in the analysis of the Grodner and Gibson (2005) results. Locality effects in the embedded verb conditions could be rooted in the distance between the subject and verb, but they could also reflect an unrelated source of difficulty rooted in the computational demands of integrating an argument that has moved to a non-argument position. Additionally, Experiment 1 offers no way to measure the influence of similarity-based retrieval interference (between the subject and object) on locality effects in the embedded verb conditions. The absence of locality effects in the matrix verb conditions fails to disambiguate the results of the embedded verb conditions. As a result, Experiment 1 provides no conclusive evidence of whether locality effects arise from constraints on working memory or from another source of complexity specific to  $\bar{A}$ -movement. Experiment 1 also cannot distinguish between retrieval interference and distance-based degradation of memory representations as sources of difficulty.

For present purposes, Experiment 1 serves the dual role of providing further motivation for the eye tracking and lexical manipulations of Experiments 2–4, and providing an SPR baseline for the Grodner and Gibson (2005) materials in the same participant population used in the subsequent experiments. We defer discussing Experiment 1's results further until we can do so in the context of the results of the next experiments.

## Experiment 2: Eye tracking version of Experiment 1

Experiment 2 was an eye tracking version of Experiment 1 (and Grodner and Gibson (2005) Experiment 2).

### *Methods*

#### *Participants*

Forty-seven University of Michigan undergraduates participated for partial course credit or for pay.

#### *Apparatus*

Fixation time measures were gathered from both eyes using an SMI (SensoMotoric Instruments) Eyelink I head-mounted eye-tracker running at a 250 Hz sampling rate. Data from the right eye was used for all analyses.

#### *Stimuli*

The stimuli for this study were the same as Experiment 1. The same two items were removed from analysis due to design problems.

### *Procedure*

Participants were seated with their eyes twenty inches in front of a 17-inch CRT computer monitor, and the eye-tracker was fitted to their head. After the eye-tracker was calibrated using Eyelink-I software, participants began the first of twenty practice trials. Participants fixated a cross in the middle of the screen before every trial to allow the experimenter to verify the calibration of the tracker. As soon as the experimenter observed stable fixation on the fixation cross, he pressed a button to replace the central cross with an identical one at the left edge of the screen. The entire sentence for the trial was presented as soon as the participant made a stable fixation on this fixation cross.

Fixation data was gathered continuously throughout each trial. When the participant finished reading the sentence, he pressed the space bar and a comprehension question appeared, and the participant proceeded as in Experiment 1.

## *Results*

### *Question accuracy*

Four participants were excluded from analyses for answering fewer than 70% of the comprehension questions correctly. The remaining participants averaged 80% accuracy on the comprehension questions for this experiment.

### *Reading time measures and covariates*

Definitions of the eye movement measures used in the analysis of Experiments 2 and 4 are given in Table 4. Note that our definition of First Fixation Duration excludes single fixations: it is the duration of the first fixation of multiple fixations, but we retain the shorter label for convenience. Mixed effects models were constructed for each measure using the contrasts given in Table 3; as described in detail above Figures 2 and 3 show the contrast estimates and associated HPD intervals.

### *Spillover*

Last-pass reading time from the word immediately before the critical verb was used to model spillover. (See Table 4 for a definition of last-pass reading time). The length and frequency of the preceding word were also used as covariates. Spillover was modeled for Single Fixation Duration, First Fixation Duration, and First-Pass Reading Time in all the results we report. Last-Pass Reading Times from the previous word accounted for a near-significant amount of variance in First Fixation Duration—which suggests that measuring spillover this way may be reasonable.



*Reading times*

Analyses were conducted with and without incorrect trials. Because excluding incorrect trials did not change any results, we report analyses over all trials.

*Locality effects (Figure 2).* There were locality effects in the matrix verb conditions in two first-pass measures—Single Fixation Duration and First-Pass Reading Time (but not First Fixation Duration)—as well as Total Fixation Time. The embedded verb conditions showed a locality effect only in Total Fixation Time and in later, regression based measures. More specifically, there was a locality effect in Re-Reading Time, and a marginal effect in Regression Path Duration and Non-zero Regression Path Duration.

There was no difference between PP- and RC-modification in the matrix verb conditions. In the embedded verb conditions, critical verbs in RC sentences were slower than in PP sentences in Regression Path Duration, Non-zero Regression Path Duration and Total Fixation Time.

*Embedding effect and interactions (Figure 3).* Reading at the embedded verb was slower than the matrix verb in all measures. The locality effect differed between the matrix and embedded verb conditions only in Re-reading Time, more specifically, it was larger in the embedded verb conditions (see the *locality by embedding* interaction in Figure 3). Additionally, the difference between PP and RC modification was greater in the embedded verb conditions in Total Fixation Time and all the later measures (Re-Reading Time, Regression Path Duration and Nonzero Regression Path Duration).

*Discussion*

Consistent with prior studies that have paired SPR and eye tracking (e.g., (Ferreira & Clifton, 1986; Ferreira & Henderson, 1990; Kennison, 2002; Trueswell, Tanenhaus, & Kello, 1993), Total Fixation Time (and Re-Reading Time) yielded times similar to SPR,

in both qualitative pattern and absolute value. This relationship was most evident in the embedded verb conditions, where both SPR and Total Fixation Time (and Re-Reading Time) monotonically increased with increased subject-verb distance, with a large increase in the most complex condition, the embedded relative clause sentences.

The most interesting results from Experiment 2 concern distinct locality patterns in the early versus late fixations. There are two major sets of results.

The first major set of results is the presence of locality effects in the simpler matrix conditions in the earlier measures. This can be appreciated by inspection of the middle row of Figure 1, which reveals a consistent monotonic increase in times across the matrix conditions for Single Fixation Duration and First-Pass Reading Times. (Nevertheless, there were no reliable locality effects found in First Fixation Duration for any experiments. This is consistent with the locality effects found in Single Fixations and First-Pass Reading Times *not* being driven by spillover from the previous word.) The PP vs. RC contrast was not reliable for the matrix condition, but there was a consistent trend of greater reading times in RC conditions across all the early measures.

The second major set of results from Experiment 2 is that only later measures (Re-Reading, Regression Path Duration and Non-zero Regression Path Duration) captured the most salient result of the self-paced reading experiment: a sharp increase in reading times in the most complex doubly-embedded condition.

The locality effect was not reliable in every eye movement measure. While Total Fixation Time showed a locality effect for both the matrix and the embedded conditions, there were differences between the matrix and embedded verb conditions in other measures. For the matrix verb conditions, there was a locality effect in first-pass measures (Single Fixation Duration and First-Pass Reading Time). For the embedded verb conditions, there was a locality effect only in Re-Reading Time.

Taken together, the results of Experiment 2 are important for two reasons. First,

they suggest that locality effects are *not* restricted to complex movement structures. Second: For the more complex embedded conditions, the locality effect found in self-paced reading appears in *regressive* eye-movements to and possibly from the critical verb, not in first-pass fixation durations. We offer the following tentative hypothesis to explain this finding. First-pass measures may reflect, in part, the duration of short-term memory retrievals that underlie successful integration, while later measures reflect recovery processes that occur when argument retrieval cannot be completed on time (i.e., before a programmed saccade must be executed). In the current materials, these retrieval failures in the most difficult of the embedded conditions may be a result of the combined effect of locality and similarity-based interference as described above. Experiment 4 offers further data relevant to assessing this hypothesis. For now, we note that SPR times do not distinguish between recovery processes that show up in regressions and other processing variables that are reflected in first-pass measures.

### Interim summary

Experiment 1 replicated the results of Grodner and Gibson (2005), and provided a baseline for evaluating the relationship between SPR and eye-tracking measures. Experiment 2 provided the first on-line evidence of locality effects in non-movement, non-extraction structures, suggesting that locality effects reflect difficulty accessing degraded memory representations to build syntactic structure. Experiments 3 and 4, presented below, use new materials to measure how intra-dependency distance affects comprehension when lexical processing difficulty and similarity-based interference are attenuated. These two experiments provide a more focused measure of how comprehension is shaped by the difficulty of building syntactic structure from degraded short-term memory representations of linguistic input. The results will also provide a necessary complement to Experiments 1 and 2 as a basis for modeling the relationship

between dependency integration, eye movements, and the button-presses involved in SPR.

**Experiment 3: Testing locality effects using self-paced reading  
with short, high-frequency words**

Experiment 3 was a replication of Experiment 1 using a new set of materials composed from a more tightly controlled set of short, high frequency words. Our motivations for this manipulation were detailed earlier. Briefly: (1) By ensuring relatively easy lexical processing, we hope to increase the likelihood of detecting locality effects in the matrix verb conditions. If this goal is met, we may find locality effects in the matrix verb conditions, where none were found in the previous self-paced reading experiment. (2) Using inanimate objects as arguments for the critical verb decreases retrieval interference between the subject and object. As a result, we predict reduced estimates of locality effects in the embedded verb conditions, as well as a reduction or elimination of the embedding effect and the interaction between locality and embedding. (3) Using inanimate objects may also make parsing easier in the embedded verb conditions by lessening the unexpectedness of the critical verb.

*Methods*

*Participants*

49 University of Michigan undergraduate students participated for partial course credit or for pay.

*Stimuli*

Thirty experimental sentences were created for use in a self-paced reading experiment (Experiment 3) and a parallel eye tracking (Experiment 4). The syntactic structure of all sentences was identical to Experiments 1 and 2, and Grodner and Gibson (2005) Experiment 2, but content words were restricted: Sentences consisted of

three-to-six letter words that had a frequency higher than fifty occurrences per-million-words in the First Release of the American National Corpus.<sup>6</sup> A comparison of the new and old materials based on lexical properties is given in Table 5

Unlike the Grodner and Gibson (2005) materials, which have a human referent as both the subject and direct object of the critical verb, the new materials have an inanimate direct object. As described above, this change should reduce similarity-based retrieval interference at the critical verb, providing better estimates of the independent effects of locality.

Using inanimate referents in the object role also helps test the possibility that, as PDC predicts, locality effects in the embedded conditions of experiment 1 and 2 partially result from difficulty processing object-extracted RCs with rare configurations of argument properties. Object-extracted structures with a previously unmentioned, animate referent in the object role (like in experiments 1 and 2) are infrequently produced, and therefore PDC predicts they should be more difficult. Object-extracted structures with an inanimate referent as the object (like in experiments 3 and 4) are, however, common. Therefore, in the new materials, increased difficulty in the non-local conditions can be interpreted more directly as the result of the distance between the critical verb and its subject.

The plausibility of materials in experiments 3 and 4 was also measured. 57 participants read each experimental item at one level of subject-modification, distributed randomly amongst 54 filler sentences, and rated plausibility on a 5-point Likert scale. Table ?? shows the mean ratings for each level of modification.

While the average rating for each level of subject-modification was above average, sentences with an unmodified subject were rated significantly more plausible than the PP-modified ( $t(994)=9.66$   $p < .05$ ) and RC-modified subjects ( $t(1070)=-5.58$ ,  $p < .05$ ). Mean plausibility ratings for each item in each condition was included as a predictor in all

statistical models of experiments 3 and 4. None of the results reported were affected by the inclusion of plausibility as a predictor.

Items were assigned to lists using a Latin square design. Experimental items never appeared consecutively, and no arguments or argument modifiers were used more than once. Table 7 gives examples of the materials.

### *Procedure*

The procedure was identical to Experiment 1. Participants pressed the space-bar on a keyboard to advance through each sentence, and then answered a comprehension question about the sentence.

## *Results*

### *Question accuracy*

Participants responded more accurately to comprehension questions in the second experiment, averaging 92% accuracy across all trials, suggesting that the lexical manipulation succeeded in reducing overall difficulty. As in Experiment 1, participants failing to meet a 70% accuracy criterion were excluded from analysis. This disqualified one participant. Data from the remaining forty-eight participants were analyzed. One item was removed from analysis because it was displayed with words missing. Another item was removed because the critical verb did not meet the word frequency criterion; a third was removed because the sentence was missing its subject. The remaining 27 items were analyzed.

### *Reading Times*

The self-paced reading times at the critical verb are presented graphically in Figure 1 (top row, middle panel, grey lines), and HPD intervals corresponding to the seven contrasts of interest are presented in Figure 2 and Figure 3.

*Locality effects (Figure 2).* There was a locality effect in both the matrix and embedded verb conditions: reading times at the critical verb were longer in the non-local conditions than the local conditions. There were no reliable differences due to modification (the PP vs. RC contrast). The RC and PP contrast was reliably larger in the original materials than the new materials. This was established by a mixed effects model combining the data from the two SPR experiments that included a contrast coding the interaction of materials set and the embedding effect (contrast estimate = -0.056, HPD intervals: *lower* = -0.107, *upper* = -0.005).

*Embedding effect and interactions (Figure 3).* Embedded verbs were read more slowly overall than matrix verbs. There were no reliable interactions, and unlike Experiment 1, these interactions did not approach conventional significance.

The embedding effect found in Experiments 1 and 2 appeared to be reduced, suggesting that replacing the object with an inanimate noun phrase made the embedded verb sentences easier to comprehend. However, this cross-experiment difference in the embedding effect, tested by a contrast coding the interaction of materials set and verb embedding, showed no reliable difference between the SPR experiments (coefficient estimate=0.007; HPD intervals: *lower* =-0.02, *upper* =0.03).

### *Discussion*

The most important result of Experiment 3 is the locality effect in the matrix verb conditions, which was not found in Experiment 1. Using short, high-frequency words, locality effects were detectable where they were not detected in Experiment 1 (which was also run in SPR).

In short, the two empirical goals were met: the materials manipulation made it possible to detect a locality effect in the matrix condition, and locality effects manifested in more similar ways in both the matrix and embedded clause conditions. The evidence

from Experiment 3 reinforced the tentative conclusion we advanced in Experiment 2, namely, that locality effects exist outside of  $\bar{A}$ -movement. The new materials also ruled out an explanation based on retrieval interference between the subject and object, and another based on the relative rarity of object-extracted structures with an animate, discourse-new direct object. Experiment 3 yielded important evidence that long-distance dependent-head integration is difficult because underlying representations of the dependent and head degrade over time. It also provided evidence suggesting that locality may interact with overall processing complexity.

#### **Experiment 4: Eye tracking version of Experiment 3**

Experiment 4 was an eye tracking version of Experiment 3. Using shorter lexical items has the further advantage in eye tracking of reducing the number of fixations on individual words (Brysbaert & Vitu, 1998; K. Rayner, 1979), which should increase the number of data points available to analyse as Single Fixations.

#### *Methods*

##### *Participants.*

Forty-five University of Michigan undergraduates participated for partial course credit or for pay.

##### *Stimuli.*

The stimuli were identical to Experiment 3.

##### *Procedure*

The procedure was identical to Experiment 2. Participants read each sentence and then answered a yes-or-no comprehension question about the sentence. Eye movement data were collected.



## Results

### *Question accuracy*

Participants averaged 92% accuracy across all conditions. All participants met the minimum accuracy criterion of 70%.

### *Reading times*

The same eye-movement measures used in the analysis of Experiment 2 were used to analyze Experiment 4 data, and these measures are plotted as solid grey lines along side the Experiment 2 results in Figure 1. The same seven contrasts in Table 3 were analysed using mixed effects models with the same structure as Experiment 2, including covariates for length and frequency of the verb and the preceding word. The contrast estimates and HPD intervals are shown in Figure 2 and Figure 3.

*Locality effects (Figure 2).* There was a locality effect for the matrix verb conditions in the first-pass measures: Single Fixation Duration, and First-Pass Reading Time. In the embedded verb conditions, there was a locality effect in Single Fixation Duration, Regression Path Duration and Total Fixation Time.

Reading times for PP and RC sentences did not differ in any measure for the matrix verb or embedded verb conditions.

*Embedding effect and interactions (Figure 3).* Embedding the verb led to increases in Re-Reading Time and Non-zero Regression Path Duration.

There was only one reliable interaction: The locality effect was smaller in the embedded verb conditions than the matrix verb conditions in First Fixation Duration<sup>7</sup>

A comparison between the two eyetracking experiments showed a smaller embedding effect in the new materials in all measures but Single Fixation Duration and Non-zero Regression Path Duration (First Fixation:  $HPD_{min} = 0.04$ ,  $HPD_{max} = 0.22$ ;

First-Pass Reading:  $HPD_{min} = 0.05$ ,  $HPD_{max} = 0.13$ ; Regression Path:  $HPD_{min} = 0.06$ ,  $HPD_{max} = 0.16$ ; Re-Reading:  $HPD_{min} = 0.15$ ,  $HPD_{max} = 0.32$ ; Total Fixation Time:  $HPD_{min} = 0.18$ ,  $HPD_{max} = 0.28$ .)

### *Discussion*

There are three key results from Experiment 4. First, there were locality effects in the matrix verb conditions, as there were in Experiments 2 and 3. As one can see in Figure 1, there was a consistent increase in reading times (denoted by the grey lines) from local (no modification) to non-local (PP and RC-modification) in the Matrix condition across all the measures except First Fixation Duration and Re-Reading Time.

Second, in contrast to Experiment 2, there was a relatively early locality effect for the embedded conditions, in addition to a locality effect in re-reading.

Third, and perhaps most striking, the main effect of embedding was eliminated in the early measures and was reliable only in Re-Reading Time and Non-zero Regression Path Duration. This suggests that embedding the verb did not cause consistent difficulty integrating the verb with its arguments. One possibility is that the embedding effects obtained in this experiment reflect retrieval failures triggering regressions from the critical verb.

One aspect of the data pattern in Experiment 4 is somewhat surprising: the absence of a locality effect in Total Fixation Time for the matrix verb conditions. However, we argue that this result does not suggest that subject-verb integration is unaffected by locality in the matrix verb conditions, because there were reliable locality effects in Single Fixation Duration and First-Pass Reading Times. Rather, the absence of a locality effect in Total Fixation Time reflects appears to reflect the very high variance in re-reading measures.

## General Discussion

Locality effects are important because they potentially inform us about the the short-term memory processes that underlie the on-line computation of linguistic relations in language comprehension. But as argued in the Introduction, the evidence for locality overall is surprisingly mixed, and the existing on-line evidence is both linguistically and methodologically narrow, while at the same time admitting alternative explanations that do not invoke memory limitations.

The four experiments presented in this paper were intended to broaden the evidential base and provide new insights into locality and its empirical manifestation. In the remainder of the paper we organize these insights under three topics: (a) conclusions about the putative ubiquity and nature of locality effects; (b) conclusions about the relationship of eye tracking and self-paced reading measures; and (c) implications for current theoretical approaches to parsing.

### *The ubiquity and nature of locality effects*

There are four main conclusions that we draw from Experiments 1–4 concerning the extent and nature of locality effects. These conclusions represent tentative answers to the motivating questions in the Introduction.

1. Locality effects are ubiquitous: they emerge not only in the computation of relatively difficult embedded structures involving  $\bar{A}$  movement (as replicated in experiment 1), but may be detected in the computation of relatively simple subject-verb relations (as shown for the matrix conditions in experiments 2 – 4). While experiment 1 replicated an earlier null finding for the matrix conditions, experiments 2 – 4 consistently showed that locality effects do exist in those structures; but they may be obscured by lexical processing variance and processing constraints of a non-cumulative display task. Experiments 3 and 4 ruled out three additional explanations for locality effects, which

were unrelated to special classes of argument movement. First,

2. Although locality effects may be observed in easy-to-process materials, the effects interact (over-additively) with overall processing complexity, including complexity with structural and lexical origins. This was demonstrated in both experiments with the new materials, where locality effects were smaller than in the experiments with the original materials, particularly in the complex, embedded verb conditions, due to a smaller difference between the PP and RC conditions.

3. The largest and most robust effects of locality that previously have been empirically observed in SPR correspond to retrieval failures during parsing. The effects are large in SPR in part because they include time to recover from failure. The very high reading times observed in the most complex conditions in SPR correspond well to patterns in Re-Reading Time and Regression Path Duration.

4. The locality effects obtained in the present experiments appear to be robust against spillover effects and plausibility differences. Locality effects emerged in the both the matrix and the embedded verb conditions when lexical properties and reading times from the pre-critical word were included in the model. Furthermore, locality effects were not evident in First Fixation Duration, where spillover effects would be expected, and where they were empirically observed. Modeling sentence plausibility for experiments 3 and 4 did not alter any results.

*The relationship of eye tracking and self-paced reading measures to  
locality effects*

There is a small literature in which parallel self-paced and eye tracking versions of the same experiment have been run (e.g., Kennedy & Murray, 1984; Ferreira & Clifton, 1986; Ferreira & Henderson, 1990; Speer & Clifton, 1998; Binder & Rayner, 1998; Pearlmutter, Garnsey, & Bock, 1999; Hoeks, Vonk, & Schriefers, 2002; Kennison, 2002;

Traxler, Pickering, & McElree, 2002), and authors have generally observed either that the two types of measurement lead to similar conclusions (because both show the predicted result), or that eye tracking has greater temporal resolution than SPR. What we add to this literature are more detailed observations and hypotheses about the relationship of eye tracking and SPR measures to locality effects, and more generally to structural complexity effects in parsing relatively unambiguous structures. In particular, the present study is in the spirit of Pearlmutter et al. (1999), which noted that regression-based analyses are both useful and unique to eye tracking.

We draw two main conclusions concerning SPR, eyetracking, and locality. First, early measures (including First Pass Reading Times, Single Fixation Durations, and First Fixation Durations) do provide meaningful measures of structural integration processes. The locality effects present in the early measures in Experiments 2 and 4 support this claim.

Second, self-paced-reading times include time spent recovering from parsing failures, and therefore may not provide good estimates of the effect of locality on successful components of parsing. The evidence for this comes from two sources: the correspondence between long SPR times and later eye-tracking measures, and the presence of locality effects in early measures on simpler structures.

*Toward a model of locality effects based on short-term memory retrieval  
and adaptive control of eye movements and button presses*

It is possible to account for the phenomena in Experiments 1–4 with a theoretical model that combines existing independently motivated proposals for short-term memory processes with adaptive control of button-presses in SPR and eye-movements in reading. We sketch the basic principles of this model here and describe how it yields several interesting predictions concerning the relationship of SPR and eye-tracking measures to

locality, structural complexity, and lexical frequency. A key advantage of this kind of model is that the relationship of the empirical measures to each other and to the underlying memory processes is not stipulated as a set of *linking assumptions* Boland (2004); Clifton, Staub, and Rayner (2007), but rather follows from a theory of adaptive control of motor responses.

In what follows we summarize the theory in two parts—assertions about the parsing and memory processes, and assertions about control—and then describe briefly how the theory accounts for empirical patterns in Experiments 1–4.

*The nature of the memory and parsing processes: the locus of locality effects*

We adopt the retrieval model of Lewis and Vasishth (2005); Lewis et al. (2006), which provides an integration of interference and locality effects; see especially Lewis and Vasishth (2005) for more details.

1. On-line sentence comprehension (in all modalities) consists in part of the word-by-word cue-based retrieval (or reactivation) of short-term representations of partial linguistic structures created at earlier points of the sentence.

2. The retrieval of these prior representations is negatively affected by (a) increasing the temporal distance between initial creation of the representation and retrieval time and (b) increasing the number and similarity of distractors to the target representation-to-be-retrieved (these distractors may have occurred before the target, leading to effects labeled as proactive interference, or after the target, leading to effects labeled as retroactive interference)<sup>8</sup>.

3. The negative effect on retrieval consists of both increased retrieval times and increased probability of retrieval failure (Lewis & Vasishth, 2005).

*The adaptive control of button presses and eye movements*

There is now clear evidence for the local adaptation of behavioral control to the joint constraints of local task structure, reward, and cognitive constraints (Howes, Lewis, & Vera, 2008 (submitted)). We assume here that button presses in self-paced reading and eye-movements in reading are also subject to locally adaptive control (cf. Reichle and Laurent (2006)). More specifically:

1. For present purposes, we simplify the adaptive control problem to the problem of finding the appropriate signal from the (partial) processing of each word to trigger the preparatory processes for advancing the eye (i.e., programming the saccade) or making a button press. Possible signals range from the completion of early stages of orthographic encoding to partial completion of lexical access to partial completion of the short-term retrieval processes required for structural composition. The optimal control signal is one that allows comprehension to proceed as swiftly and accurately as possible under the prevailing paradigm constraints and reward structure.

Figure 4 — Figure ?? present a simplified illustration of possible control signals coordinating visual and linguistic processes. In the figures there are four different possible timelines associated with reading a pair of words ( $n$  and  $n + 1$ ) in a sentence. The boxes depict a cascaded pipeline of processes; each row of processes corresponds to a type of visual or linguistic processing for which there might plausibly be constraints on parallel processing, so that the most efficient processing arises when each stream is maximally occupied. Early signals are favored over late to the extent that they allow for more efficient parallel pipelining of comprehension processes. Late signals are favored over early to the extent that they ensure sufficient time for comprehension sub-processes to complete so that incoming information from the next word does not interfere with processing from the prior word. Figure 4 depicts a situation where lexical processing is the bottleneck, and

the optimal control signal for advancing the eye derives from some degree of completion of lexical processing. This would lead (probabilistically) to fixation durations on word  $n$  primarily determined by lexical-level properties associated with word  $n$ . The idea of the *optimal* control signal can be appreciated by contrasting the timeline in the first figure with the next two. Figure 5 depicts a situation where the control signal comes too late; all processing stages complete but *slack time* has been introduced indicating that the processing is slower than it could be. Figure 6 depicts a situation where the control signal comes too early so that lexical processing of word  $n$  has not had time to complete before lexical processing of word  $n + 1$  has begun, hypothesized to increase the probability of processing failure.

2. The control of self-paced button presses in the non-cumulative moving window paradigm will tend to be more conservative (i.e., tend to adopt later control signals) than the control of eye-movements, because the probability and cost of error recovery failure is relatively higher. When sentence viewing is not restricted by a moving window, eye movement control can afford to be more aggressive (i.e., tend to adopt earlier control signals), in part because parsing errors that arise from moving the eyes too quickly may be reliably recovered from by re-reading via regressive eye-movements.

3. Given the non-cumulative display, recovery from short-term retrieval failure may happen via exploiting an alternative memory trace: the phonological loop (Baddeley, 1992). But doing so requires a serial reactivation or rehearsal of early parts of the sentence, and this takes considerably more time than the elementary short-term retrieval processes that normally underlie effective comprehension.

4. If the computation of multiple relations is required (such as the embedded verb in the present experiments), and if there is a logical dependency between the relations that imposes a serial order on their computation, then this ordering will be reflected in the eye-movement measures: factors affecting the relation(s) computed earlier will be reflected



in early measures, factors affecting relations(s) computed later will be reflected in later measures.

*Accounting for the key phenomena*

This model accounts for all six of the basic results from Experiments 1–4, described above.

1. Ubiquitous locality effects should be present even in simple structures because all structural integrations require short-term retrievals that are slowed by increased distance.

2. But it may be difficult to observe such effects in materials with lexical items that take relatively long to process because it is more likely that lexical processing is the bottleneck (or, in scheduling terms, on the *critical path*), making it correspondingly more likely that the optimal control signal for advancing the eyes is derived from lexical processes rather than higher-level retrieval processes. Conversely, it is more likely that locality effects will emerge consistently when lexical processing is made relatively easier; then the optimal control signal is more likely derived from higher-level retrieval processes. This can be seen in the contrast between the timeline in Figure 4, where lexical processing is the bottleneck, and the timeline in Figure 7, where structural processing is the bottleneck.

3. But the magnitude of the locality effects will tend to be reduced when lexical processing is made easier, because reading will be faster overall, mitigating effects of short-term memory decay. This gives rise to an over-additive interaction of locality and lexical frequency—but one that does not have its source in a direct effect of frequency on the underlying memory processes that yield the locality effect.

4. Points of parsing difficulty in SPR will be associated with disproportionately long reading times to the extent that retrieval failures are the source of the parsing difficulty, because the error recovery from such failures involves costly phonological rehearsal

processes. Furthermore, such effects will be associated with regressions in the eye-movements, rather than sharply increased reading times in first-pass measures.

5. Points of both high interference and distal relations increase the probability of retrieval failure because both effects combine to determine the activation of the target item and distractors. Combined with the subsequent costs of error recovery, this gives rise to an over-additive interaction of locality and structural complexity—but one that does not have its source in a direct effect of complexity on the underlying memory processes that yield the locality effect.

6. The computation of the antecedent for the “gap” at the embedded verb in the relative clause will be associated with later eye-movement measures because that relation must be computed *after* the successful retrieval of the verb prediction itself.

In short, this model explains both the emergence and compression of locality effects in the simpler lexical materials, the relationship of SPR and early vs. late eye-movements, and the *empirical* over-additive interaction of locality and both structural and lexical processing complexity—without appealing to an underlying interaction at the level of memory processes. This has the virtue of keeping the quantitative range of empirical effects that more directly reflect short-term memory retrieval processes in a range consistent with estimates of short-term memory retrieval durations in the general STM literature (McElree, 2006).

### *Open questions*

The present experiments raise important new questions and bring some lingering questions into focus. While the results of experiments 1 – 4 show that degraded short-term memory representations result in locality effects, even in simple structures, a worthy next goal is to flesh out how memory limitations might interact with other sources of comprehension difficulty such as violation of expectations. The present results are

consistent with memory-based accounts like DLT and the ACT-R inspired account of Lewis et al. (2006), and inconsistent with at least one variety of statistically-driven expectation account (PDC), but there is no compelling logical or experimental evidence that memory retrieval dynamics and expectations must have additive effects on comprehension difficulty. A useful extension of this work might be to test dependency locality and strength of expectations directly against each other.

The eye movement model outlined here must also be empirically tested. In particular, future work might target the purpose of regressions. Since regressions sometimes reflect major comprehension difficulty due to factors such as locality, it is important to pursue more detailed models of both the triggers of regressions and the cognitive events that follow. To understand what regressions tell us about locality effects, we must know whether failed long-distance structural integration results in selective reanalysis or relatively unguided back-tracking in the sentence. It is possible that both the trigger of a regression and the efficiency of the regression (in terms of spatial precision and information-finding) may differ between long-distance dependency integration and other instances of high processing load or memory load.

In the short term, the experiments presented here provide important and novel support for the claim that locality effects are a ubiquitous phenomenon in English, and that they can be traced to retrieval difficulty in long-distance dependencies that occurs because the short-term representations of elements in the dependency become less accessible as a function of time. On the basis of these four experiments, we argue that a complete model of parsing processes must account for the dynamics of activation in working memory, particularly at points where linguistic dependencies are established.

## References

- Altmann, G., Nice, K. van, Garnham, A., & Henstra, J. (1998). Late closure in context. *Journal of Memory and Language*, *38*, 459–484.
- Baddeley, A. (1992). Working memory. *Science*, *255*, 556–559.
- Bates, D., & Sarkar, D. (2007). lme4: Linear mixed-effects models using s4 classes [Computer software manual]. (R package version 0.9975-11)
- Binder, K., & Rayner, K. (1998). Contextual strength does not modulate the subordinate bias effect: Evidence from eye fixations and self-paced reading. *Psychonomic Bulletin and Review*, *5*(2), 271-276.
- Boland, J. (2004). *Linking eye movements to sentence comprehension in reading and listening*. Routledge.
- Brysbaert, M., & Vitu, F. (1998). Word skipping: Implications for theories of eye movement and control in reading. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 125–148). Oxford, England: Elsevier.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Clifton, C., 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. 341–372). Amsterdam: Elsevier.
- Ferreira, F., & Clifton, J., C. (1986). The independence of syntactic processing. *Journal of Memory and Language*, *25*, 348–368.
- Ferreira, F., & Henderson, J. (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*(4), 555–568.
- Frazier, L., & Fodor, J. D. (1978). The sausage machine: A two-stage parsing model. *Cognition*, *6*, 291-325.
- Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical*

*models*. Cambridge, UK: Cambridge University Press.

Gennari, S., & MacDonald, M. (2009). Linking production and comprehension processes: the case of relative clauses. *Cognition*, *111*, 1–23.

Gibson, E. (1991). *A computational theory of human linguistic processing: Memory limitations and processing breakdown*. Unpublished doctoral dissertation, Carnegie Mellon University, Pittsburgh, PA.

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

Gibson, E. (2000). Dependency locality theory: A distance-based theory of linguistic complexity. In A. Marantz, Y. Miyashita, & W. O'Neil (Eds.), *Image, language, brain: Papers from the first mind articulation project symposium*. Cambridge, MA: MIT Press.

Gibson, E., Pearlmutter, N., Canseco-Gonzales, E., & Hickock, G. (1996). Recency preference in the human sentence processing mechanism. *Cognition*, *59*, 23–59.

Gibson, E., Pearlmutter, N., & Torrens, V. (n.d.). Recency and lexical preference in spanish. *Memory and Cognition*, *27*, 603–611.

Gibson, E., & Warren, T. C. (2004). Reading time evidence for intermediate linguistic structure in long-distance dependencies. *Syntax*, *7*, 55–78.

Gordon, P. C., Hendrick, R., Johnson, M., & Lee, Y. (2006). Similarity-Based Interference During Language Comprehension: Evidence from Eye Tracking During Reading. *Journal of Experimental Psychology: Learning Memory and Cognition*, *32*(6), 1304–1321.

Grodner, D., & Gibson, E. (2005). Consequences of the serial nature of linguistic input. *Cognitive Science*, *29*, 261–290.

Grodner, D., Gibson, E., & Tunstall, S. (2002). Syntactic complexity in ambiguity resolution. *Journal of Memory and Language*, *46*, 267–295.

- Grodzinsky, Y. (2000). The neurology of syntax: Language use without broca's area. *Behavioral and Brain Sciences*, *23*, 1–21.
- Hoeks, J., Vonk, W., & Schriefers, H. (2002). Processing coordinated structures in context: The effect of topic-structure on ambiguity resolution. *Journal of Memory and Language*, *46*, 99–119.
- Howes, A., Lewis, R. L., & Vera, A. (2008 (submitted)). *Rational behavior under task and processing constraints: Implications for testing theories of cognition and action*.
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*(1), 122–149.
- Kennedy, A., & Murray, W. (1984). Inspection times for words in syntactically ambiguous sentences under three presentation conditions. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(6), 833–849.
- Kennison, S. (2002). Comprehending noun phrase arguments and adjuncts. *Journal of Psycholinguistic Research*, *31*(1), 65–81.
- Kimball, J. (1973). Seven principles of surface structure parsing in natural language. *Cognition*, *2*, 15–47.
- Konieczny, L. (2000). Locality and parsing complexity. *Journal of Psycholinguistic Research*, *29*, 627–645.
- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177.
- Lewis, R. L., & Vasishth, S. (2005, May). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, *29*, 1–45. Available from <http://www.ling.uni-potsdam.de/vasishth/Papers/cogsci05lewisvasishth.pdf>
- Lewis, R. L., Vasishth, S., & Van Dyke, J. (2006, October). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Science*, *10*(10),

447-454.

- Mahajan, A. K. (1990). *The A/A-bar Distinction and Movement Theory*. Unpublished doctoral dissertation, MIT, Cambridge, MA.
- McElree, B. (2006). Accessing recent events. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 46). San Diego: Academic Press.
- Mitchell, D. C. (1984). An evaluation of subject-paced reading tasks and other methods of investigating immediate processes in reading. In D. E. Kieras & M. Just (Eds.), *New methods in reading comprehension research*. Hillsdale, N.J.: Erlbaum.
- Nakatani, K., & Gibson, E. (2004). *An online study of Japanese nesting complexity*. (MS)
- Pearlmutter, N. J., Garnsey, S. M., & Bock, K. (1999). Agreement processes in sentence comprehension. *Journal of Memory and Language*, *41*, 427-456.
- Pearlmutter, N. J., & Gibson, E. (2001). Recency in verb phrase attachment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 574-590.
- Pritchett, B. L. (1992). *Grammatical competence and parsing performance*. Chicago: University of Chicago Press.
- R Development Core Team. (2006). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Available from <http://www.R-project.org> (ISBN 3-900051-07-0)
- Rayner, K. (1979). Eye guidance in reading: Fixation locations within words. *Perception*, *8*, 21-30.
- Rayner, R., & Duffy, S. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory and Cognition*, *14*, 191-201.
- Reichle, E. D., & Laurent, P. A. (2006). Using reinforcement learning to understand the emergence of "intelligent" eye-movement behavior during reading. *Psychological Review*, *113*, 390-408.

- Speer, S. R., & Clifton, C. (1998). Plausibility and argument structure in sentence comprehension. *Memory and Cognition*, *26*(5), 965-978.
- Traxler, M., Pickering, M., & McElree, B. (2002). Coercion in sentence processing: Evidence from eye-movements and self-paced reading. *Journal of Memory and Language*, *47*, 530-547.
- Trueswell, J., Tanenhaus, M., & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden paths. *Journal of Experimental Psychology*, *19*, 528-553.
- Van Dyke, J. (2007). Interference effects from grammatically unavailable constituents during sentence processing. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *33*, 407-430.
- Van Dyke, J., & Lewis, R. L. (2003). Distinguishing effects of structure and decay on attachment and repair: A cue-based parsing account of recovery from misanalyzed ambiguities. *Journal of Memory and Language*, *49*, 285-316.
- Vasishth, S. (2002). *Working memory in sentence comprehension: Processing Hindi center embeddings*. Unpublished doctoral dissertation, Ohio State University, Columbus, OH. Available from <http://www.ohiolink.edu/etd/view.cgi?osu1023402958>
- Vasishth, S., & Lewis, R. L. (2006, December). *Argument-head distance and processing complexity: Explaining both locality and antilocality effects*. Available from <http://www.ling.uni-potsdam.de/vasishth/Papers/LangVasLewFinal.pdf>  
(To appear in *Language*, Vol. 82.4)
- Wanner, E., & Maratsos, M. (1978). An ATN approach to comprehension. In M. Halle, J. Bresnan, & G. A. Miller (Eds.), *Linguistic theory and psychological reality*. Cambridge, MA: MIT Press.
- Warren, T. C., & Gibson, E. (2002). The influence of referential processing on sentence complexity. *Cognition*, *85*, 79-112.



Wu, H. I., & Gibson, E. (2008). Processing chinese relative clauses in context. In *Proceedings of the 21st cuny conference on sentence processing*. University of North Caroline.

sectionPlots of reading time in all regions

### Footnotes

<sup>1</sup>There are further important distinctions to be made here about the nature of these representations—whether they involve *predictions* of the verb (Lewis & Vasishth, 2005; Gibson, 2000), the degree to which they are semantic (Van Dyke, 2007), etc.—but these distinctions are not relevant for present purposes.

<sup>2</sup>It has also been suggested that there is ambiguity as to whether the critical verb is the head of the critical subject-verb dependency or the beginning of a reduced relative clause. However, this ambiguity would only complicate interpretation of Grodner and Gibson (2005)’s results and other apparent locality effects if the reduced relative interpretation was more expected at the end of a subject-modifying relative clause than after an unmodified relative clause. This does not seem likely, and theories based on expectations or experience do not present any clear evidence of such a pattern.

<sup>3</sup>For present purposes we remain neutral about the precise nature of the subject relation—under some accounts it may also involve movement from within the verb-phrase to an argument position outside it. Under any analysis, the subject-verb dependency here is qualitatively different from the extracted object dependency.

<sup>4</sup>Each contrast vector was normalized by dividing it by the difference between the positive and negative coefficients coding the two groups. For example, to normalize the vector [ -2 1 1 0 0 0 ], we divide it by the difference between the positive coefficient 1 and the negative coefficient -2, or  $1 - (-2) = 3$ . The normalized vector is thus [  $-\frac{2}{3}$   $\frac{1}{3}$   $\frac{1}{3}$  0 0 0 ].

<sup>5</sup>Reading times for all regions are plotted in , for each experiment. Because those plots do not change our interpretation of the results, those plots will not be discussed further.

<sup>6</sup><http://www.americannationalcorpus.org/FirstRelease/>

<sup>7</sup>In fact, First Fixations show an anti-locality trend in the embedded verb conditions, although this trend is difficult to interpret in light of Total Fixation Time, which shows a

larger locality effect for the embedded verb conditions than the matrix verb conditions.

<sup>8</sup>In the ACT-R model of Lewis and Vasishth (2005) the distance effects generally emerge from decay, but it may be possible to dispense with decay given that retroactive interference also plays a role.

Table 1

*Extant experimental evidence for locality effects.*

PUBLICATION	LINGUISTIC STRUCTURES	METHODOLOGY
Gibson (1998)	subject- and object- relative clauses	self-paced reading
Grodner et al. (2002)	reduced-relative ambiguities	self-paced reading
Gibson & Warren (2004)	extraction across VP or NP	self-paced reading
Grodner & Gibson (2005)	subject- and object- relative clauses	self-paced reading
Wu & Gibson (2008)	subject- and object- relative clauses	self-paced reading
Van Dyke & Lewis (2003)	NP/S ambiguities	self-paced reading
Van Dyke (2007)	NP/S ambiguities	self-paced reading
Warren & Gibson (2002)	nested vs. non-nested structures	complexity ratings

Table 2

*Examples sentences from the six conditions in Experiments 1 and 2; the critical verb is underlined.*

CONDITION	EXAMPLE
<i>Unmodified</i>	The nurse <u>supervised</u> the administrator while...
<i>PP-modified</i>	The nurse from the clinic <u>supervised</u> the administrator while...
<i>Matrix</i>	
<i>RC-modified</i>	The nurse who was from the clinic <u>supervised</u> the administrator while...
<i>Unmodified</i>	The administrator who the nurse <u>supervised</u> scolded the medic while...
<i>PP-modified</i>	The administrator who the nurse from the clinic <u>supervised</u> scolded the medic while...
<i>Embedded</i>	
<i>RC-modified</i>	The administrator who the nurse who was from the clinic <u>supervised</u> scolded the medic while...

Table 3

*Two sets of contrasts used in the mixed effects models to analyse reading times from Experiments 1–4. Set 2 was a full matrix of five orthogonal contrasts, but only the theoretically interesting and non-redundant contrasts are shown here.*

CONTRAST		MATRIX			EMBEDDED		
		$\emptyset$	PP	RC	$\emptyset$	PP	RC
SET 1	Embedding effect (overall)	-0.50	-0.50	-0.50	0.50	0.50	0.50
	Local vs. non-local (matrix)	-0.67	0.33	0.33	0.00	0.00	0.00
	PP vs. RC (matrix)	0.00	-0.50	0.50	0.00	0.00	0.00
	Local vs. non-local (embedded)	0.00	0.00	0.00	-0.67	0.33	0.33
	PP vs. RC (embedded)	0.00	0.00	0.00	0.00	-0.50	0.50
SET 2	Locality $\times$ embedding interaction	0.75	-0.38	-0.38	-0.75	0.38	0.38
	Modification type $\times$ embedding interaction	0.00	-0.50	0.50	0.00	0.50	-0.50

Table 4

*Definitions of the eye tracking measures used in the analysis of Experiments 2 and 4.*

MEASURE	DEFINITION
<i>First Fixation Duration</i>	Time between the initial landing in a region and the first saccade; excludes trials where only one fixation was made.
<i>Single Fixation Duration</i>	Time spent fixating a region when only one fixation was made therein.
<i>First-Pass Reading Time</i>	The summed duration of all fixations made within a region before exiting to the right or left.
<i>Regression Path Duration</i>	The sum of all fixations within a region $n$ and in any regions to its left before fixating to the right of $n$ .
<i>Non-zero Regression Path Duration</i>	Identical to Regression Path Duration, but Non-Zero Regression Path Duration excludes cases where no regressions occurred.
<i>Re-reading Time</i>	The sum of all fixations in a region after it is exited (to the right or left). Re-reading analyses include zero-millisecond re-reading times.
<i>Last-Pass Reading Time</i>	The sum of all fixations in the last run of fixations within a region.
<i>Total Fixation Time</i>	The sum of all fixations within a region during a trial.

Table 5

*Lexical properties of each set of materials, through the critical verb position. The new materials included many plural forms of content words, not including the verb, whose singular forms met all length and frequency criteria. Statistics for those content words were computed for the plural forms participants saw. Frequency counts displayed are occurrences per-million-words in the American National Corpus.*

CRITICAL VERB	Old materials	New materials
Median length	8.0	4.0
<i>Std. Deviation</i>	1.6	.91
Median frequency	5.0	112.0
<i>Std. Deviation</i>	13.2	166.3
CONTENT WORDS	Old materials	New materials
Median length	7.00	5.00
<i>Std. Deviation</i>	2.56	.97
Median frequency	12.50	77.0
<i>Std. Deviation</i>	53.10	88.78



Table 6

*Mean plausibility ratings for each level of subject-modification used in the new materials for experiments 3 and 4.*

Subject-modification	Example	Mean rating
none	The child played sports...	4.29
PP	The child from the school played sports...	3.55
RC	The child who was from the school played sports...	3.9

Table 7

*Examples sentences from the six conditions for Experiments 3 and 4. The critical verb is underlined.*

CONDITION	EXAMPLE
<i>Unmodified</i>	The child <u>played</u> the sports that were hard to master.
<i>PP-modified</i>	The child from the school <u>played</u> the sports that were hard to master.
<i>Matrix</i>	
<i>RC-modified</i>	The child who was from the school <u>played</u> the sports that were hard to master.
<i>Unmodified</i>	The sports that the child <u>played</u> were hard to master.
<i>PP-modified</i>	The sports that the child from the school <u>played</u> were hard to master.
<i>Embedded</i>	
<i>RC-modified</i>	The sports that the child who was from the school <u>played</u> were hard to master.

### Figure Captions

*Figure 1.* Reading time measures from Experiments 1–4 and the original Grodner & Gibson (2005) self-paced reading study. Error bars are one standard error around condition means. Black lines indicate data collected using the Grodner & Gibson materials; grey lines indicate data collected using the materials composed of short, high-frequency words. The top row shows self-paced reading times from the Grodner & Gibson study (top left), self-paced reading times from Experiments 1 and 3 (top middle), and total fixation times from eye tracking Experiments 2 and 4 (top right). The middle row show the early eye tracking measures, and the bottom row shows the late eye tracking measures. Note that the scale for the early measures has a smaller range.

*Figure 2.* HPD (highest posterior density) intervals for the locality contrasts in Table 3 for Experiments 1–4. Black lines indicate results obtained from data collected using the Grodner & Gibson materials, grey lines indicate results obtained from data collected using the materials composed of short, high-frequency words. HPD intervals that do not include zero, indicating a conventionally reliable non-zero coefficient estimate for the contrast, appear as solid lines.

*Figure 3.* HPD (highest posterior density) intervals for the embedding contrast and interaction contrasts in Table 3 for Experiments 1–4. Black lines indicate results obtained from data collected using the Grodner & Gibson materials, grey lines indicate results obtained from data collected using the materials composed of short, high-frequency words. HPD intervals that do not include zero, indicating a conventionally reliable non-zero coefficient estimate for the contrast, appear as solid lines.

*Figure 4.* Coordination of visual and linguistic processes, with optimal timing of saccade-triggering control signal based on partial completion of lexical processing.

*Figure 5.* Wasted slack time due to late control signal.

*Figure 6.* Overlapping processing of two words, due to early saccade-triggering control signal.

*Figure 7.* Coordination of visual and linguistic processes, with optimal timing of saccade triggering signal based on partial completion of higher-level structural processing.

*Figure 8.* Mean RT from each region: Experiment 1.

*Figure 9.* Mean FPRT from each region: Experiment 2.

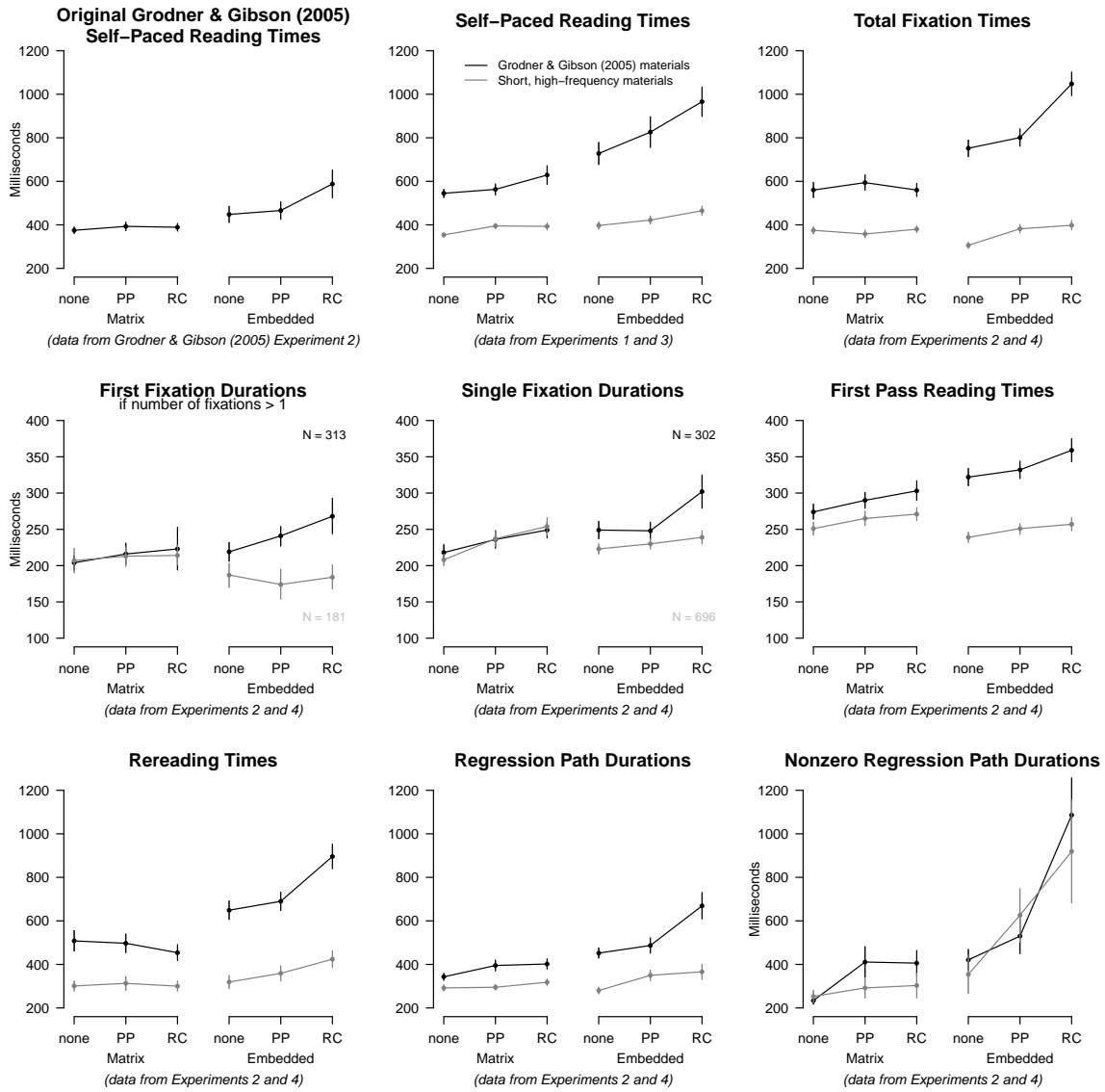
*Figure 10.* Mean TFT from each region: Experiment 2.

*Figure 11.* Mean RT from each region: Experiment 3.

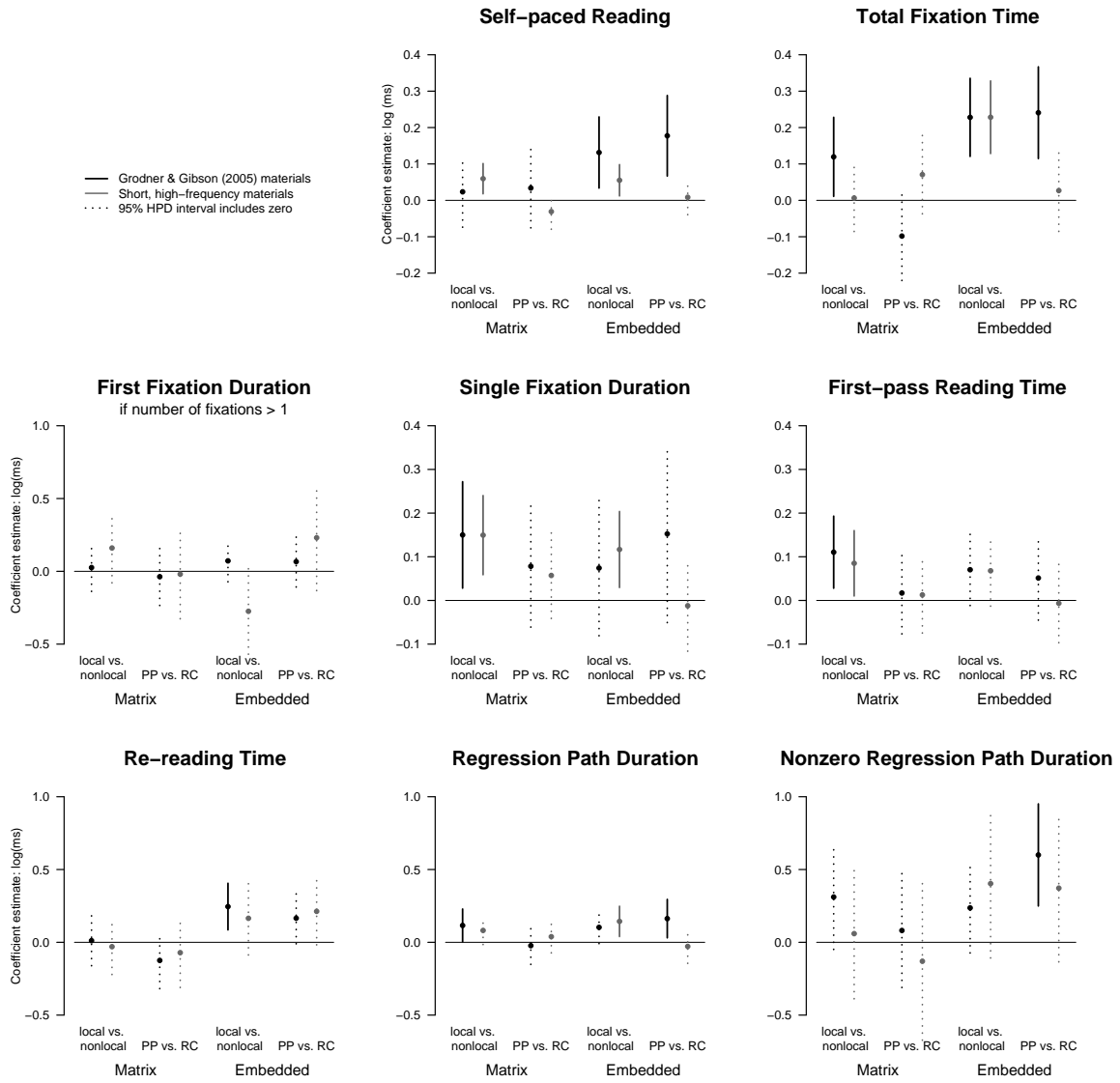
*Figure 12.* Mean FPRT from each region: Experiment 4.

*Figure 13.* Mean TFT from each region: Experiment 4.

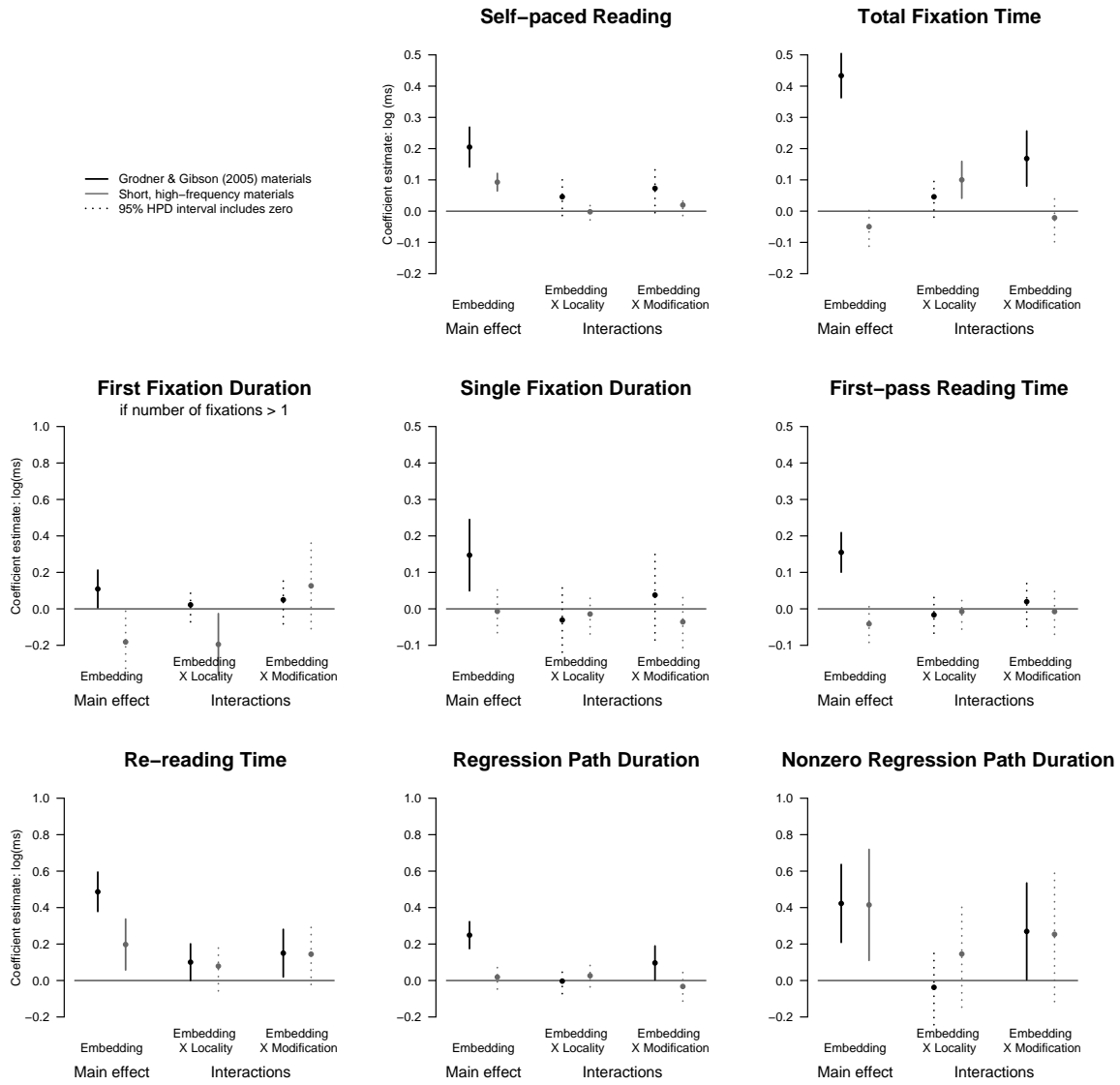
In search of on-line locality effects, Figure 1



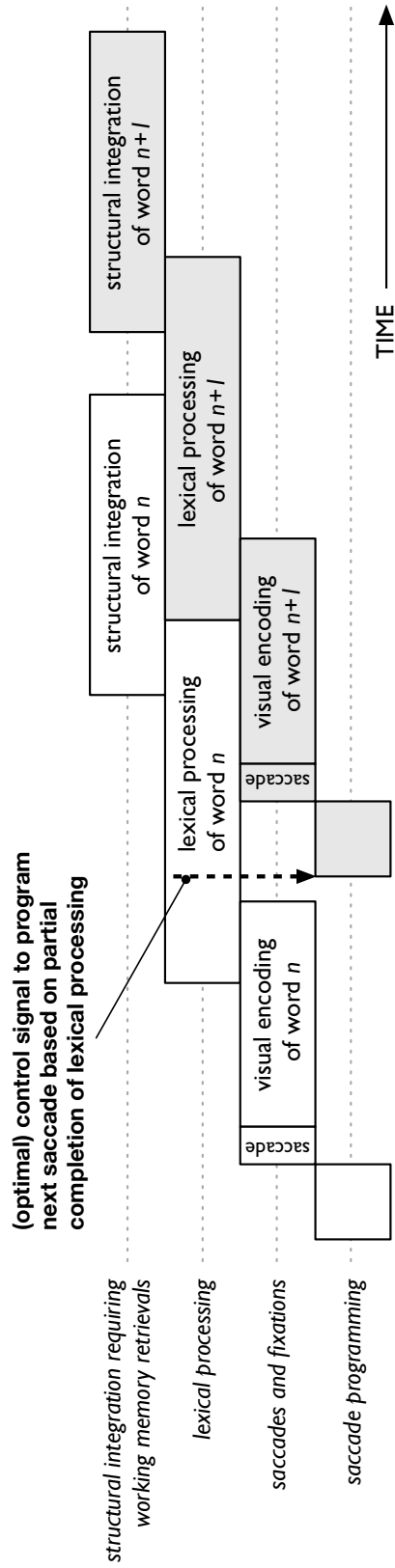
In search of on-line locality effects, Figure 2



In search of on-line locality effects, Figure 3

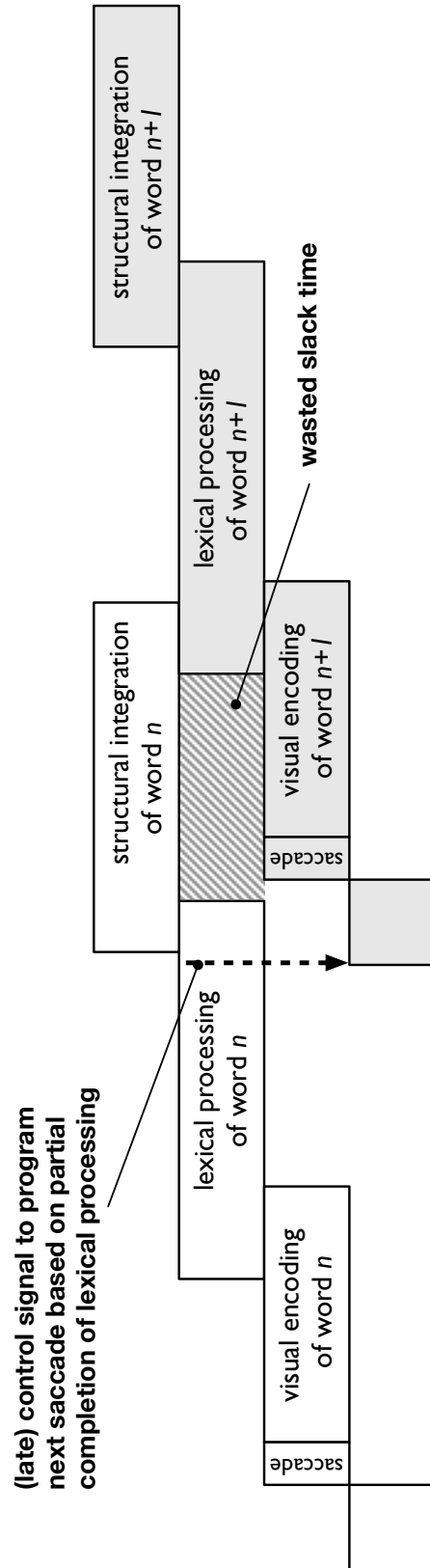


In search of on-line locality effects, Figure 4

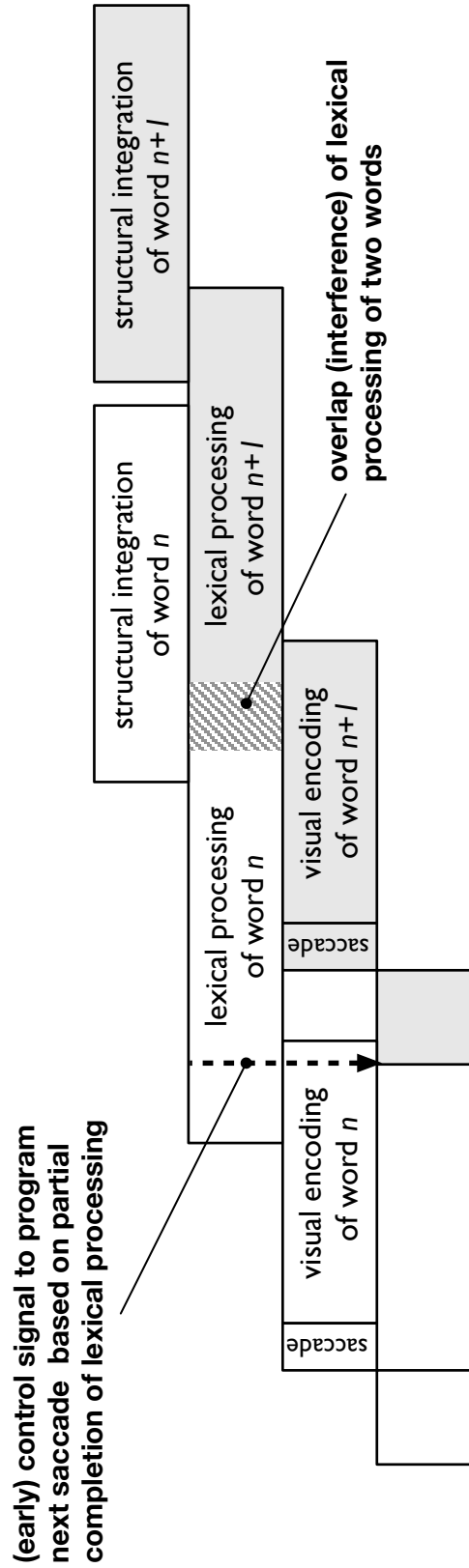


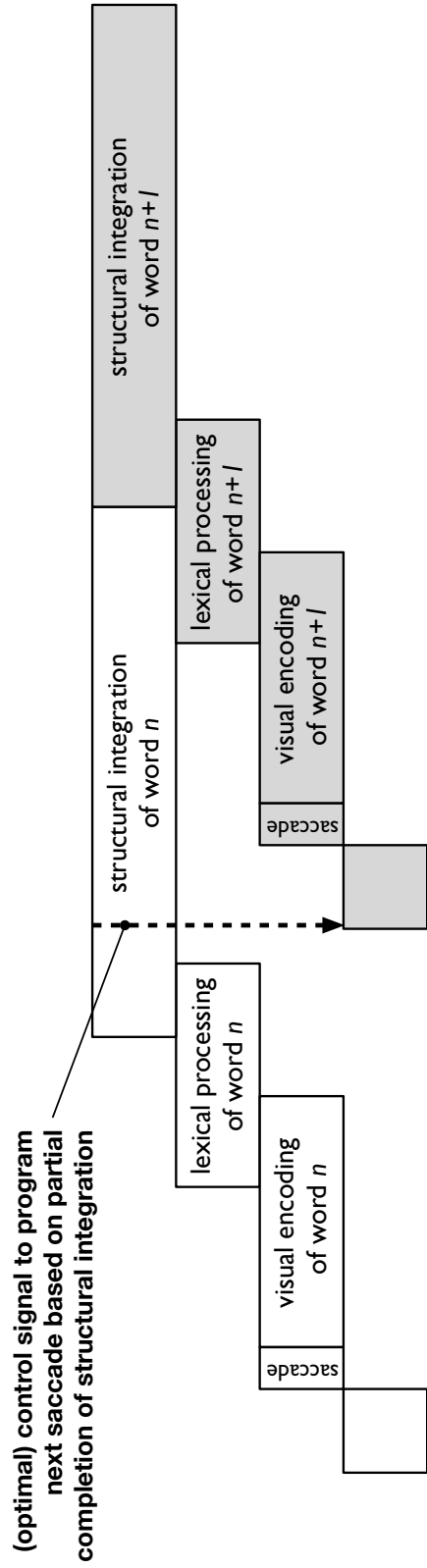


In search of on-line locality effects, Figure 5

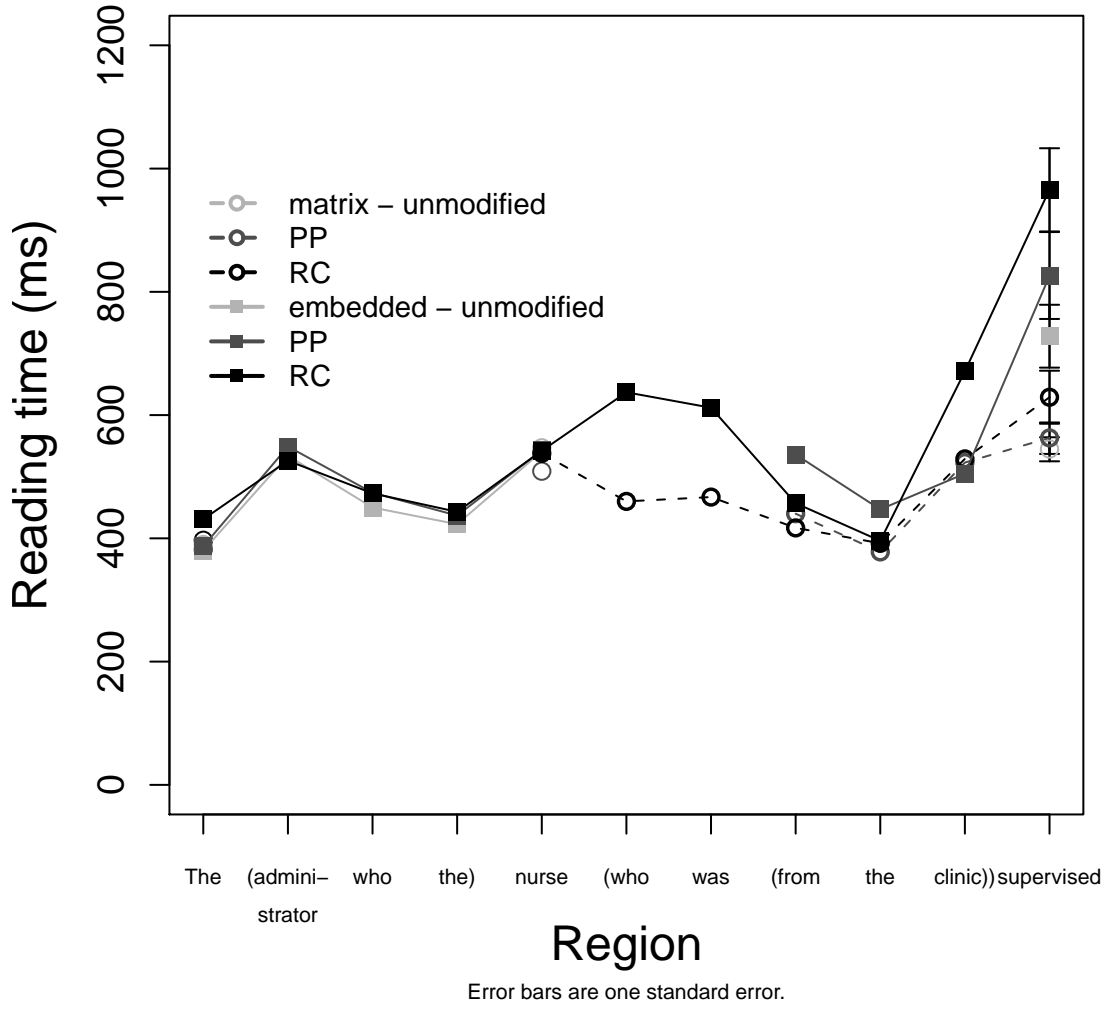


In search of on-line locality effects, Figure 6

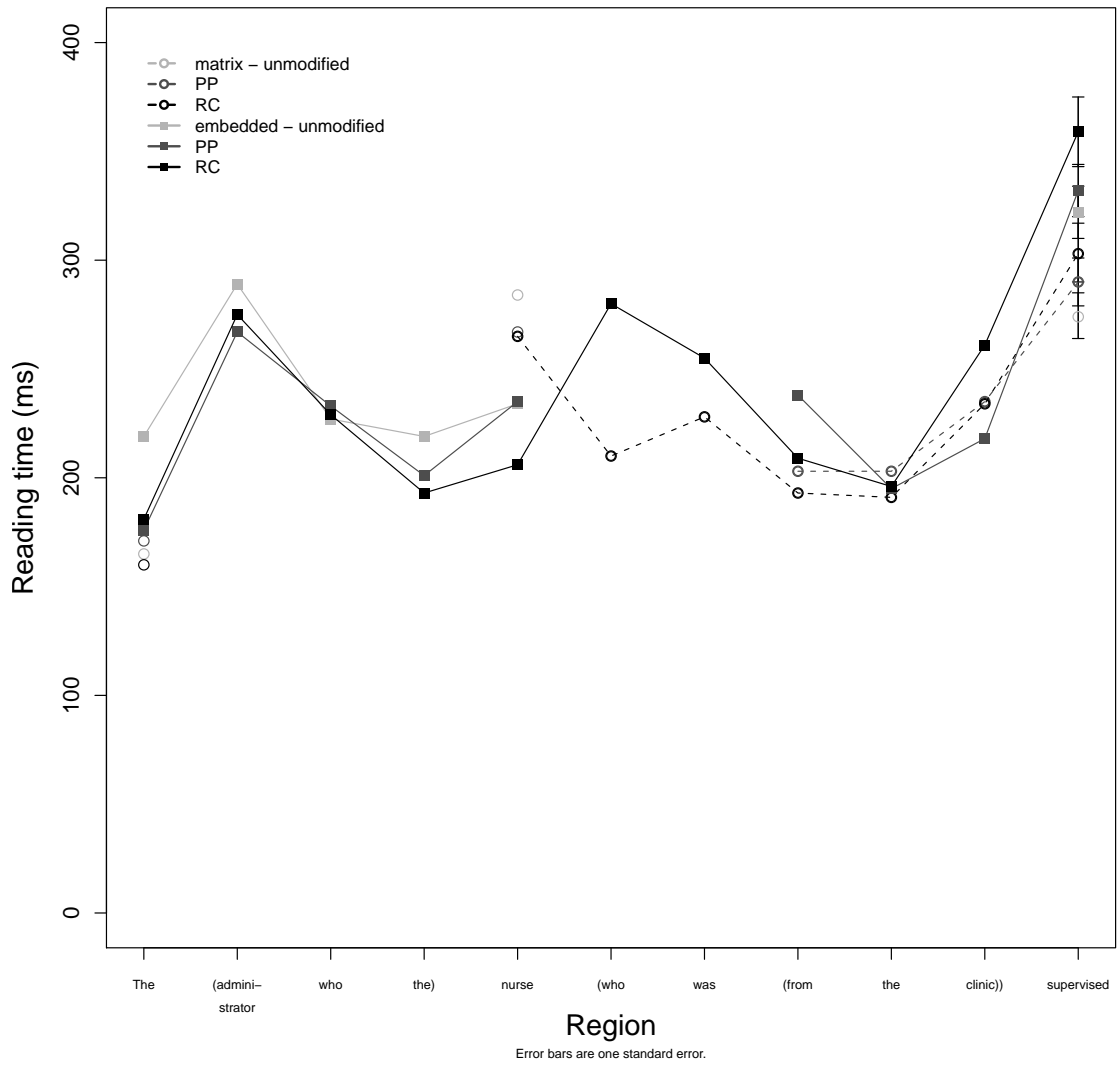




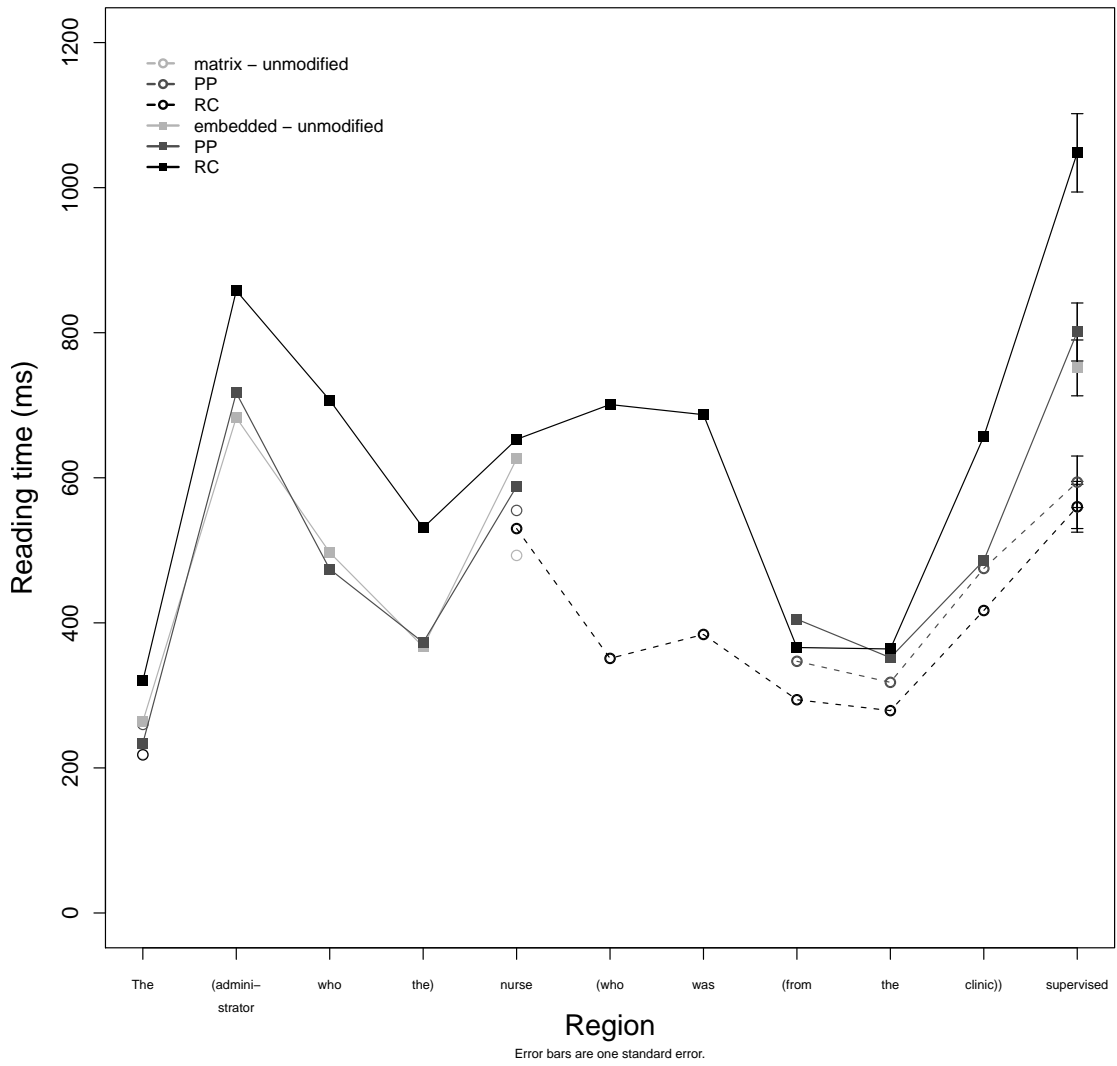
### Experiment 1 self-paced reading times



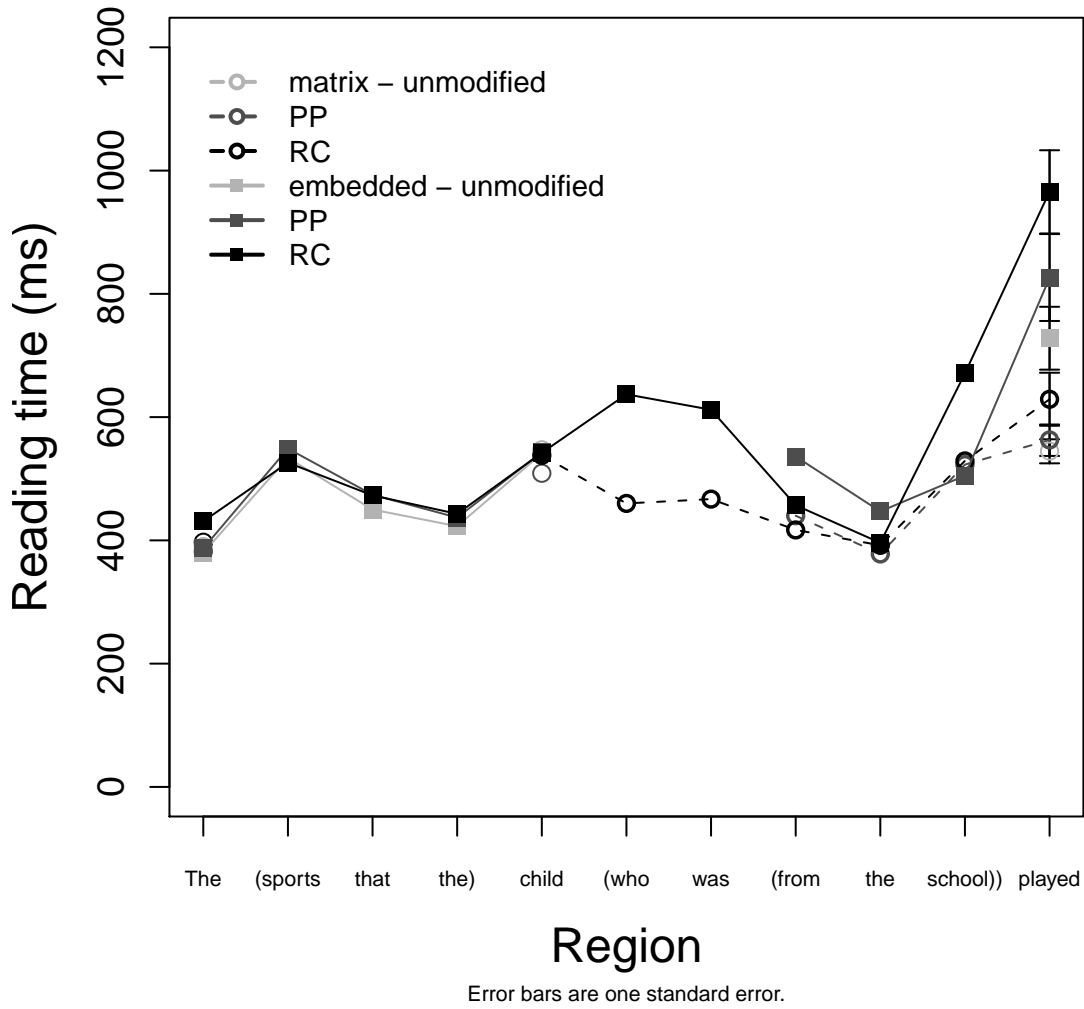
### Experiment 2 first-pass reading times



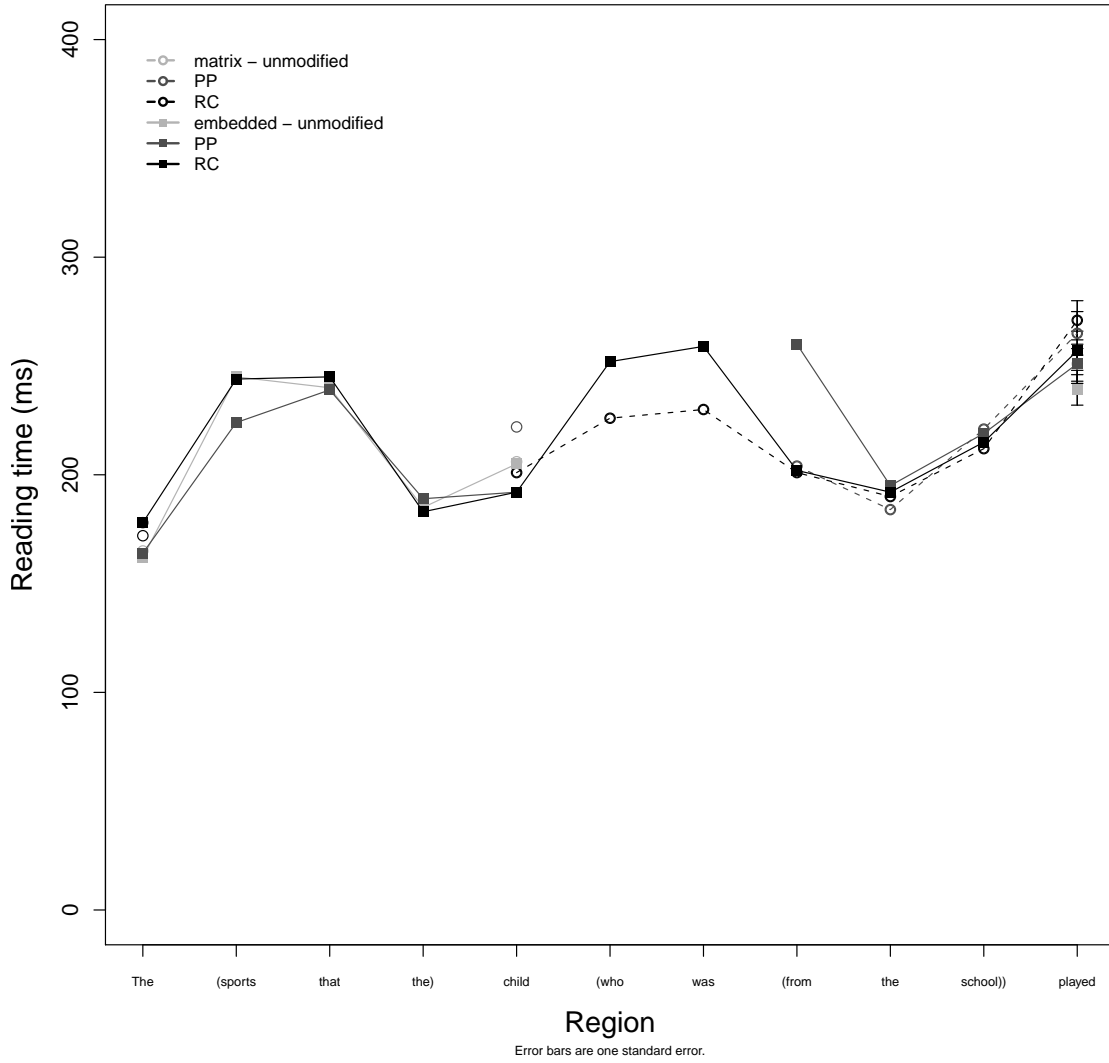
### Experiment 2 total fixation times



### Experiment 3 self-paced reading times



### Experiment 4 first-pass reading times





### Experiment 4 total fixation times

