S Vasishth <vasishth.shravan@gmail.com>

Fwd: Manuscript Submitted - JEP: Learning, Memory, and Cognition

1 message

Brian Bartek <bartek@umich.edu> **Tue, Oct 26, 2010 at 2:34 AM**
To: vasishth.shravan@gmail.com, "Richard L. Lewis" <rickl@umich.edu>

Begin forwarded message:

> **From:** jeplmc@rice.edu
> **Date:** October 25, 2010 11:12:03 AM GMT-04:00
> **To:** bartek@umich.edu
> **Subject: Manuscript Submitted - JEP: Learning, Memory, and Cognition**
>
> Dear Dr. Bartek-
>
> We are in receipt of your revised manuscript titled In search of on-line locality effects in sentence comprehension. You should receive notice of the status of your manuscript within 90 days. During this period, you should not submit your manuscript to another journal. If your manuscript is accepted for publication, you will be required to transfer your copyright to APA, provide full disclosure of any conflict of interest, and certify compliance with APA ethical principles.
>
> Your manuscript number is 2009-1682-R. If you should have any questions regarding your manuscript, please email me directly at jeplmc@rice.edu.
>
> To check the history and status of your manuscript, please visit http://www.jbo.com/jbo3/dsp_checkhistory.cfm?journal_code=xlm2 and enter your last name as the username and 2009-1682-R as the password.
>
> Please also read the APA's Open Letter to Authors located at http://www.jbo.com/jbo3/OpenLetterToAuthors.doc
>
> Sincerely,
>
> Emma Tsai
> Editorial Associate
>
> Journal of Experimental Psychology: Learning, Memory, & Cognition
> jeplmc@rice.edu

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
Stanford University, Stanford, CA

Draft of October 24, 2010

Abstract

Many comprehension theories assert that increasing the distance between elements participating in a linguistic relation (e.g., a verb and an NP argument) increases the difficulty of establishing that relation during on-line comprehension. Such *locality effects* are expected to increase 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, 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 relation. 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, and are of much shorter duration than previously reported effects. To account for the observed empirical patterns we outline a processing model of the *adaptive control* of button pressing and eye-movements. This model makes progress toward the goal of eliminating linking assumptions between memory constructs and empirical measures, in favor of explicit theories of the coordinated control of motor responses and parsing.

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 <u>manager</u> unexpectedly <u>quit</u> her job yesterday.
b. The <u>manager</u> who the supervisor admired unexpectedly <u>quit</u> 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. We refer to this general class of effects on reading times as *locality effects*. While this paper presents evidence that locality effects are consistent with memory-based parsing theories, we call them locality effects without intending to associate them exclusively with the details of DLT or any other specific parsing model. Locality effects are important and relevant to a very broad range of extant memory and parsing theories (see Lewis et al. (2006) or Gibson (2000) 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 sentence processing. 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 eyetracking measures, which we will show has advantages over self-paced reading (SPR) for investigating locality effects. Furthermore, we adopt an approach of running identical materials in both paradigms. This facilitates 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 eyetracking as empirical measures). We sketch the beginnings of such a theory in the main discussion of this paper.

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.

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 eyetracking experiments. Finally, we discuss the theoretical and methodological implications of the results in the General Discussion, concluding with an outline of a theory of the adaptive control of eye-movements and button pressing that provides a framework for understanding the effects of underlying memory processes on the observable measures used in reading studies.

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 (as we summarize below), but *anti-locality* 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, 2003; Nakatani & Gibson, 2008), as well as English (Levy, 2008, reporting an unpublished experiment by Jaeger and colleagues). Although anti-locality 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 locality and anti-locality effects; 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. In other work we have outlined a theoretical model that provides an integrated explanation of both locality and anti-locality (Vasishth & Lewis, 2006; Lewis et al., 2006; Lewis & Vasishth, 2005).

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, 1999; Altmann, Nice, Garnham, & Henstra, 1998) and in garden path reanalysis (garden paths involving longer ambiguous regions are typically more difficult to recover from; Pritchett, 1992; Gibson, 1991; Van Dyke & Lewis, 2003; Ferreira & Henderson, 1991). 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 (putatively) globally unambiguous structures. (In the General Discussion we take up the issue of possible local ambiguity in our materials in some detail).

Table 1: Extant experimental evidence for locality effects in (relatively) unambiguous structures.

Source	Linguistic structures	Methodology
Gibson (1998), Exp. 1	subject- and object-relative clauses	self-paced reading
Grodner et al. (2002), Exps. 1 & 2	reduced-relative ambiguities	self-paced reading
Gibson & Warren (2004), Exp. 1	extraction across VP or NP	self-paced reading
Grodner & Gibson (2005), Exp. 1 & 2	subject- and object-relative clauses	self-paced reading
Wu & Gibson (2008), Exp. 1	subject- and object-relative clauses	self-paced reading

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

Table 1 provides an overview of the existing experimental evidence for locality effects in relatively unambiguous structures. The evidence is restricted to English (a cross-linguistic gap that we do not fill in this paper), and to points of extraction—more specifically, to relations conventionally analyzed as $\overline{A}$-movement (of an argument) from an argument position to a non-argument position (Mahajan, 1990). It has been speculated in Grodner and Gibson (2005, p. 284) and elsewhere (Gibson, 2007) that $\overline{A}$-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 structures that require $\overline{A}$-movement, such as object-extracted relative clauses. 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 augmented transition network (ATN) model of Wanner and Maratsos (1978), and continuing with the Grodzinsky (2000) theory of function in Broca's area.

Prior experiments that could have determined if locality effects generalize beyond object-relatives, and 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, $\overline{A}$ movement occurs when the object is moved from its base position (adjacent to 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 ∅ symbol denotes the null string (nothing interposed).

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

$$\text{The administrator who the nurse} \left\{ \begin{array}{c} \emptyset \\ \text{from the clinic} \\ \text{who was from the clinic} \end{array} \right\} \text{supervised}\ldots$$

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,[1] 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.

But the reliable locality effect observed in (2) may have been driven entirely by the sharp increase in reading times for condition (2c): a case of double center-embedding of relative clauses, an effect that can be explained in ways that have nothing to do with locality (e.g. similarity-basd interference (Lewis & Vasishth, 2005)). How can we be sure that the observed effects in (3) generalize beyond object extractions over embedded relative clauses? 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 …

[1] 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.

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*

$$\text{The nurse} \left\{ \begin{array}{c} \emptyset \\ \text{from the clinic} \\ \text{who was from the clinic} \end{array} \right\} \text{supervised the administrator...}$$

If a locality effect is observed at *supervised* in (5), this would provide evidence that dependencies that are *not* the result of $\overline{A}$-movement relations are also subject to locality effects.[2] In other words, the presence of such effects in both kinds of structures would mean that increasing locality increases the processing cost of resolving simple subject-verb dependencies as well as object extractions. 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 extend the empirical base of locality effects beyond relative clauses. Unfortunately, they do not. Separate 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.[3] In short, it is quite possible that the locality effects are driven by independent sources of difficulty resulting from embedding the verb and from center-embedding the relative clauses.[4]

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 explore locality effects, 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

[2]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.

[3]The possibility of a spillover effect from the preceding was not taken into account in the Grodner and Gibson (2005) study; we address this in the analysis of the new data presented here.

[4]Grodner and Gibson (2005) also reported a linear regression analysis of the relation between reading times and integration cost (the DLT locality metric), but that analysis does not provide independent evidence for possible locality effects in the simple conditions of (5), because it includes data points from all the conditions.

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 increased reading times.

This difference between SPR and eyetracking turns out to be crucial for interpreting SPR reading time data such as that in Grodner and Gibson (2005). 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 caused by the center-embedding. More specifically, the observed 125–150ms increase in reading time may not be due to longer integration or memory processes affected by locality, but primarily recovery processes—perhaps covert rehearsal—triggered by retrieval *failures* (not reanalysis of ambiguity). To anticipate one of the findings reported in this paper: the combined results of our experiments provide support for this interpretation of existing SPR locality effects.

Why does it matter whether observed effects are associated with recovery or initial retrieval or integration? It matters for the purpose of building a cumulative quantitative base of results on which to build computational theories of the underlying memory processes. We should, in principle, be able to use the empirical results from reading studies along with our developing models of memory in parsing to converge on stable estimates of memory retrieval processing rates that may be meaningfully compared (and combined with) processing rate estimates obtained through other methodologies, such as speed-accuracy-tradeoff paradigms (McElree, Foraker, & Dyer, 2003). Such quantitative integration is important not simply because we desire quantitative predictions but because it facilitates theoretical integration.

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 $\overline{A}$ movement or interference between multiple arguments, and observing locality effects under conditions of relatively easy processing. We employ 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 eyetracking 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 measures (if they exist) vs. late measures 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 observe the early manifestations of locality.

4. In the new set of stimuli, we use only inanimate nouns in the extracted object position. As described above, both the subject and extracted object in the original Grodner and Gibson (2005) materials were noun phrases referring to humans. Thus in addition to increasing locality, the embedding manipulation also potentially increased 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 eyetracking 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 naïve 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* 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 $\overline{A}$-movement), and the rest of the sentence became an RC modifying that subject. A clausal connective always followed the matrix object.

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

Condition		Example
Matrix	*Unmodified*	The nurse supervised the administrator while...
	PP-modified	The nurse from the clinic supervised the administrator while...
	RC-modified	The nurse who was from the clinic supervised the administrator while...
Embedded	*Unmodified*	The administrator who the nurse supervised scolded the medic while...
	PP-modified	The administrator who the nurse from the clinic supervised scolded the medic while...
	RC-modified	The administrator who the nurse who was from the clinic supervised scolded the medic while...

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.

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 spacebar 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 spacebar 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 spacebar during sentence presentation was used as a reaction time measure 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% confidence interval is usually (incorrectly) interpreted. For details on how the HPD intervals are computed, see Gelman and Hill (2007); for an accessible description of posterior density estimates, see Kruschke (2010).

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, yielding approximately normal distributions.

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.[5] 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

[5]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 $[\ -\frac{2}{3}\ \frac{1}{3}\ \frac{1}{3}\ 0\ 0\ 0\ ]$.

Table 3: Two sets of contrasts used in the linear mixed models to analyze 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.

		Matrix			Embedded		
	Contrast	∅	pp	rc	∅	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 × embedding interaction	0.75	-0.38	-0.38	-0.75	0.38	0.38
	Modification type × embedding interaction	0.00	-0.50	0.50	0.00	0.50	-0.50

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.[6] This left twenty-eight experimental items.

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 linear mixed models that include length and log 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).[7]

[6] 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. Both design errors were present in the original Grodner and Gibson study.

[7] Reading times for all regions through the critical verb are shown for each experiment in the Appendix. These plots show that difficulty increases at the onset of an embedded relative clause and at the critical verb; and that the difficulty of the critical verb relative to preceding parts of the sentence is reduced in Experiments 3 and 4, where lexical processing is easier. Since they show no other systematic trends and do

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 eyetracking 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 eyetracking 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

not change our interpretation of the results, these plots are not discussed further.

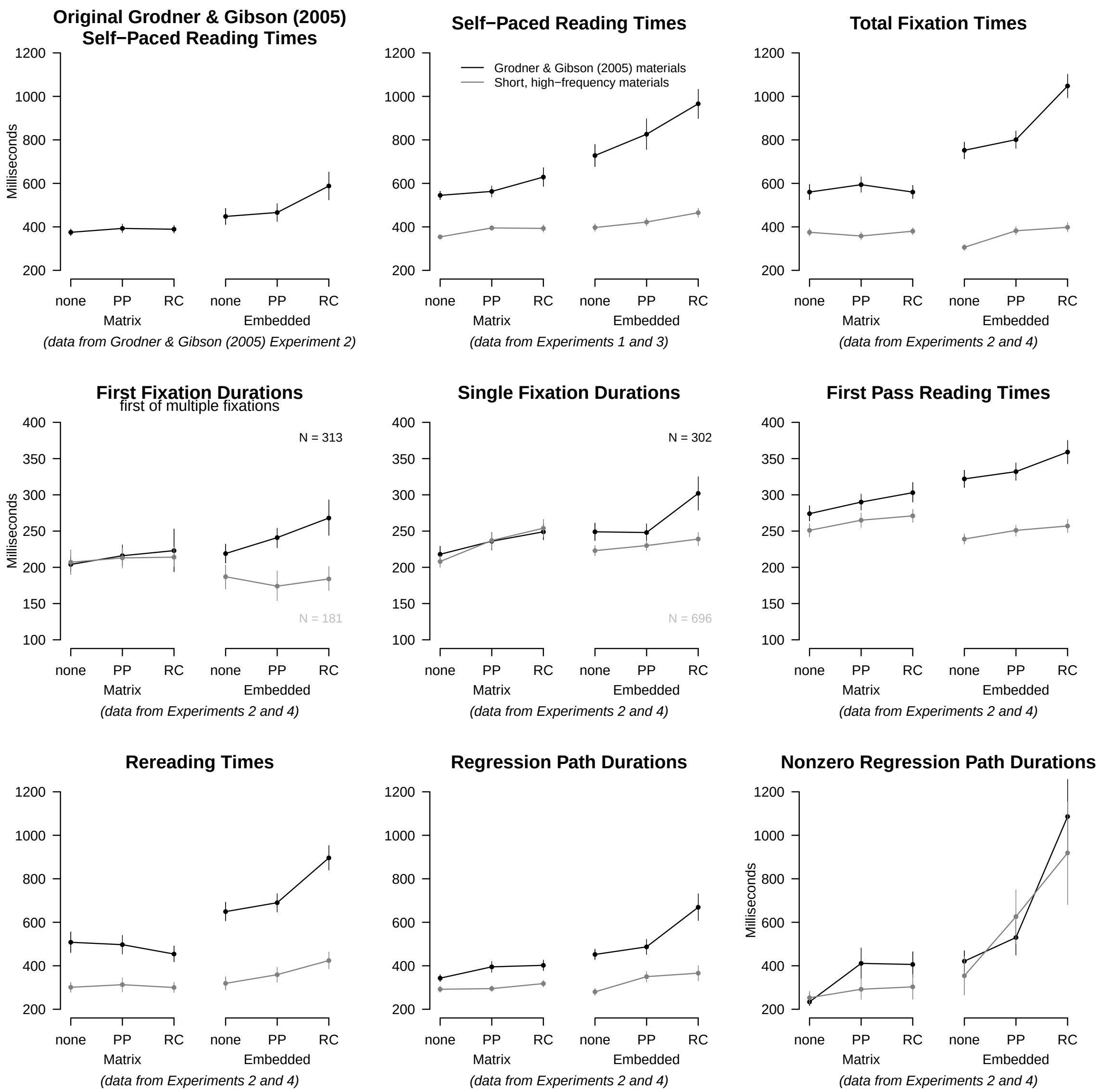

Figure 1. Reading time measures from Experiments 1–4 and the original Grodner & Gibson (2005) self-paced reading study. Error bars are one standard error around condition means. Black lines indicate data collected using the Grodner & Gibson materials; grey lines indicate data collected using the materials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

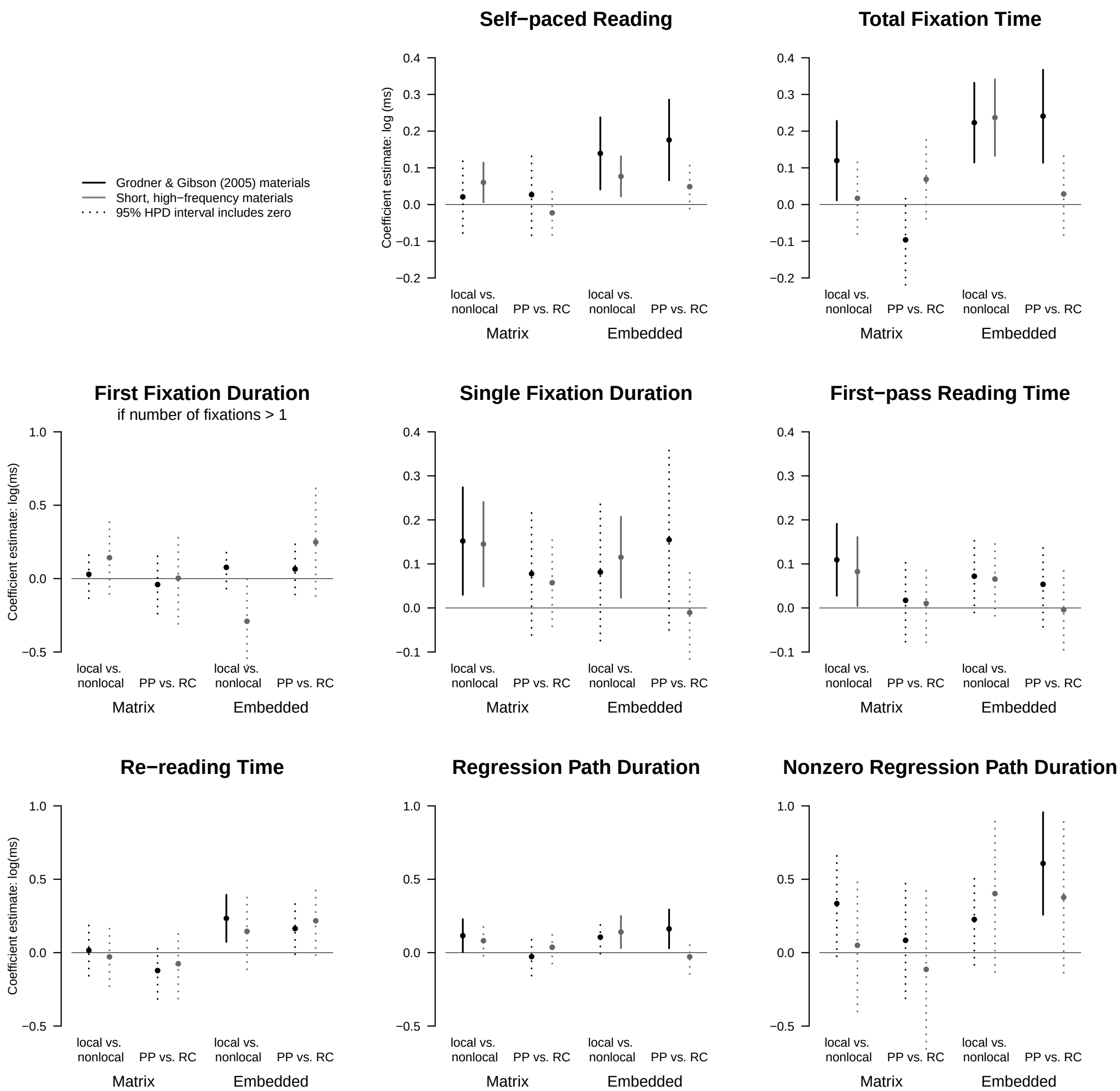

Figure 2. HPD (highest posterior density) intervals for the locality contrasts in Table 3 for Experiments 1–4. Black lines indicate results obtained from data collected using the Grodner & Gibson materials, grey lines indicate results obtained from data collected using the materials composed of short, 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.

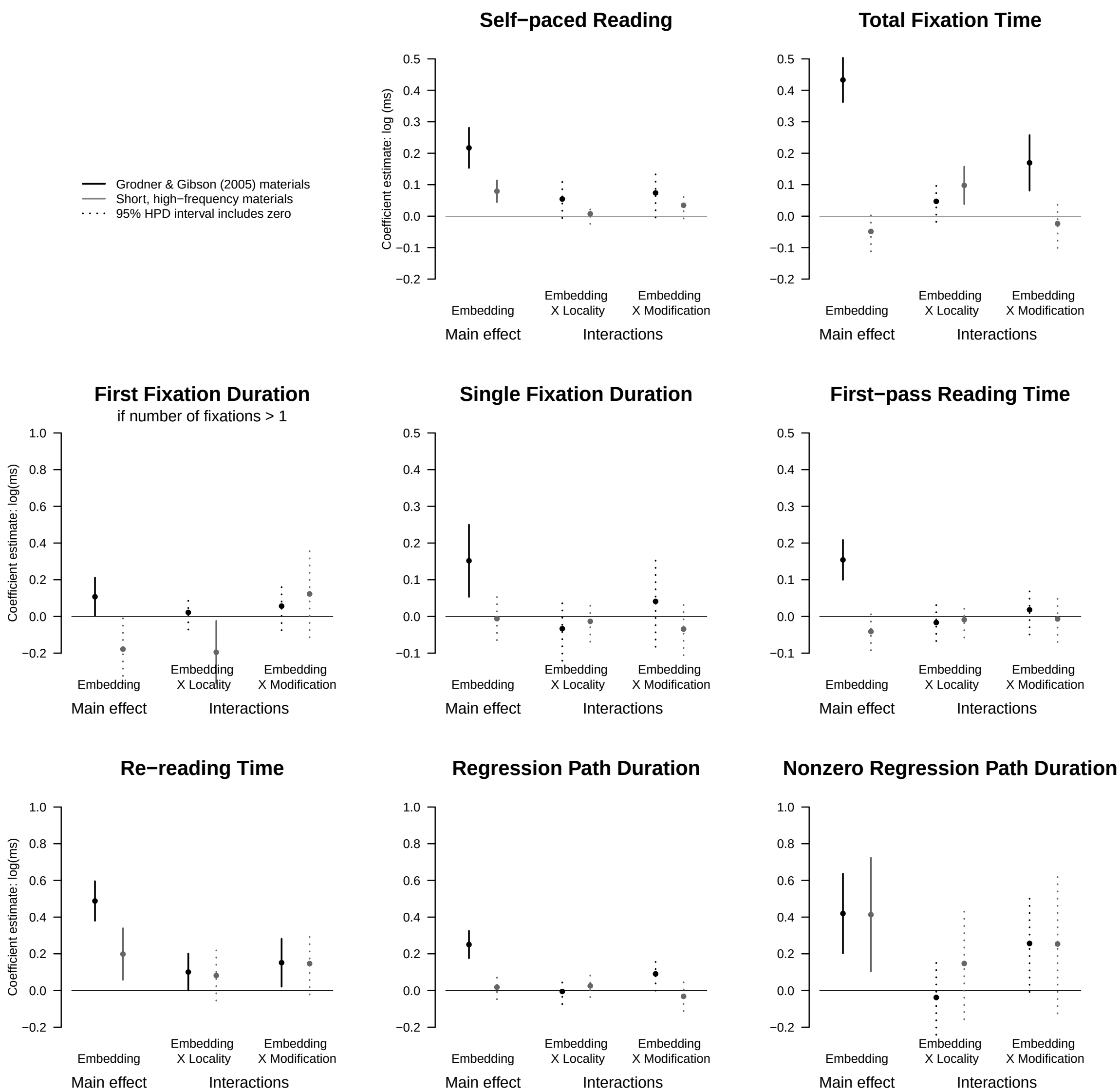

Figure 3. HPD (highest posterior density) intervals for the embedding contrast and interaction contrasts in Table 3 for Experiments 1–4. Black lines indicate results obtained from data collected using the Grodner & Gibson materials, grey lines indicate results obtained from data collected using the materials 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.

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. The interaction of locality and embedding was marginally significant.

These results are thus ambiguous concerning the nature of locality effects in the ways detailed above in the analysis of the Grodner and Gibson (2005) results. The observed locality effects in the embedded verb conditions may be directly related to the increased distance between the subject and verb, but they could also be explained by retrieval interference between the two relative pronouns (*who*) in the embedded RC conditions, by interference between the object (*administrator*) and subject (*nurse*), or by other sources of difficulty related to center-embedding and object-relative extraction. The matrix conditions do not help to disambiguate the results of the embedded conditions. No locality effects were found in these simpler sentences. It is of course possible that locality effects are present but harder to detect in the simpler sentences due to other sources of variance in the materials or methodological limitations of self-paced reading

For present purposes, Experiment 1 serves the dual role of providing further motivation for the eyetracking and materials 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 remaining experiments.

Experiment 2: Eyetracking version of Experiment 1

Experiment 2 was an eyetracking 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 payment.

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 spacebar 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.

Table 4: Definitions of the eyetracking measures used in the analysis of Experiments 2 and 4.

Measure	Definition
First Fixation Duration	Time between the initial landing in a region and the beginning of the first saccade out of the region; excludes trials where only one fixation was made.
Single Fixation Duration	Time spent fixating a region when only one fixation was made therein.
First-Pass Reading Time	The summed duration of all fixations made within a region before exiting to the right or left.
Regression Path Duration	The sum of all fixations within a region n and in any regions to its left before fixating to the right of n.
Non-zero Regression Path Duration	Identical to Regression Path Duration, but Non-Zero Regression Path Duration excludes cases where no regressions occurred.
Re-reading Time	The sum of all fixations in a region excluding first-pass reading time. Re-reading analyses include zero-millisecond re-reading times.
Last-Pass Reading Time	The sum of all fixations in the last run of fixations within a region.
Total Fixation Time	The sum of all fixations within a region during a trial.

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. Linear mixed 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 eyetracking (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 locality patterns in both early and late fixations. The first such result 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. The PP vs. RC contrast was not reliable for the matrix condition, but there was a consistent trend of greater reading times in the RC conditions across all the early measures.

The second interesting result from Experiment 2 is that only later measures (Re-Reading, Regression Path Duration and Non-zero Regression Path Duration) mirror 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 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.

There were no reliable locality effects found in First Fixation Duration (in either Experiment 2 or 4). 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.

Interim summary and motivations for Experiments 3 & 4

Experiment 1 replicated the results of Grodner and Gibson (2005), and provided a baseline for evaluating the relationship between SPR and eyetracking measures. The results of Experiment 2 are important for two reasons. First, the effects observed in the simple matrix conditions in Experiment 2 provide the first on-line evidence of locality effects in non-extraction structures, suggesting that locality effects are *not* restricted to complex movement structures, and that they do not rely on interference between possible retrieval targets or between multiple relative pronouns.

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 processes that are reflected in first-pass measures.

The primary goal for both Experiments 3 and 4 is to increase our ability to detect locality effects across the conditions, and especially in the early eye-movement measures. We attempt to do this by minimizing overall comprehension difficulty, especially the difficulty associated with the embedded conditions.

Two lexical changes to the materials were made to accomplish these aims while maintaining the structure of the six conditions:

1. All words prior to and including the critical verb are restricted to short (3–6 letters), high-frequency (greater than fifty occurrences per million) words (Table 6).

2. The object of the critical verb, which was always an animate, human referent in the original materials, was made uniformly inanimate in the new materials. This change is expected to make processing of the embedded conditions easier in two ways. First, inanimate referents in the object position should reduce retrieval interference at the verb. Second, using inanimate referents as object may ease processing at the verb by biasing the reader towards an object-relative reading.

This manipulation to increase the bias toward the object relative reading is important because experience-based parsing theories predict local comprehension difficulty at points where new input signals a relatively unlikely continuation of the sentence (see Gennari and MacDonald (2008) for a summary). In particular, the constraint-satisfaction account of Gennari and MacDonald (2008, 2009) predicts difficulty in the embedded structures of our Experiments 1 and 2 on this basis. These studies demonstrate that object-relatives beginning with an animate head noun like *administrator* are difficult to comprehend because the parser learns that structures other than object relatives are more likely to follow in such contexts (such as passives, e.g. *The administrator who the nurse was supervised by ...*). Encountering the verb *supervised* rules out more likely parses in favor of the unexpected object-relative. Thus, the verb creates difficulty by violating the parser's implicit expectations. However, object-relatives are frequently produced in sentences where an inanimate head noun fills the object role (Gennari & MacDonald, 2008), and there is evidence that these constructions are nearly as easy to process as subject-relative clauses (Traxler, Morris, & Seely, 2002).

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 set of short, high frequency words. Our motivations for this manipulation were detailed above.

Methods

Participants

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

Stimuli

Thirty experimental sentences were created for use in a self-paced reading experiment (Experiment 3) and a parallel eyetracking (Experiment 4). The syntactic structure of all sentences was identical to Experiments 1 and 2, and Grodner and Gibson (2005) Experiment

Table 5: Examples sentences from the six conditions for Experiments 3 and 4. The critical verb is underlined.

Condition		Example
Matrix	*Unmodified*	The child played the sports that were hard to master.
	PP-modified	The child from the school played the sports that were hard to master.
	RC-modified	The child who was from the school played the sports that were hard to master.
Embedded	*Unmodified*	The sports that the child played were hard to master.
	PP-modified	The sports that the child from the school played were hard to master.
	RC-modified	The sports that the child who was from the school played were hard to master.

Table 6: Lexical properties of each set of materials, through the critical verb position. The new materials for Experiments 3 & 4 included 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 the participants saw. Frequency counts displayed are occurrences per-million-words in the American National Corpus.

	Critical verb		Content words	
	Exps. 1 & 2	Exps. 3 & 4	Exps. 1 & 2	Exps. 3 & 4
Median length	8.0	4.0	7.00	5.00
Std. deviation	1.6	.91	2.56	.97
Median frequency	5.0	112.0	12.50	77.0
Std. deviation	13.2	166.3	53.10	88.78

2, but content words were restricted to 3–6 letter words that had a frequency higher than fifty occurrences per-million-words in the First Release of the American National Corpus.[8] A comparison of the relevant lexical properties of the new and old materials is given in Table 6.

Table 5 gives examples of the materials. 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.

Plausibility norming

In these materials locality is manipulated via nominal modifications that unavoidably change the semantic content of the sentences. To control for possible plausibility effects that may be confounded with the locality manipulations, we conducted a separate norming study with 57 participants from the same population who did not participate in the reading

[8] http://www.americannationalcorpus.org/FirstRelease/

Table 7: Mean plausibility ratings on a 5-point scale 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

experiments themselves. Participants read each experimental item at one level of subject-modification, distributed randomly among 54 filler sentences, and rated plausibility on a 5-point Likert scale. Table 8 provides the mean ratings for each level of modification.

To test whether dependency locality predicted plausibility ratings, a linear mixed model including two orthogonal locality contrasts was run. One contrast tested the unmodified-subject condition against both types of subject modification; the other tested PP modification against RC modification. Both contrasts were significant (HPD: local vsňon-local (-0.45,-0.56); PP vs. RC (0.21, 0.45). Although there are plausibility differences, they are relatively small and we control for their effects on reading times in all the subsequent analyses by including item-level plausibility predictors in the mixed-effect models. None of the results reported below were affected by the inclusion of plausibility as a predictor.

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, top row, middle panel, grey lines). 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 contrasts were larger in the original materials than the new materials. This was established by a linear mixed 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 (min,max)= (-0.107, 0.005)).

Embedding effect and interactions (Figure 3, top row, middle panel, grey lines). 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 interval = (-0.02, 0.03).

Discussion

The most important result of Experiment 3 is the locality effect in the matrix verb conditions. Using short, high-frequency words, locality effects were detected where they were not apparent (in SPR) in Experiment 1. The joint analysis of Experiment 1 and 3 also provide evidence suggesting that locality may interact with overall processing difficulty—here manipulated by lexical processing difficulty.

The empirical goals of this study were thus met: the materials change produced faster overall reading times *and* made it possible to detect a locality effect in the matrix condition. Furthermore, the size of the locality effect in both the matrix and embedded clause condition is comparable. The evidence from Experiment 3 thus supports the tentative conclusion we advanced in Experiment 2: locality effects exist outside of $\overline{A}$-movement and may be detected under conditions of relatively rapid and easy comprehension. Finally, the effects in Experiment 3 cannot be explained by the relative rarity of object-extracted structures with an animate, discourse-new direct object (because these sentences used inanimate objects).

Experiment 4: Eyetracking version of Experiment 3

Experiment 4 was an eyetracking version of Experiment 3. Using shorter lexical items has the further advantage in eyetracking of reducing the number of fixations on individual words (Brysbaert & Vitu, 1998; 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 payment.

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 analyzed using linear mixed 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.[9]

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 (HPD(*min*, *max*): First Fixation (0.04, 0.22); First-Pass Reading

[9] 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.

(0.05, 0.13); Regression Path (0.06, 0.16); Re-Reading (0.15, 0.32); Total Fixation Time: (0.18, 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, a locality effect for the embedded conditions emerged in an early measure (Single Fixation).

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. One possibility is that the embedding effects obtained in this experiment reflect only regressions triggered by retrieval failure.

One aspect of the data pattern in Experiment 4 remains surprising: the absence of a locality effect in Total Fixation Time for the matrix verb conditions. However, we do not believe that this negative result should be taken to mean 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 appears to be a function of the high variance and null-locality effect in the re-reading measures, which contribute to the Total Fixation measure.

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 involve mechanisms affected by locality.

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 review the main conclusions, consider alternative explanations, and outline a theoretical model of how locality effects might arise as features of adaptive policies for controlling eye-movements and button-presses in reading.

The ubiquity and nature of locality effects

There are three 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 may indeed be ubiquitous: they emerge not only in the computation of relatively difficult embedded structures involving $\overline{A}$ movement (as replicated in Experiment 1), but can be detected in the computation of relatively simple subject-verb

relations (as shown for the matrix conditions in Experiments 2–4). Experiment 1 replicated an earlier null finding for the matrix conditions, but Experiments 2–4 consistently showed that locality effects may be detected in those structures using eyetracking (Experiments 2 and 4) and using lexical items designed to ease overall processing.

2. The locality effects obtained in the present experiments appear to be robust against spillover effects and plausibility differences. Locality effects emerged in 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 in fact observed. Including item-level plausibility for Experiments 3 and 4 in the analysis models did not alter the estimates of the locality effects.

3. The largest and most robust effects of locality previously observed in SPR correspond well with the pattern observed in rereading and regression measures in the eyetracking record. This is consistent with our hypothesis that the long SPR times correspond to recovery from short-term retrieval failures during parsing—the effects are large in SPR in part because they include time to recover from failure.

Alternative explanations

We briefly consider here two possible alternative explanations for the observed locality effects: local ambiguity and experience-based accounts.

Local ambiguity explanations. In some of the items in the matrix conditions, there is a temporary attachment ambiguity at the critical verb: the verb may be parsed as either the main verb or the beginning of a reduced relative clause (as in *The child (from the school/who was from the school) played by his friends as a fool ...*). Could this local ambiguity give rise to the locality effects found in our experiments?

Local ambiguity is unlikely to be the source of the locality effects for two reasons. Consider first how the ambiguity might in principle give rise to the effect. In animate-subject contexts such as these, there is an overwhelming bias for a main verb continuation (MacDonald, Perlmutter, & Seidenberg, 1994). The post-nominal modifications could thus give rise to a locality effect if they made the relative clause continuation more likely, producing either greater competition times for a single-path parser or longer reading times associated with pursuing the relative clause structure for a ranked parallel parser. But such post-nominal modifications make the onset of the matrix verb *more likely*, not less likely (Levy, 2008). Put another way, shorter subject phrases are more likely than longer ones (a point we take up again below when considering experience-based approaches).

Second, the ambiguity in question rests on a morphological ambiguity between the active and past-participle form of the verb—an ambiguity that is present in twenty-three of the items in Experiments 3 and 4 (such as *played/played*) but not in seven of the items (such as *wrote/written*). When we analyze the effect of morphological ambiguity in a linear mixed model, we find no interaction between morphological ambiguity and locality.[10]

Experience-based explanations. We consider here how two prominent experience-based theories might account for the observed effects: the *Production-Distribution-Comprehension*

[10] A table of coefficient estimates and their HPD intervals is included in the appendices.

(PDC) Theory of Gennari and MacDonald (2009), and the *surprisal* metric of Hale (2001). The central claim of PDC is that pressures on the production mechanism create distributional regularities in natural language, and comprehension performance is shaped by exposure to these distributional regularities. Thus, a mechanism that created a preference for producing short phrases might result in sentences with the non-local conditions being less probable, and more costly to parse, than the unmodified matrix or embedded condition baseline. The locality effects here are in principle consistent with this account, but it is presently not specified in enough detail to make clear predictions concerning the direction of the effects.

To see why, it is useful to consider an existing experience-based parsing account that is both consistent with the overall PDC theory, and is specified in enough detail to make on-line processing predictions: surprisal (Hale, 2001). Under the surprisal account, a contextual manipulation will make reading time on a word increase to the extent that the manipulation makes the word less likely[11]—a clear and natural assumption of the effect of the probabilistic encoding of experience on reading time that is consistent with PDC. For the materials in the experiments presented here, locality effects would be expected if the post-nominal modification—increasing the length of the subject noun phrase—makes the matrix verb less likely. Working out the precise predictions of surprisal depends upon assumptions about grammar and parsing algorithm, but at least one implementation of surprisal has been shown to predict exactly the opposite pattern (Levy, 2008). The reason is simply that longer noun phrases are less likely than shorter ones, and so the longer the noun phrase, the more likely the matrix verb is to appear.

The point of considering PDC and surprisal together here is not to argue that experience-based theories are unable to account for the observed effects—but simply to observe that, even under the very plausible assumption that we have more experience with shorter rather than longer phrases, additional processing assumptions are required to generate specific reading time predictions that flow from this assumption. And at least one experience-based processing account (surprisal) has been instantiated in a way that does not make the correct predictions for the materials in Experiments 1–4.

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 the adaptive control of eye-movements in reading and button-presses in self-paced reading. We sketch the basic principles of this model here and describe how it yields several interesting predictions concerning the relationship of SPR and eyetracking 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, & Rayner, 2007), but rather follows from a theory of the control of motor responses. We briefly summarize the theory in two parts—assertions about the

[11]See Hale (2001) and Levy (2008) for the precise mathematical formulation of surprisal, which we need not appeal to here.

parsing and memory processes, and assertions about control—and then describe briefly how the theory accounts for empirical patterns in Experiments 1–4. A full exploration of this theory requires extensive computational modeling; what follows is meant to be a suggestive start.

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

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

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).

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 external task structure and utility, and internal cognitive resources and limitations (Howes, Lewis, & Vera, 2009). We assume here that button presses in self-paced reading and eye-movements in reading are also subject to locally adaptive control (Reichle & 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 7 present a simplified illustration of possible control signals coordinating visual and linguistic processes. In each figure there is a different possible timeline associated with reading a pair of words (n and $n+1$) in a sentence; each figure corresponds to a different control signal. 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. Although the figures depict discrete stages, nothing in this account depends on such an assumption. Early signals are favored over late to the extent that they allow for more efficient parallel pipelining of comprehension processes. Late signals

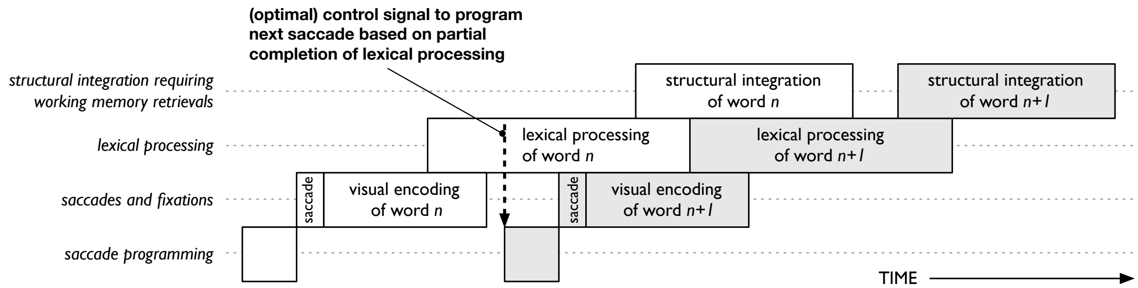

Figure 4. Coordination of visual and linguistic processes, with optimal timing of saccade-triggering control signal based on partial completion of lexical processing.

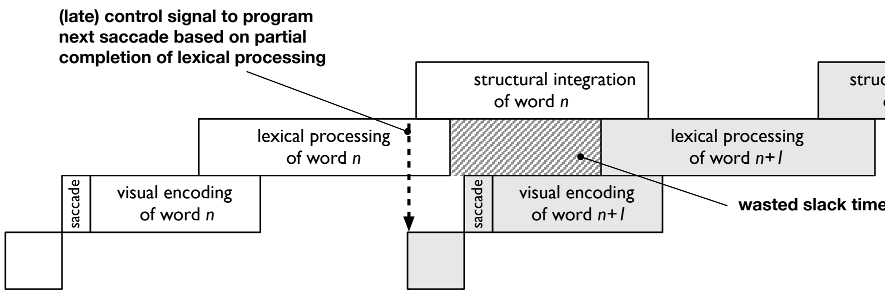

Figure 5. Wasted slack time due to late control signal.

are favored over early to the extent that they ensure sufficient time for comprehension subprocesses 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

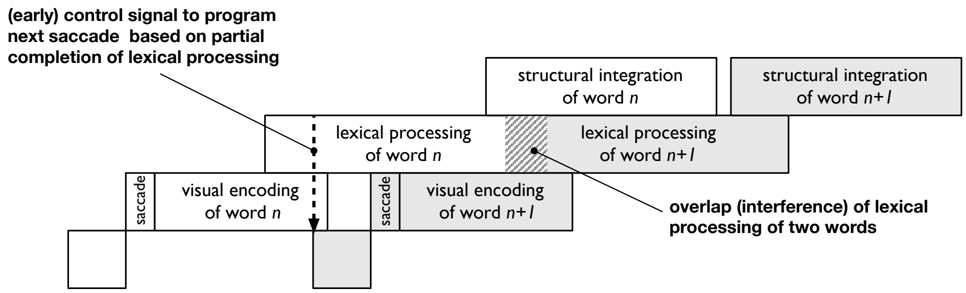

Figure 6. Overlapping processing of two words, due to early saccade-triggering control signal.

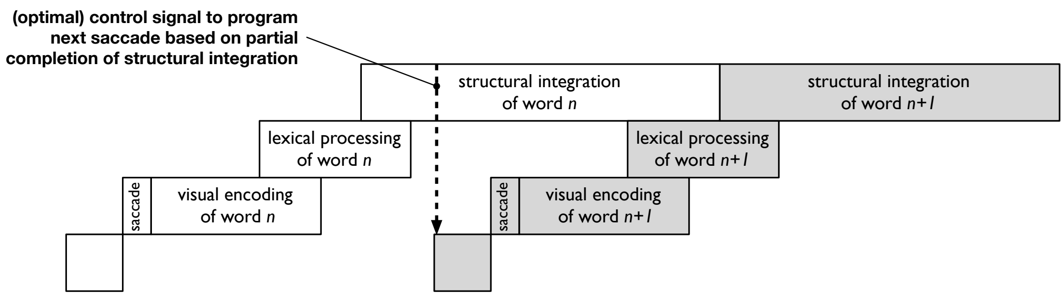

Figure 7. Coordination of visual and linguistic processes, with optimal timing of saccade triggering signal based on partial completion of higher-level structural processing.

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 the key results from Experiments 1–4:

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. It may be difficult to observe such locality 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*; e.g. Schweickert, 1980), 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. 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.

This model thus 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 reflect posited short-term memory retrieval processes in a range consistent with estimates of short-term memory retrieval durations in the general STM literature (McElree, 2006).

This is merely the start of a substantial combined modeling and empirical effort, but it does suggest that there is theoretical payoff in the empirical approach pursued here of combining eyetracking and self-paced reading, and in systematically exploring the effects of lexical manipulations on (relatively) unambiguous structural processing (as has been demonstrated repeatedly for ambiguous processing). Finally, whatever theoretical approach is pursued—whether one that emphasizes experience or memory processes or both—the evidence from the four experiments presented here is relevant because it suggests that locality effects may indeed be a ubiquitous feature of human sentence processing.

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)

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.

Ferreira, F., & Henderson, J. M. (1991). Recovery from mis-analyses of garden-path sentences. *Journal of Memory and Language*, *30*, 725–745.

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. (2008). Semantic indeterminacy in object relative clauses. *Journal of Memory and Language*, *58*, 161–187.

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. (2007). *Locality and anti-locality effects in sentence comprehension.* Leipzig, Germany. (Presented at Workshop on processing head-final languages)

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. (1999). 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 Coginition*, *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.

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics.* Pittsburgh, PA.

Howes, A., Lewis, R. L., & Vera, A. H. (2009). Rational adaptation under task and processing constraints: Implications for testing theories of cognition and action. *Psychological Review*, *116*(4), 717–751.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99(1)*, 122–149.

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.

Kruschke, J. K. (2010). What to believe: Bayesian methods for data analysis. *Trends in Cognitive Sciences*, *14*, 293–300.

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.

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.

MacDonald, M. C., Perlmutter, N. J., & Seidenberg, M. S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101(4)*, 676–703.

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.

McElree, B., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory and Language*, *48*, 67–91.

Nakatani, K., & Gibson, E. (2008). Distinguishing theories of syntactic expectation cost in sentence comprehension: evidence from japanese. *Linguistics*, *46*, 63–86.

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.

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.

Schweickert, R. (1980). Critical-path scheduling of mental processes in a dual task. *Science*, *209*(4457), 704-706.

Traxler, M., Morris, R., & Seely, R. (2002). Processing subject and object relative clauses: evidence from eye movements. *Journal of Memory and Language*, *35*, 69–90.

Trueswell, J., Tanenhaus, M., & Kello, C. (1993). Verb-specific contraints 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 Expermiental 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. (2003). *Working memory in sentence comprehension: Processing Hindi center embeddings.* New York: Garland Press. (Published in the Garland series Outstanding Dissertations in Linguistics, edited by Laurence Horn)

Vasishth, S., & Lewis, R. L. (2006). Argument-head distance and processing complexity: Explaining both locality and antilocality effects. *Language*, *82*, 767–794.

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.

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 Carolina.

Plots of reading time in all regions

Figure 8. Mean reading time from each region: Experiment 1.

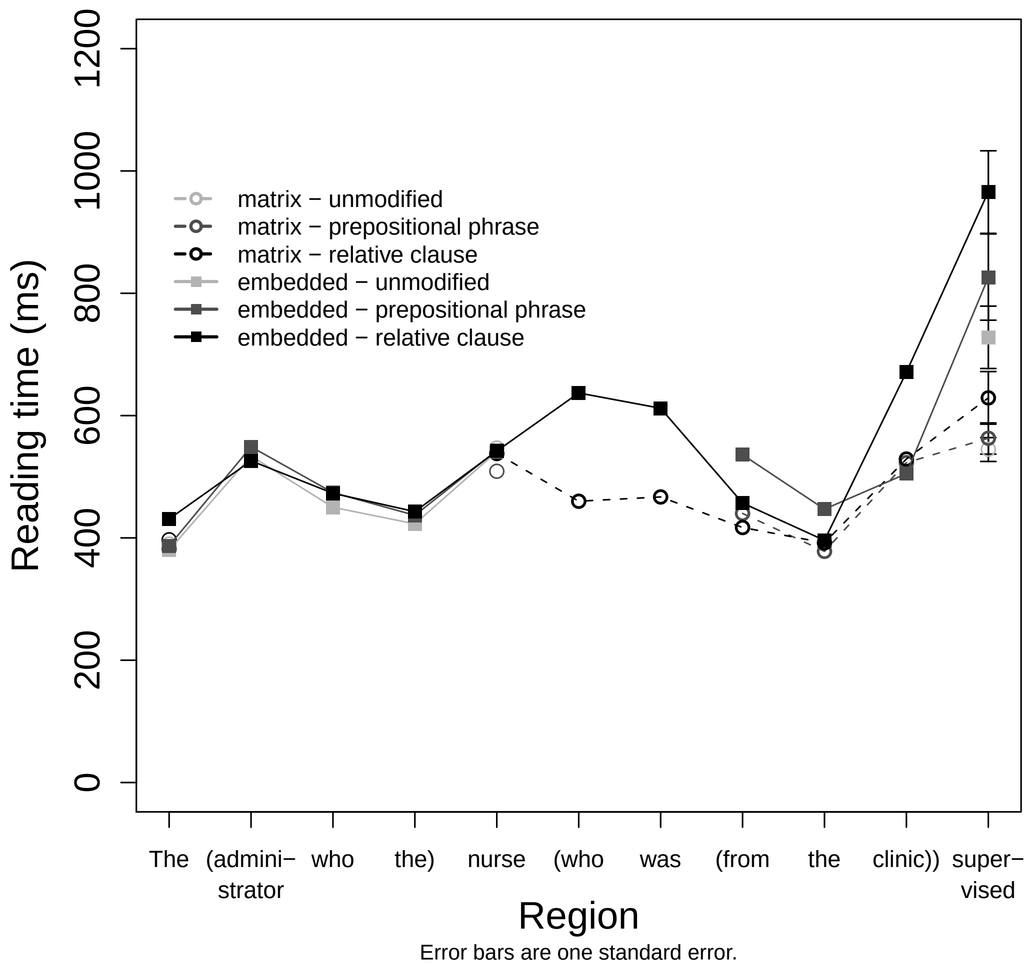

Figure 9. Mean first-pass reading time from each region: Experiment 2.

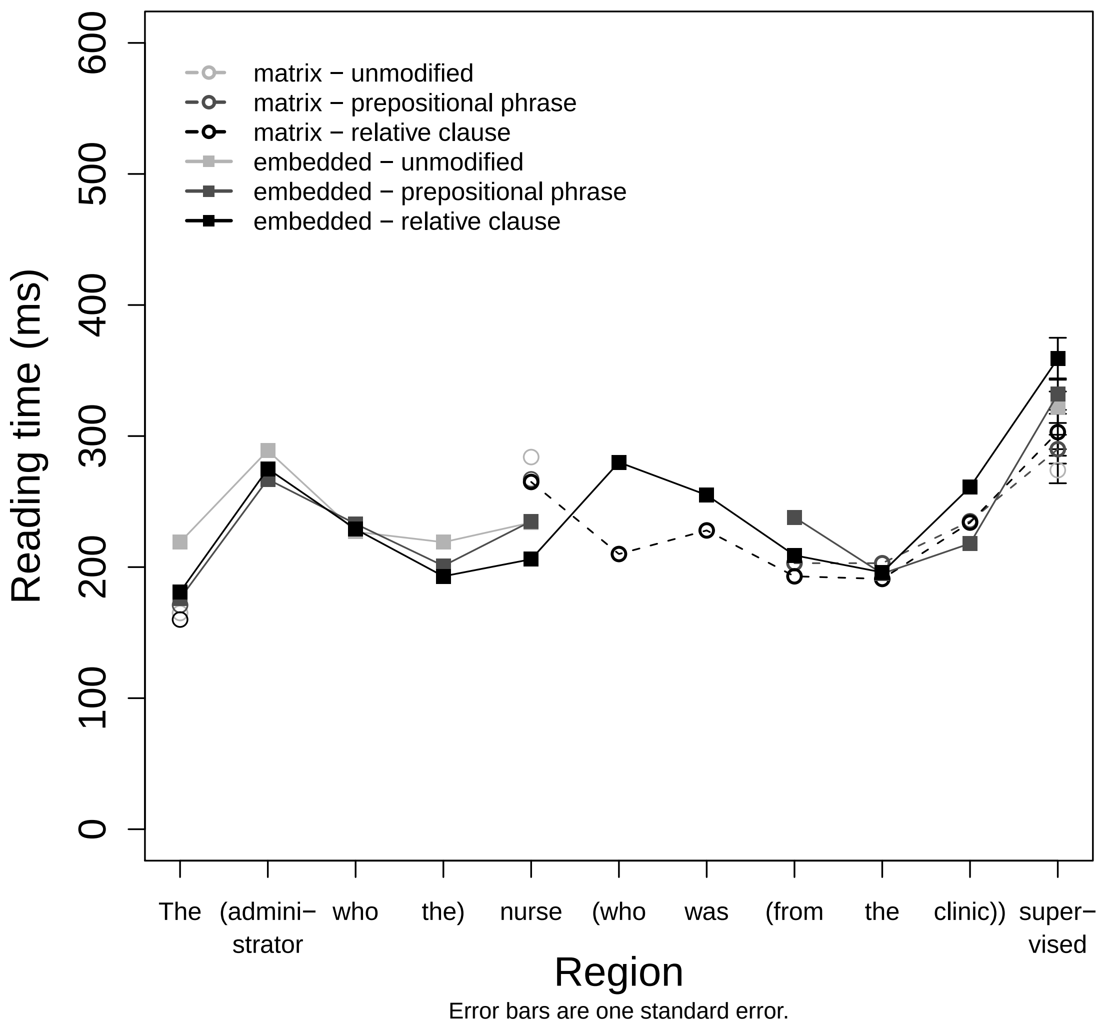

Figure 10. Mean total fixation time from each region: Experiment 2.

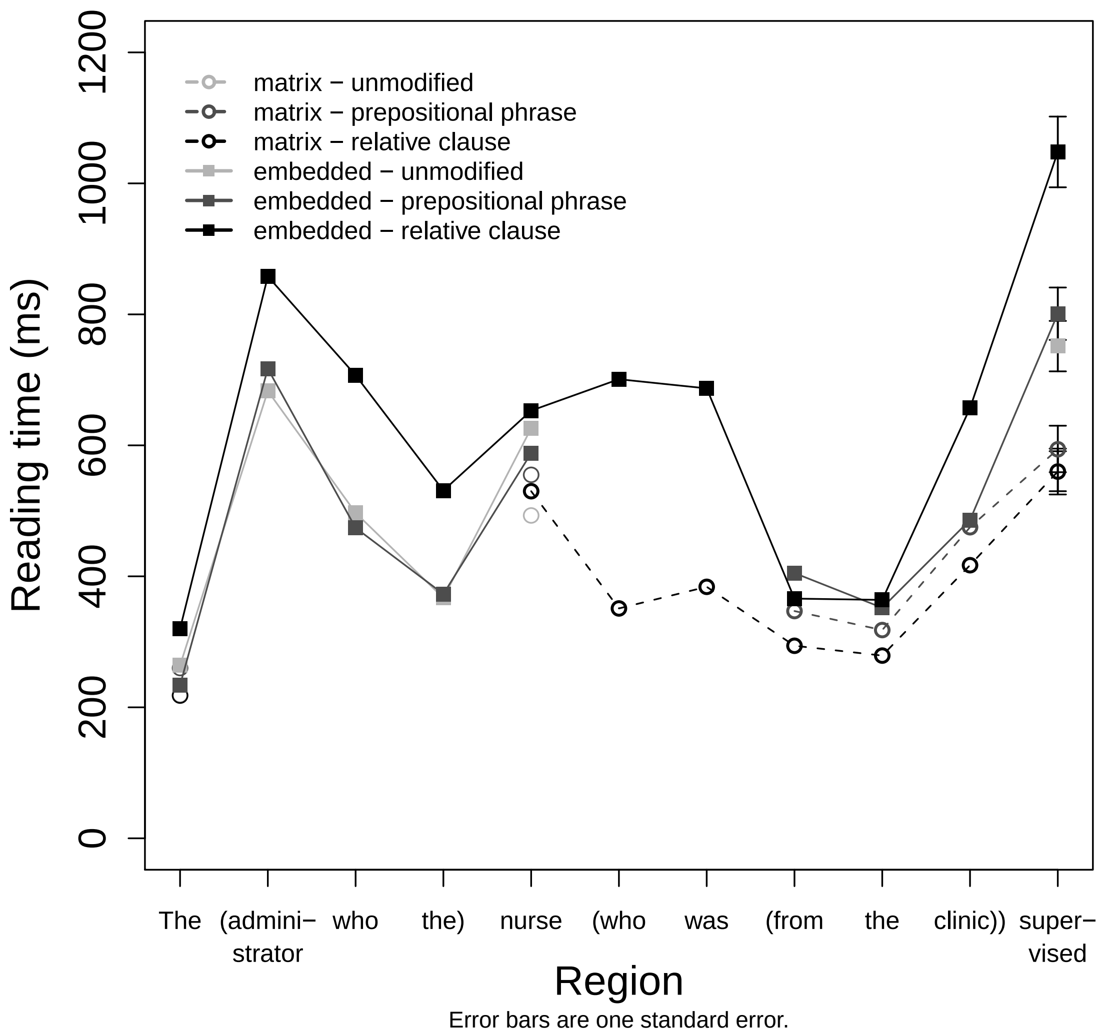

Figure 11. Mean reading time from each region: Experiment 3.

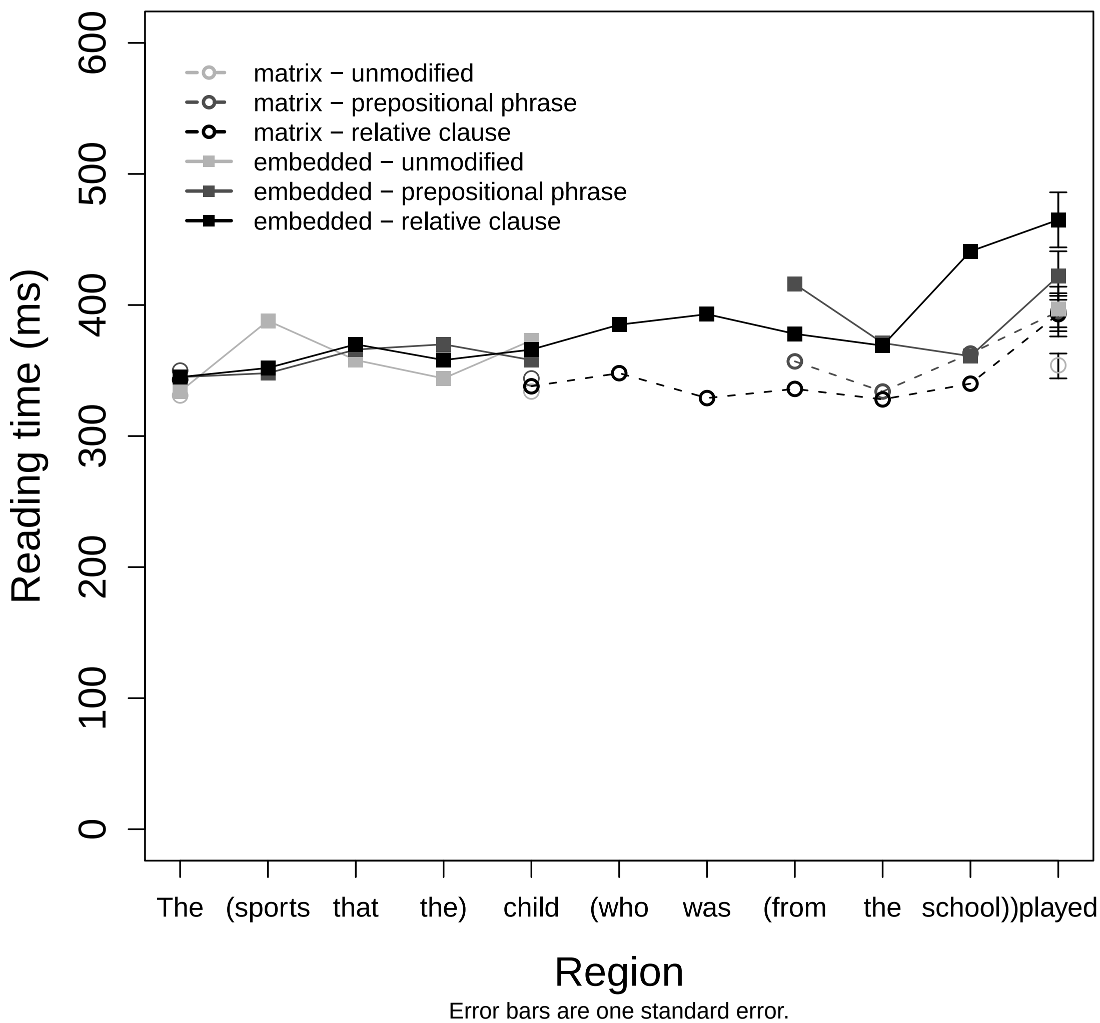

Figure 12. Mean first-pass reading time from each region: Experiment 4.

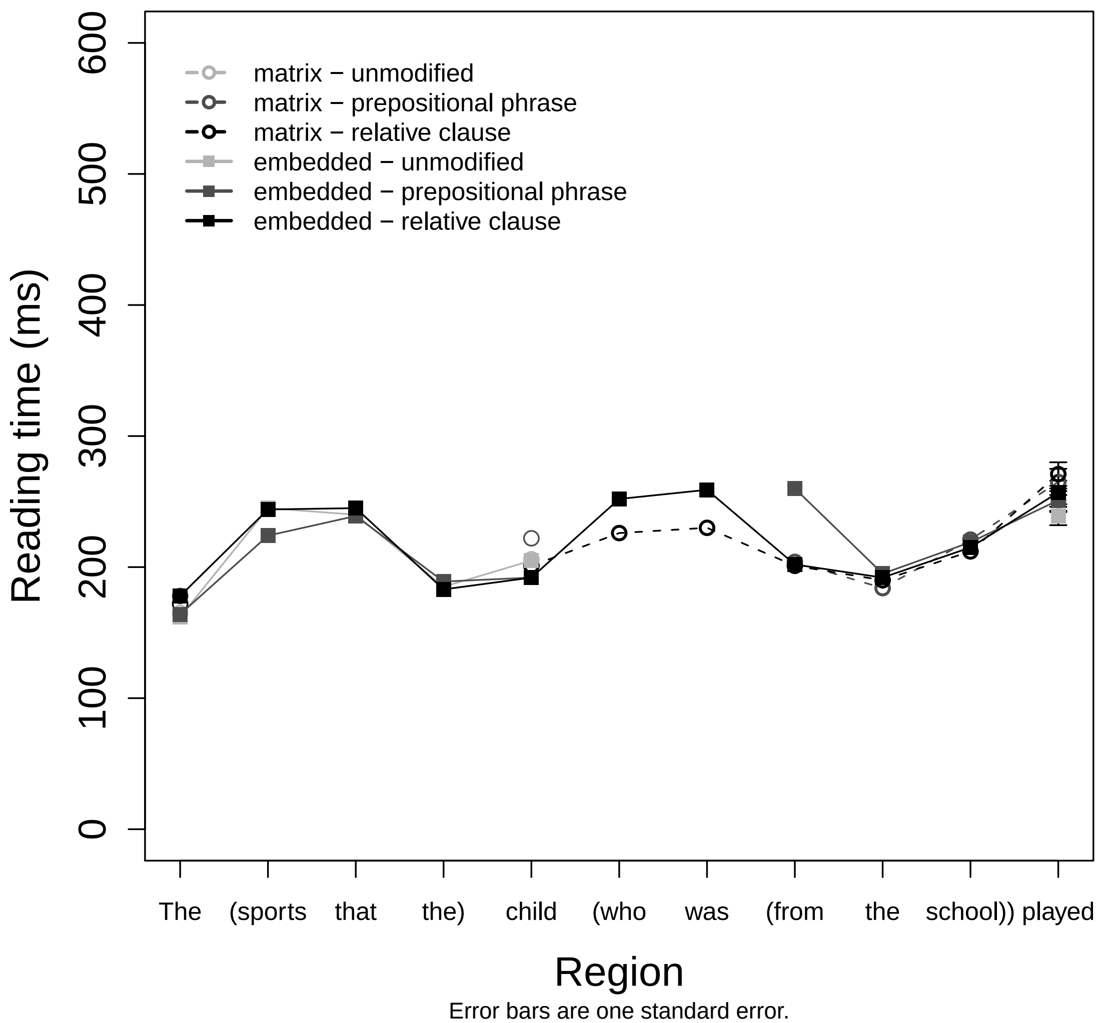

Figure 13. Mean total fixation time from each region: Experiment 4.