Consequences of the Serial Nature of Linguistic Input for Sentenial Complexity

Daniel Grodner[a], Edward Gibson[b]

[a]*Brown University*
[b]*Massachusetts Institute of Technology*

Abstract

All other things being equal the parser favors attaching an ambiguous modifier to the most recent possible site. A plausible explanation is that locality preferences such as this arise in the service of minimizing memory costs—more distant sentential material is more difficult to reactivate than more recent material. Note that processing any sentence requires linking each new lexical item with material in the current parse. This often involves the construction of long-distance dependencies. Under a resource-limited view of language processing, lengthy integrations should induce difficulty even in unambiguous sentences. To date there has been little direct quantitative evidence in support of this perspective. This article presents 2 self-paced reading studies, which explore the hypothesis that dependency distance is a fundamental determinant of reading complexity in unambiguous constructions in English. The evidence suggests that the difficulty associated with integrating a new input item is heavily determined by the amount of lexical material intervening between the input item and the site of its target dependents. The patterns observed here are not straightforwardly accounted for within purely experience-based models of complexity. Instead, this work supports the role of a memory bottleneck in language comprehension. This constraint arises because hierarchical linguistic relations must be recovered from a linear input stream.

Keywords: Parsing; Syntactic complexity; Locality; Memory and language

1. Introduction

In 1951 Karl Lashley famously argued that the complexity of human sequential behavior could not be explained by associative chaining mechanisms (Lashley, 1951), but rather indicated the existence of abstract hierarchical schemata. This insight has perhaps had its greatest effect on the study of linguistic behavior. Language is the prototypical serialized cognitive sys-

Requests for reprints should be sent to Daniel Grodner, Department of Cognitive and Linguistic Sciences, Brown University, Box 1978, Providence, RI 02912. E-mail: dgrodner@brown.edu

tem, and most models of language production take as their primary explanatory focus the fact that constituent structures must be filtered through a linear output (e.g., Dell, Burger, & Svec, 1997). However, less attention has been given to how the seriality of language might affect perceptual processes. This article addresses this issue by investigating a resource bottleneck in comprehension. This bottleneck arises because hierarchical relations must be extracted from a linear input stream.

To a first approximation, linguistic units are encountered in succession both in reading and listening (Rayner & Pollatsek, 1995). As each unit is input to the processing mechanism, it is immediately incorporated into a partial sentence interpretation (e.g., Marslen-Wilson & Tyler, 1980; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). Although lexical items are processed in order, linguistic dependencies can span several words, such as when a subject is separated from its verb by a modifier. Thus, when a new word is encountered, it sometimes must be integrated with a dependent element that came much earlier in the input. Because mental representations become degraded over time or with more input (e.g., Anderson, 1983), the difficulty of this computation is, in part, determined by the amount of lexical material intervening between the new item and the target site of the attachment. In the following sections we review evidence that this simple factor is important in structural ambiguity resolution. We then present two experiments that indicate that this factor is a fundamental determinant of unambiguous sentence complexity.

1.1. Locality in ambiguity resolution

All other things being equal, the parser favors attaching an ambiguous constituent to the most recent possible site. For instance, the sentences in (1) contain an element that can be linked structurally with either a more or less recent verb. There is a strong preference to attach to the more recent verb in each case, leading to a semantic anomaly (cf. Altmann, van Nice, Garnham, & Henstra, 1998, for experimental evidence from related examples).

(1a) *Mary said that the kids will swim yesterday.*
(1b) *Joe figured the puzzle that Sue took out.*

These types of preferences have motivated a number of theoretical mechanisms (cf. Frazier's late closure, 1979; Gibson's recency, 1991; Kimball's right association, 1973). In general, these theories claim that the parser favors a particular structural analysis over all others. Subordinate interpretations are not distinguished from one another. Contrary to this view, Pearlmutter and Gibson (2001) provided evidence for graded preferences over several sites. They looked at modifier attachment in sentences containing three verbal predicates (VPs) such as (2):

(2) *The judge noted that the guard confirmed that the prisoner confessed the crime last week.*

Participants experienced the greatest difficulty when the modifier (e.g., "last week") was semantically compatible with only the highest VP, somewhat less difficulty when the modifier went with the middle VP, and the least difficulty when the modifier attached low.

Observations from noun phrase (NP) modification are somewhat messier, but when there are three sites as in (3), there appears to be a cross-linguistic preference to attach relative clauses (RCs) to the most local of the potential nominal heads (e.g., Gibson, Pearlmutter, Canseco-Gonzalez, & Hickok, 1996; Miyamoto, Gibson, Pearlmutter, Aikawa, & Miyagawa, 1999).

(3) *The computer next to the painting of the houses which was damaged in the flood. …*

There is by now some consensus that a locality preference exists and is responsible for the preceding results, but there is less agreement on the origins and pervasiveness of locality effects. Some researchers have deemphasized the role of locality, invoking it only for the purposes of selecting among candidate readings of an ambiguity. Indeed, Frazier's (1979) variant on locality, Late closure, operates only if other grammatical decision principles make no commitments. This view is tenable because the most convincing evidence for locality effects so far comes from situations involving ambiguous modifier attachment. Deciding among possible attachment sites for a modifier may amplify the role of locality because one of the few ways to distinguish among the candidate structures is the serial position of the attachment sites.

Locality has also figured in descriptions of what perceivers do after an interpretation of a partial sentence turns out to be incorrect. Sturt, Scheepers, and Pickering (2002) examined the processing of sentences that caused misanalysis and in which there were initially two points at which reanalysis could proceed. Individuals chose to reanalyze at the more recent site. Consistent with this, Frazier and Clifton (1998) proposed that more recent phrases are more "visible" to reanalysis. The longer an individual has been holding onto an incorrect analysis, the harder reanalysis will be. In this view recency affects *uninterpreted* material. The sensory trace of each word in a string decays rapidly after being input and interpreted. When a partial analysis turns out to be incorrect, some portion of it must be jettisoned, and the sensory trace has to be reactivated to permit reanalysis. Significantly, this view makes no claims about the decay of elements that are already interpreted. It makes claims about reanalysis, not first-pass parsing.

A rather different view holds that locality preferences in comprehension emerge as a by-product of constraints on production. For instance, MacDonald and colleagues have argued that ordering preferences in production are responsible for the aforementioned effects (e.g., MacDonald, 1999). Short constituents tend to be produced before long constituents (e.g., Stallings, MacDonald, & O'Seaghdha, 1998; Thornton & MacDonald, 1997). Because longer constituents tend to appear at the end of sentences, long-distance dependencies will be infrequent. To illustrate, this view predicts that the adverb attachment anomaly in (1a) arises because there are alternative word orderings that convey the high attachment, as in (4):

(4a) *Mary said yesterday that the kids will swim.*
(4b) *Yesterday, Mary said that the kids will swim.*

These orderings are more frequently associated with the high attachment than that in (1a), because the shorter constituent (*Yesterday*) precedes the lengthy sentential complement (*that the kids will swim*). Locality preferences in comprehension then arise because comprehenders are attuned to the relative frequency with which the surface forms map onto high and low attachments.

All of the preceding characterizations of locality assign to it a limited or secondary role in comprehension. This work takes a different approach. It argues that locality is a ubiquitous property of comprehension processes. Namely, more distant sentential material is more difficult to reactivate in memory than more recent material. Recency preferences in structures such as (1) and (2) arise in the service of minimizing these memory costs. But these are just special cases of a more general property. Processing *any* sentence requires linking each new lexical item with material in the current parse. Often this will involve the construction of long-distance dependencies. Under a resource-based view, lengthy integrations should induce difficulty even in unambiguous sentences (cf. Gibson, 1998; Stevenson, 1994; Vosse & Kempen, 2000). To date there has been little direct quantitative evidence in support of this perspective. This article explores the hypothesis that locality is a fundamental determinant of reading complexity in unambiguous constructions.

1.2. Locality in unambiguous sentences

According to this locality hypothesis, integrating a newly input syntactic head *h* into the current structural analysis of the input requires reactivation or retrieval of the element or elements in this structure to which *h* connects syntactically (e.g., as part of a head-dependent relation, or as part of a syntactic chain). The difficulty of reactivating previous elements in the structure depends on how far back in the input they have occurred, what kind of elements have occurred in the interim, and how much they interfere with one another and the head to be connected. This view can explain offline relative complexity judgments for a wide array of structures (see Gibson, 1998, 2000, and Hawkins, 1994, for overviews). The studies that follow provide evidence that word-by-word integration costs are affected by a graded form of locality as formalized in (5) (adapted from Gibson, 1998).

(5) *Integration cost:* The resources consumed by incorporating each new lexical item into a particular parse is a monotonically increasing function, $I(n)$, where n represents the distance between the item and the site to which it is being attached to form a larger constituent.

Distance could be measured in a number of ways, such as the number of words (Hawkins, 1994), syllables, or the number of interfering similar elements (Gordon, Hendrick, & Johnson, 2001; Lewis, 1996). As an initial hypothesis, we follow Gibson (1998) and Warren and Gibson (2002) by computing distance in terms of the complexity of the intervening discourse structure. Distance here is calculated as the number of new discourse referents that are introduced between the endpoints of an integration, where a discourse referent is defined to be a new discourse object (an NP) or a tensed verbal element. Notwithstanding, the predictions and results that follow are the same regardless of whether distance is measured in terms of words, referents, or similar NPs. To simplify calculations, we assume a linear model of integration cost where $I(n) = n$.[1] Although there may be exceptions, we assume that words can be identified before they can be syntactically and semantically integrated into a grammatical structure. Therefore the new lexical item is included in the measure of distance.

It is natural to inquire what memory processes might cause attachment sites to become degraded. Locality claims that target sites degrade as a function of the amount of intervening lin-

guistic material. This suggests that the representations of elements within partially connected structures can be disrupted by incoming elements. Such interference has been the cornerstone of several recent models of sentence complexity (Gordon et al., 2001; Lewis, 1993, 1996; Van Dyke & Lewis, 2003). However, as noted previously, certain locality-based effects have been attributed to the decay of the linguistic input. These claims have mostly been in the context of sentential reanalysis effects (Bader, 1998; Frazier & Clifton, 1998; Sturt et al., 2002; but see Gibson, 1998; Stevenson, 1994; Vosse & Kempen, 2000; for models where decay plays a fundamental role in structural complexity). Our work examines unambiguous structures and does not attempt to distinguish between interference and decay as mechanisms underlying locality effects. One attractive possibility is that both mechanisms contribute to processing complexity: Partially interpreted material may be subject to interference, and uninterpreted sensory traces may be subject to decay. If so, then the factors that make unambiguous structures complex and the factors that make recovery from structural misanalysis more difficult should be different despite superficial similarity. More detailed examination of this issue is needed.

It is important to distinguish this resource-based explanation for locality-based processing costs from an experience-based account. For the latter, processing complexity is determined by how closely the input conforms to structural distributions the perceiver has been exposed to in their environment. Preferences in ambiguity resolution and difficulty with unambiguous sentences might then arise because nonlocal attachments are less frequent than more local ones. This view shifts the burden of explaining locality-based difficulty onto theories of production. Such production-centered theories have only been elaborated for adverb attachment (MacDonald, 1999) and heavy NP shift (Stallings et al., 1998; Wasow, 1997), but there has been speculation that this line of reasoning could extended to account for other locality effects. In particular, if a central component of sentence processing is predicting upcoming words (Elman, 1991), then perceivers should experience difficulty whenever the input deviates from a canonical word order. Many such arrangements involve nonlocal dependencies—for instance, when adjuncts intervene between arguments and predicates (Konieczny, 2000), or *wh* pronouns must be associated with object gaps in relative clauses (RCs; MacDonald & Christiansen, 2002). However, experience and resource-based theories of complexity make different predictions about *where* difficulty will be encountered. The experience-based account predicts that perceivers should have the greatest difficulty when there is the largest degree of uncertainty or the greatest amount of prediction error. This should typically occur at the onset of rare or unexpected configurations (e.g., the onset of an embedded clause or modifier). In contrast, if locality costs arise from resource-based perceptual limitations, then the greatest points of difficulty will surface when associations must be made between distant dependents. This typically occurs *after* regions of uncertainty and, in fact, at points where the syntactic category is highly predictable given the preceding context.

2. Experiment 1: Subject- versus object-extracted RCs

In the first experiment, participants read sentences in which the matrix subject was modified either by a subject- or object-extracted RC. Previous work has established that, ceteris paribus, object-extracted RCs are more difficult to process than subject-extracted RCs (Ford, 1983; Gordon et al., 2001; Hakes, Evans, & Brannon, 1976; Holmes & O'Regan, 1981; Just, Carpen-

ter, & Keller, 1996; King & Just, 1991; Mak, Vonk, & Schriefers, 2002; Stromswold, Caplan, Alpert, & Rauch, 1996; Traxler, Morris, & Seely, 2002; Wanner & Maratsos, 1978; Waters, Caplan, & Hildebrandt, 1987). These studies were designed to compare qualitative patterns of difficulty across sentence types. The primary contribution of this experiment is to test for quantitative word-by-word effects of a graded form of locality. A sample item is given in (6) along with predicted integration costs at each word.

(6a) Subject extraction
      The reporter who sent the photographer to the editor hoped for a story.
      0   1   0   1   0   1   1   0   1   4   0   0 1
(6b) Object extraction
      The reporter who the photographer sent to the editor hoped for a story.
      0   1   0   0   1   1+2 0   0   1   4   0   0 1

In the subject-extracted condition (6a), all integrations are local until encountering the preposition *to. Reporter, sent,* and *photographer* each incur a cost of 1, because they introduce new discourse referents. The preposition attaches to the embedded verb, *sent,* as its second argument. This link spans an intervening photographer referent, so the integration cost is 1. Integrations at *the* and *editor* are also local. The matrix verb, *hoped,* must be linked with the subject noun, *reporter.* This integration crosses four lexicalized discourse referents: (a) a photographer, (b) an editor, (c) a sending event, and (d) a hoping event. Integration costs for the object-extracted condition, (6b), differ from the subject-extracted condition primarily at the embedded verb (*sent*). Two integrations must be performed at this point. First, the verb must be locally linked with its subject, crossing 1 discourse referent. Second, the object of the verb must be associated with the *wh* pronoun co-indexed with *the reporter.* This integration traverses the embedded verb and embedded subject, a distance of 2 discourse referents. The total cost at this point is therefore $I(1) + I(2) = 3$. Thus locality-based integration costs predict that the primary locus of difference between the two conditions will be at the embedded verb.

The predictions of experience-based theories of sentential complexity are somewhat different. For instance, MacDonald and Christiansen (2002) hypothesized that perceivers try to predict the next word in an utterance by approximating structural frequencies in the environment. Complexity in such models can be equated with uncertainty or prediction error, and it arises especially when rare constructions are encountered (cf. Hale, 2003, for a formalization of a similar idea, which makes similar predictions). For the constructions in (6), MacDonald & Christiansen claim that object-extracted RCs are harder, because they do not follow the standard word order found in simple active one-clause sentences. Specifically, the logical object precedes both the embedded verb and the embedded subject. Subject-extracted RCs are alleged to be simpler, because they conform to the predominant subject–verb–object (SVO) pattern of English.[2] This account predicts the object-extracted RCs will be more difficult over the embedded clause, but this difficulty should arise at the embedded subject, where the departure from canonical word order begins. In contrast to the integration cost account, processing at the embedded verb might actually be facilitated for the object-extracted conditions. This is because the presence of a verb in English is likely to be more probable after the initial sequence *NP who*

NP … than after *NP who* … . Note that the latter string may initiate either a subject- or object-extracted RC, so that a verb or a subject might be the next constituent. The former string cannot be a subject extraction because there is an overt NP occupying the subject position. At this point a tensed verb is needed to act as the embedded predicate.

The experience-based account might also predict difficulty in the object-extracted RC because the embedded verb is not followed by an NP object.[3] The ditransitive argument structure of the verb and the dominant verb–object order of matrix clauses in English might provide local pressure for an object in this position. The potential conflict between local transitional probabilities and larger configurational patterns, which support positing a gap at this position, might result in increased processing load. This would lead to elevated reading times at the preposition immediately following the verb where the absence of an overt NP is signaled.

Another theory that predicts difficulty at the embedded subject of object-extracted RCs is the active filler strategy (AFS; Frazier, 1987; Frazier & Flores d'Arcais, 1989). Under this account, the increased complexity of object-extracted RCs arises because they are initially misanalyzed as subject-extracted RCs. The AFS claims that gaps are posited for fillers as soon as is grammatically possible. Thus, on seeing the *wh* pronoun in the string *NP who* … , the parser inserts a subject gap. For a subject-extracted RC, this is the correct analysis, so that no special difficulty arises in processing this structure. For an object-extracted RC, this subject-gap analysis is proved wrong as soon as the embedded subject is observed. As a result, the parser has to reanalyze the input resulting in increased processing complexity. Thus, like the experience-based theories, the AFS predicts maximal complexity for the object-extractions at the embedded subject. Pinpointing the locus of difficulty in these constructions will serve to distinguish the locality account from the experience-based and AFS accounts.

Two previous studies have examined word-by-word processing behavior on subject- and object-extracted RCs: King and Just (1991) and Gordon et al. (2001). Each found evidence of elevated reading times over regions containing the verbs in the object-extracted conditions. However, the methodologies and materials employed in these studies were not ideal for evaluating this hypotheses. The data reported in King and Just include a condition in which participants read sentences while simultaneously holding in memory the final word for several sentences. It is not clear how this secondary task may have affected primary reading processes. Further, the matrix subject head noun was analyzed along with the first three words of the embedded clause in a single region. This obfuscates reading times over the embedded verb in the subject-extracted condition, and the embedded subject in the object-extracted-condition. Gordon et al. reported words in isolation and had no secondary task, but displayed each word in the center of the screen. This unnatural presentation likely led to highly inflated reading times. In fact, participants spent over 950 msec per word on verbs for singly embedded object-extracted RCs. Moreover, in both of these studies, the matrix and embedded verbs appeared consecutively without any intervening material. It is possible that elevated reading times at the main verb occurred due to spillover difficulty from processing the embedded verb. To see whether integrating the main verb elicited difficulty, our items separated the main and embedded verbs by interposing a prepositional phrase (PP).

2.1. Methods

2.1.1. Participants

Forty-two individuals from the Massachusetts Institute of Technology (MIT) community were paid for their participation in this experiment.

2.1.2. Materials and design

Sixteen stimuli such as (6) were constructed, consisting of semantically reversible ditransitive verbs, so that the subject- and object-extracted variants contained identical lexical items. The second argument of the embedded verb was a three-word PP. This served to put distance between the predicted highest points of complexity (the matrix and embedded verbs) in the object-extracted condition. Stimuli were divided into two lists using a Latin-squares design. The stimuli were intermixed with filler items and comprised less than 35% of the sentences read by participants.[4]

2.1.3. Procedure

Sentences were presented using a noncumulative, self-paced, word-by-word display on a Macintosh computer running purpose-built software (Just, Carpenter, & Woolley, 1982). Each trial began with dashes standing in for all non-white-space characters in a sentence. Participants pressed the space bar to replace the next series of dashes with the word they concealed. With the exception of the first press, this action caused the previous word to disappear. A yes-or-no comprehension question followed each sentence. If the question was answered incorrectly, the word *INCORRECT* appeared briefly on the screen. There was no feedback for correct responses. Participants were instructed to read at a normal rate in a manner that would enable them to answer the comprehension questions. The computer recorded the time between button presses to the nearest millisecond. All analyses of interest were conducted on material appearing prior to a line break.

2.2. Results

2.2.1. Question accuracy

Overall 86.6% of questions were answered accurately. Accuracy rates for the subject- and object-extracted conditions were similar at 88.4% and 84.8%, respectively (Fs < 2, ps > .20). Two items had response rates just slightly above chance. Further examination revealed that these questions were ambiguous or presupposed information that was not attested in the sentence. Because of the high accuracy rate without these three items, no subjects or items were omitted on the basis of response accuracy.

2.2.2. Reading times

All reading time results are reported as the average per word in a region. Reading times that differed from the mean of a condition and region by more than 3 *SD*s were omitted from analyses. This adjustment discarded 1.6% of the data. Reading times at each word position are given in Appendix A. Four analyses were performed to explore the extent to which locality-based integration costs influence the processing of subject-modifying RC structures. The first three

analyses compare the qualitative predictions of integration costs with the alternatives described previously. The fourth test is intended to evaluate the quantitative fit of integration costs to the profile of reading complexity across the entire sentences.

For the first analysis, items were broken into regions consisting of major constituents as in (7):

(7a) The reporter | who | the photographer | sent | to the editor | hoped for a story.
(7b) The reporter | who | sent | the photographer | to the editor | hoped for a story.

To localize the source of differential difficulty between the conditions, the four regions of the embedded clauses were submitted to a 4×2 analysis of variance (ANOVA) crossing region by condition. Table 1 lists reading times by region and condition. As expected, object-extracted RCs were reliably harder than subject-extracted RCs, $F1(1, 41) = 7.9$, mean square error [MSE] = 6120, $p < .01$; $F2(1, 15) = 6.2$, $MSE = 2720$, $p < .05$. There were also significant effects of region and an interaction between region and condition (all Fs > 10, ps < .001). Planned comparisons between the two conditions revealed significant differences at the embedded verb, $t1(1, 41) = 11.9$, $MSE = 8000$, $p < .001$; $t2(1, 15) = 14.3$, $MSE = 2400$, $p < .01$, but not at the embedded NP or relative pronoun (ts < 1). This is consistent with the integration cost account, but runs contrary to the predictions of the experience-based account. Elevated reading times for object-extracted RCs persisted into the clause-final PP, $F1(1, 41) = 15.2$, $MSE = 2340$, $p < .001$; $F2(1, 15) = 8.6$, $MSE = 1540$, $p < .05$. This may be attributable to the self-paced reading methodology, where slowdowns often spill over into regions immediately downstream of a point of difficulty. In accord with this, reading times at the postverbal preposition differed between conditions (376 msec vs. 347 msec), although not as dramatically as for the verbs. We revisit the difficulty in the PP in the object-extracted condition at the end of this section.

Note that the order of the embedded NP and verb are reversed for the subject-extracted and object-extracted RCs. For this reason, it is possible that the embedded object in the subject-extracted RC is not an appropriate control for the embedded subject in the object-extracted RC. To confirm that there was no special difficulty at the onset of the object-extracted RC, an additional test was conducted comparing the first two of the two-word regions after the relative pronoun across conditions. To compensate for differences in word length, the comparison was performed on residual reading times computed by subtracting variance due to length from unadjusted reading times (using the procedure described by Ferreira & Clifton, 1986). Values are given in Table 2.

Table 1
Reading times per word (msec) by region over the embedded clause

Condition	Region Within the Embedded Clause			
	Relative Pronoun *who*	Noun Phrase *the photographer*	Verb *sent*	Prepositional Phrase *to the editor*
Object-extracted	343 (14)	353 (17)	422 (26)	398 (21)
Subject-extracted	350 (15)	360 (19)	355 (16)	356 (17)

Note. Standard errors given in parentheses.

Table 2
Length-adjusted residual reading times per word (ms) by two-word region over the RC

	Region Within the RC	
Condition	First Region	Second Region
Object-extracted	*the photographer* –25 (7)	*sent to* 25 (10)
Subject-Extracted	*sent the* –33 (5)	*the photographer* –10 (8)

Note. RC = related clause. Standard errors given in parentheses.

Table 3
Reading times in msec over the embedded and matrix verbs

RC Type /Verb	Embedded	Main Verb
Object-extracted	422 (26)	401 (22)
Subject-extracted	355 (16)	404 (26)

Note. Standard errors given in parentheses. RC = related clause.

There was an interaction of region and extraction type, which was reliable by participants, $F1(1, 41) = 6.3$, $MSE = 1300$, $p < .05$, and marginal by items, $F2(1, 15) = 3.4$, $MSE = 890$, $p < .10$. This reflects the fact that the object-extraction was significantly more difficult to process in the second region, $t1(1, 15) = 6.2$, $MSE = 690$, $p < .05$; $t2(1, 15) = 6.4$, $MSE = 690$, $p < .05$, but not in the first region ($Fs < 1.5$, $ps < .25$).

To establish whether the main or embedded verb was the primary locus of difficulty, a third analysis was conducted (Table 3). Distance-based integration costs predict that differential difficulty arises at the embedded verb. Consistent with this, reading times in the object-extracted sentences were slower over the embedded verb than for the subject-extracted sentences, but were essentially identical at the main verb. A 2×2 ANOVA crossing verb type (matrix vs. embedded) and extraction type confirmed that this interaction was significant, $F1(1, 41) = 5.25$, $MSE = 10050$, $p < .05$; $F2(1, 15) = 7.35$, $MSE = 2650$, $p < .01$. Looking just at the embedded verb, the difference between conditions was also reliable, $t1(41) = 18.8$, $MSE = 3330$, $p < .001$; $t2(15) = 27.9$, $MSE = 817$, $p < .001$.

The fourth analysis compared the predictions of distance-based integration costs with reading time difficulty across regions. Regions were determined as follows:

(8a) The reporter | who the photographer | sent to | the editor | hoped for | a story.
(8b) The reporter | who sent | the photographer | to the editor | hoped for | a story.

These regions are similar to those used in (7) except that the subsequent word is included at points of high integration cost (two or more cost units). This is because elevated reading times in self-paced reading are often observed a word downstream of points of difficulty. To minimize the number of short regions, boundaries were further adjusted so that function words did not appear in isolation (i.e., the relativizer *who* was grouped with the succeeding region). Figure 1 plots integration costs by region against reading times. Figure 2 depicts locality-based integration costs along with reading times by region separately for the two con-
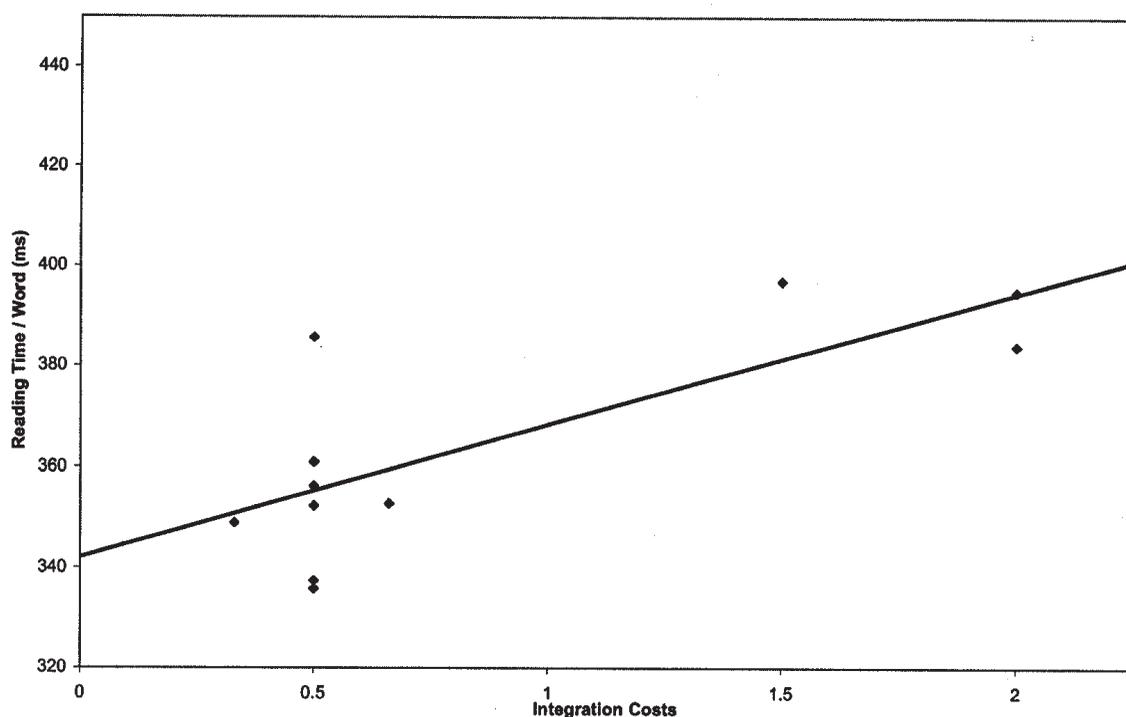
Locality as Predictor of RTs by Region

Fig. 1. Locality as a predictor of reading times in Experiment 1 by regions.

ditions (The slope and intercept were obtained by minimizing the mean squared error of a linear model predicting RTs with integration costs.). The qualitative predictions of distance-based integration difficulty resemble reading behavior across regions. There was a significant linear relation between these factors accounting for 59% of the variance in reading times ($p < .01$). Integration costs were also regressed against word-by-word reading times, explaining 33.8% of the variance ($p < .01$). If integration distance is measured by the number of words, rather than the number of new discourse referents, integration costs are still reliably correlated with reading times by regions ($r^2 = .487$, p < .01) and marginally by words ($r^2 = .15$, $p = .06$).

Predicted locality difficulty diverges most from reading times over the PP in the object-extracted condition. One plausible explanation for this effect is that the PP marks the end of the embedded clause. Clause-final slowdowns are often observed and are thought to reflect interpretive processes (cf. Just & Carpenter, 1980). Indeed, there were elevated reading times on the final word in the embedded clause for each condition. This was more apparent in the clause-final region for the object-extracted condition, in part, because there were only two words in this region as opposed to three words for the subject-extracted condition. In addition, the difficulty at the verb region in the object-extracted condition appears to have spilled over to the clause-final region.
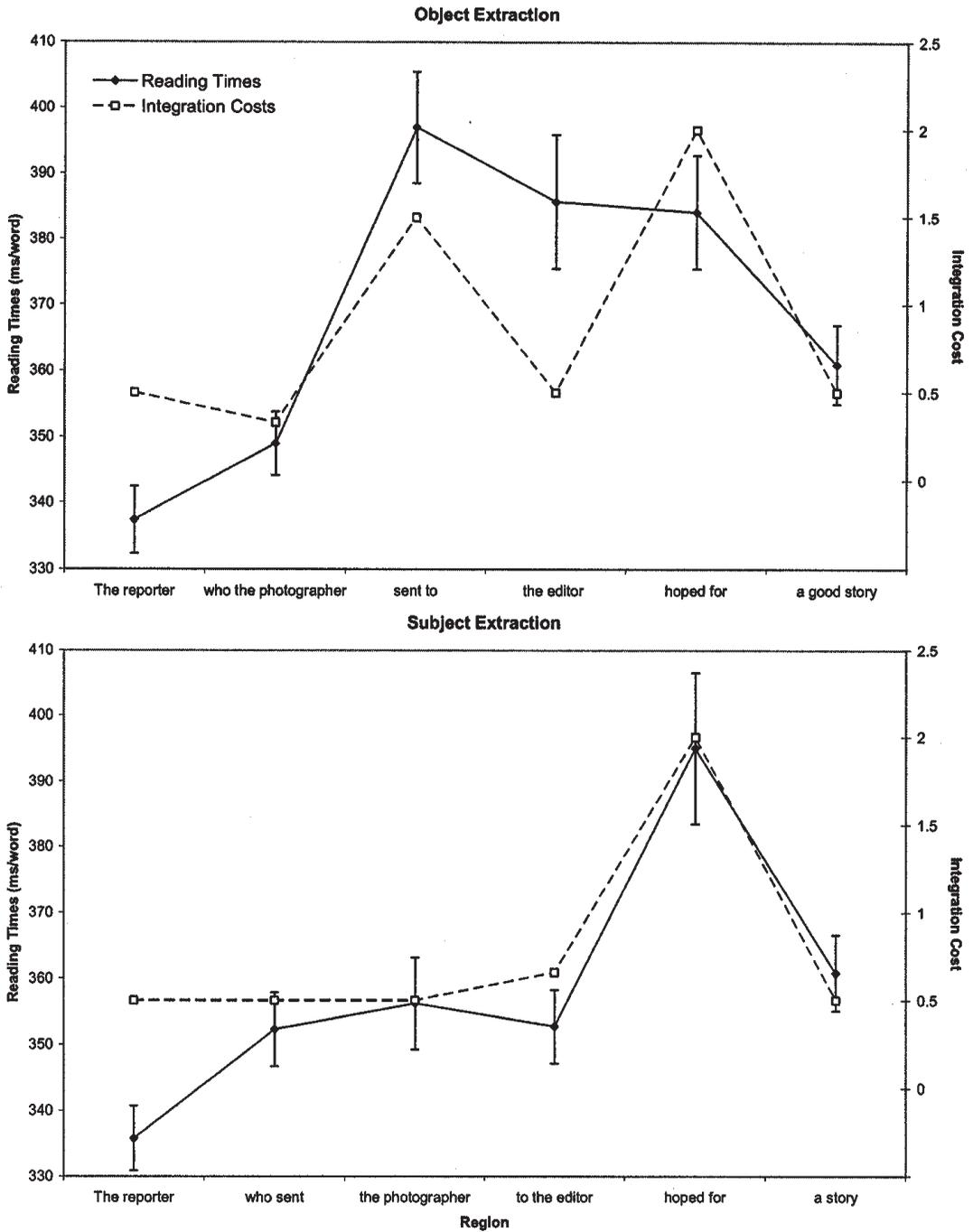
Fig. 2. Reading times and predicted integration costs for Experiment 1 plotted for consecutive regions.

2.3. Discussion

The data indicate that the processing load for object-extracted RCs diverges most from subject-extracted RCs at the embedded verb. This supports the thesis that locality effects derive from resource constraints imposed on perceptual processes rather than emerging as by-products of structural frequency matching. At this point in the clause, multiple nonlocal structural integrations must be performed. Thus locality can explain the differential difficulty between the two types of RCs. There was no special difficulty observed over the embedded subject of the object-extracted RC. This disconfirms the predictions of both the experience-based account and the AFS.

An additional finding was that a linear model of locality captured a large portion of the variance in region-by-region reading times. One caveat in drawing inferences from this result is that correlations can be driven by outliers in the data. Figure 1 illustrates that only 3 of the 11 regions were predicted to be substantially different from the others. Experiment 2 addressed this concern by looking at a wider variety of subject-modified constructions to test more rigorously the idea that reading difficulty can be manipulated by varying integration distance.

3. Experiment 2: Various forms of subject modification

In the second experiment, participants read a variety of nested and non-nested structures such as those given in (9). Distance-based integration costs are given at each word:

(9a) Matrix—unmodified subject
   The nurse supervised the administrator while …
   0   1   1   0   1   1
(9b) Matrix—PP-modified subject
   The nurse from the clinic supervised the administrator while …
   0   1   0   0   1   2   0   1   1
(9c) Matrix—RC-modified subject
   The nurse who was from the clinic supervised the administrator while …
   0   1   0   1   0   0   1   3   0   1   1
(9d) Embedded—unmodified subject
   The administrator who the nurse supervised scolded the medic while …
   0   1   0   0   1   3   3   0   1   1
(9e) Embedded—PP-modified subject
   The administrator who the nurse from the clinic supervised scolded the medic while …
   0   1   0   0   1   0   0   1   5   4   0   1   1
(9f) Embedded—RC-modified subject
   The administrator who the nurse who was from the clinic supervised scolded the medic while …
   0   1   0   0   1   0   1   0   0   1   7   5   0   1
   1

Points of high integration cost occurred at the main and embedded verbs. The magnitude of predicted integration cost was manipulated in two ways: (a) by inserting or omitting modifiers for the matrix and embedded subjects and (b) by varying the type of modification (RC or PP). In addition to examining a wider array of locality predictions, Experiment 2 compared the contributions of locality to reading behavior to several nonstructural factors. Specifically, thematic plausibility, lexical frequency, and word length are each known to affect processing difficulty. Indeed, it has been proposed that much of structural complexity emerges from the interaction of these lexical and contextual factors (MacDonald, Pearlmutter, & Seidenberg, 1994; Trueswell, Tanenhaus, & Garnsey, 1994). The hypothesis pursued here is that locality is an independent configurational property that applies similarly across sentences regardless of lexical and contextual idiosyncrasies. Thus in structures such as in (9), where locality difficulty is predicted to vary widely, the influence of distance-based integration costs will outweigh that of nonstructural factors.

Embeddings and modifiers create structures that are relatively infrequent. Rohde (2002) compiled structural statistics of sentences contained in the parsed Brown and *Wall Street Journal* corpora (Marcus, Santorini, & Marcinkiewicz, 1993). He found that 83.3% of all NPs have no postmodifier, 12.7% are PP modified, and only 4.1% are RC modified. Models that equate complexity with structural rarity will predict difficulty as soon as these modifiers can be identified. There is no reason to anticipate difficulty at the verbs in such a model.

This experiment also serves to evaluate certain prediction-based accounts of integration complexity. These accounts hold that the more predictable a particular syntactic head is, given the preceding context, the easier processing should be (cf. Hale, 2003; Konieczny, 2000, for accounts in this vein). Under these hypotheses, nominal modifiers of subject NPs, such as RCs and PPs, may actually *help* the perceiver anticipate the location and identity of the verb for syntactic and semantic reasons, in contrast to the integration-based account. Syntactically, the likelihood of seeing a predicted verb probably increases after each successive modifier. This is because, at each level of embedding, encountering a nominal postmodifier likely becomes less probable. For instance, in the Brown and *Wall Street Journal* corpora, 87.3% of NPs contained in RCs have no postmodifier compared to 80.9% of NPs in matrix clauses (Rohde, 2002). Semantically, modifiers specify certain properties of the NP, thereby making some verbs more expected than others. For instance, in the absence of modification or context, there is a certain probability distribution of expected verbs for a main-clause subject, such as *the administrator* in the embedded items in example (9). In this situation, the verbs *scolded* and *drove* might be equally likely, both of which would be more likely that the verb *fished.* Given a modifier describing a situation in which a nurse is supervising the administrator, the verb *scolded* becomes semantically more likely, and other possibilities become less likely. Because the conditions in Experiment 2 varied with respect to how many modifiers were interposed between subjects and verbs, and between fillers and gaps, they provide a test of the prediction-based hypotheses. These proposals predict that processing the obligatory matrix and embedded verbs in these items might be easier with more modification. This prediction is essentially the opposite of what is made by the integration-based account.

3.1. Methods

3.1.1. Participants
Forty-nine members of the MIT community were paid to participate in this study.

3.1.2. Materials and design
Thirty items were prepared containing six conditions. Each item was built around a core sentence consisting of a transitive verb with human NP arguments. The subject was either unmodified (9a), modified with a PP (9a), or modified with an RC (9c) containing the PP as a predicate. The RC conditions were constructed from the PP conditions by inserting the words *who was.* Semantic differences between the PP and RC conditions were thus minimal. An embedded variant of each type of sentence was also created. In these, the object NP of the core sentence—that is, *the nurse* in (9a)—became the subject of the matrix clause, and the remainder of the core formed an RC modifying that subject. The matrix clause in the embedded conditions also contained a simple transitive verb with an NP object. In all conditions the word after the matrix object was a clausal connective, which signaled a clause boundary. Stimuli were divided into six lists according to a 3 × 2 Latin-squares design crossing modification (bare, PP, or RC) with embedding (main or embedded). Sixty filler items were also included, containing a wide variety of embedded and unembedded structures.

3.1.3. Procedure
The task and stimulus presentation were identical to Experiment 1.

3.2. Results

Three items were eliminated from analyses, because they were constructed incorrectly and contained different words across conditions.

3.2.1. Comprehension questions
Question response accuracy by condition is given in Table 4. A 2 × 3 ANOVA revealed that the embedded conditions were significantly harder to understand than the unembedded conditions, $F1(1, 48) = 38.4$, $MSE = .03$, $p < .001$; $F2(1, 26) = 37.9$, $MSE = .02$, $p < .001$. There were no other reliable effects or interactions ($Fs < 1$).

Table 4
Response accuracy by condition for Experiment 2

	Main	Embedded
Bare	89.8% (2.2)	77.5% (2.2)
PP	90.6% (2.0)	76.7% (3.1)
RC	88.5% (2.2)	75.5% (3.0)

Note. PP = prepositional phrase; RC = related clause. Standard errors given in parentheses.

Some of the embedded conditions were quite difficult. It is not clear what sorts of strategies individuals adopt after a sentence has become too complicated to understand. In an effort to focus on processes during successful thematic and syntactic integrations, data from trials where participants responded incorrectly were excluded from reading time analyses.

3.2.2. Reading times

The primary regions of interest were the main and embedded verbs. Figure 3 depicts reading times over the verbs for each condition. A 2×3 ANOVA crossing modification with embedding revealed a main effect of modification by participants, $F1(2, 94) = 3.1$, $MSE = 22300$, $p < .05$, but not by items, $F2(2, 50) = 1.5$, $MSE = 27000$, $p = .24$. There was also a significant effect of embedding, $F1(1, 48) = 18.8$, $MSE = 58900$, $p < .001$; $F2(1, 25) = 37.8.5$, $MSE = 20400$, $p < .001$. There was no indication of an interaction ($Fs < 1$).[5] Planned individual comparisons revealed that verbs in the unmodified condition were reliably faster than in the RC condition, $F1(1, 48) = 4.0$, $MSE = 15200$, $p < .05$; $F2(1, 26) = 3.2$, $MSE = 11300$, $p < .05$. The difference between the unmodified and PP-modified conditions was marginally significant in the participants analysis, $F1(1, 48) = 2.7$, $MSE = 12800$, $p = .06$, but unreliable by items, $F2(1, 26) = 1.2$, $MSE = 9250$, $p = .28$. There was a numerical trend for verbs in the PP condition to be read faster than verbs in the RC condition, which was not anticipated by the locality hypothesis. This was not reliable ($Fs < 1$; see Figures 3 and 4.)

A number of regression analyses were performed to evaluate how well graded-integration costs predicted word-by-word reading times across conditions. Figure 5 plots reading times
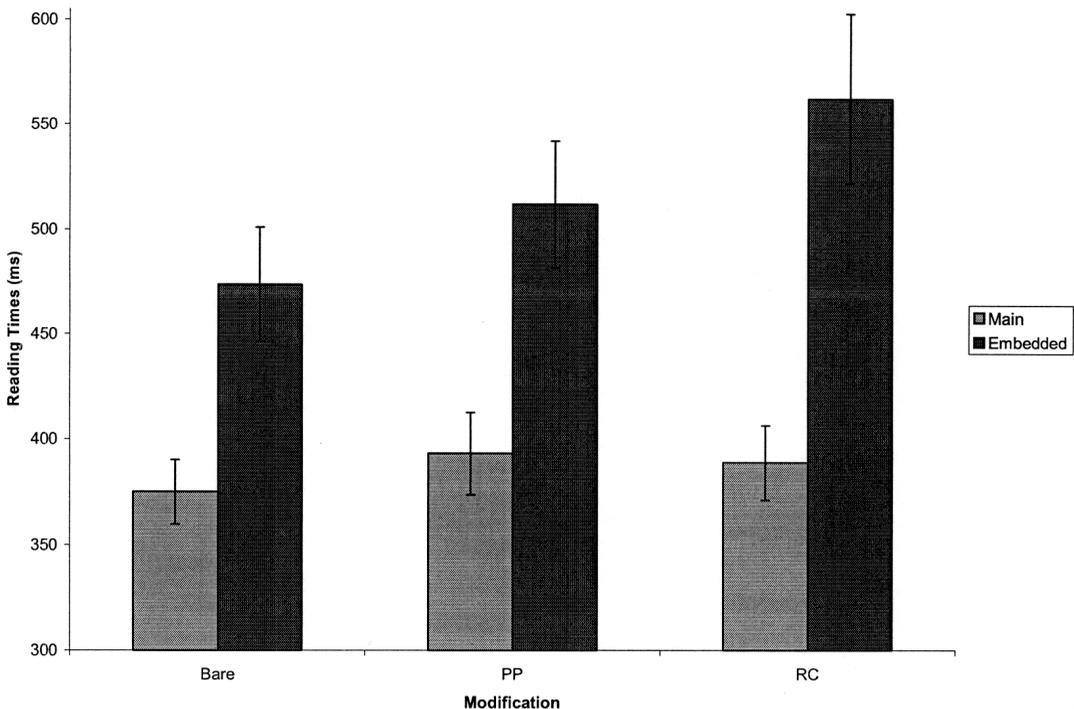

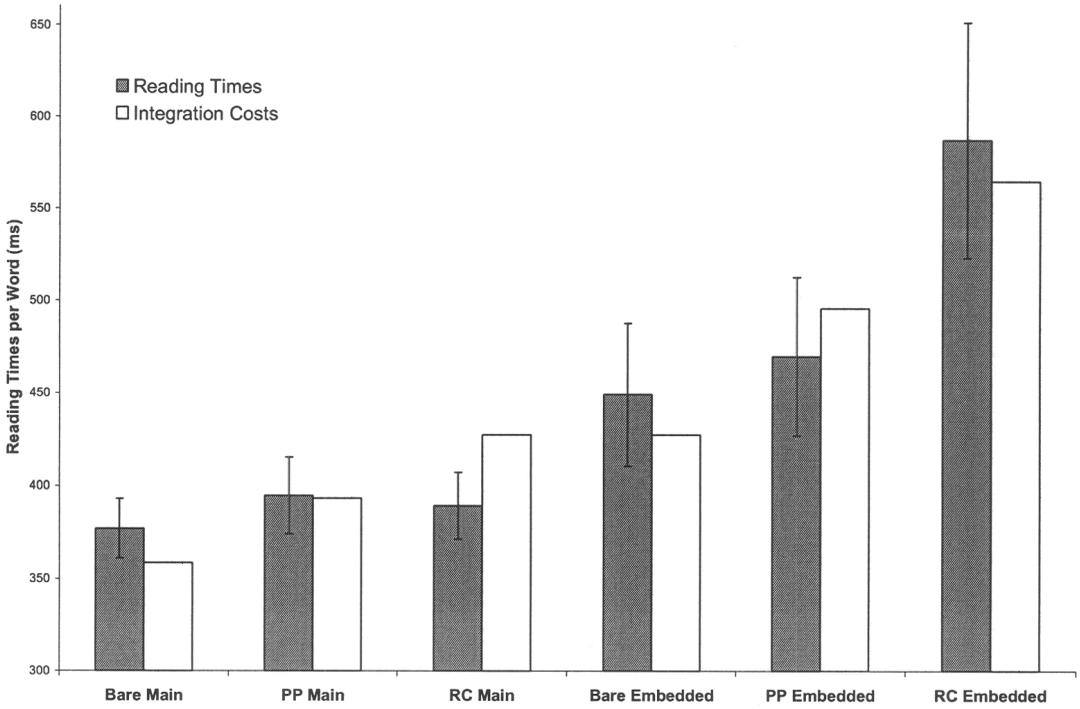Fig. 3. Average reading times on the verbs in Experiment 2 by condition.

Fig. 4. Reading times and predicted locality costs at the first verb for each condition in Experiment 2 ($r^2 = .889$).
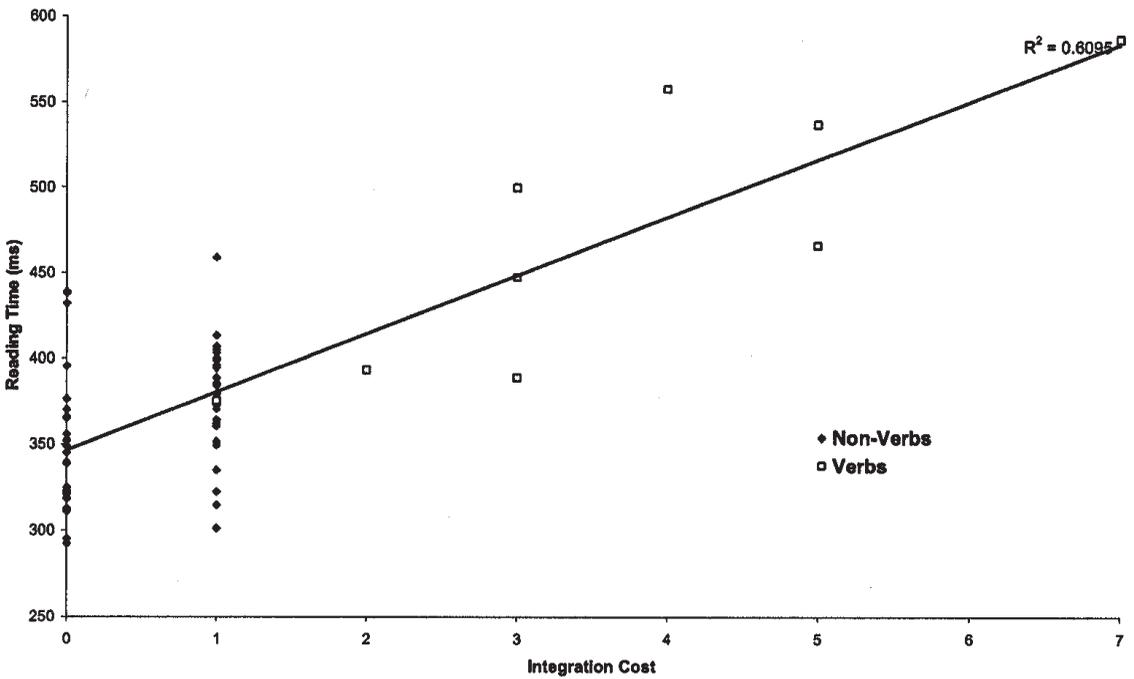

Fig. 5. Integration cost as a linear predictor of reading times per word in Experiment 2.

per word against integration costs. Overall, locality accounted for 61.0% of the variance in reading times, $F(1, 62) = 96.8$, $p < .001$. It is clear from Figure 5 that the distribution of costs is skewed toward points of 0 or 1 unit of integration cost. Costs at the verbs were evenly spread at unit intervals from 1 to 7. A linear regression including only reading times at the verbs revealed a highly significant linear relation (Figure 5), $r^2 = .69$ $F(1, 7) = 15.4$, $p < .01$. Looking just at the first verb where integration costs varied most widely (Figure 4), distance-based integration costs captured 88.9% of the variance in reading times, $F(1, 4) = 31.9$, $p < .01$.

Items were also divided into regions as in (10), following the same guidelines as in Experiment 1:

(10a) *The nurse | supervised | the doctor | while …*
(10b) *The nurse | from the clinic | supervised the | doctor while …*
(10c) *The nurse | who was | from the clinic | supervised the | doctor while …*
(10d) *The doctor | who the nurse | supervised | scolded the | medic while …*
(10e) *The doctor | who the nurse | from the clinic | supervised | scolded the | medic while …*
(10f) *The doctor | who the nurse | who was | from the clinic | supervised | scolded the | medic while …*

As illustrated in Figures 6 and 7, integration cost patterns resembled reading times across regions. Overall, integration costs predicted 72.9% of differences in regional reading times, $F(1, 29) = 77.9$, $p < .001$. Looking at only those regions containing verbs (Figure 6), this value was 62.5%, $F(1, 7) = 11.6$, $p < .01$.
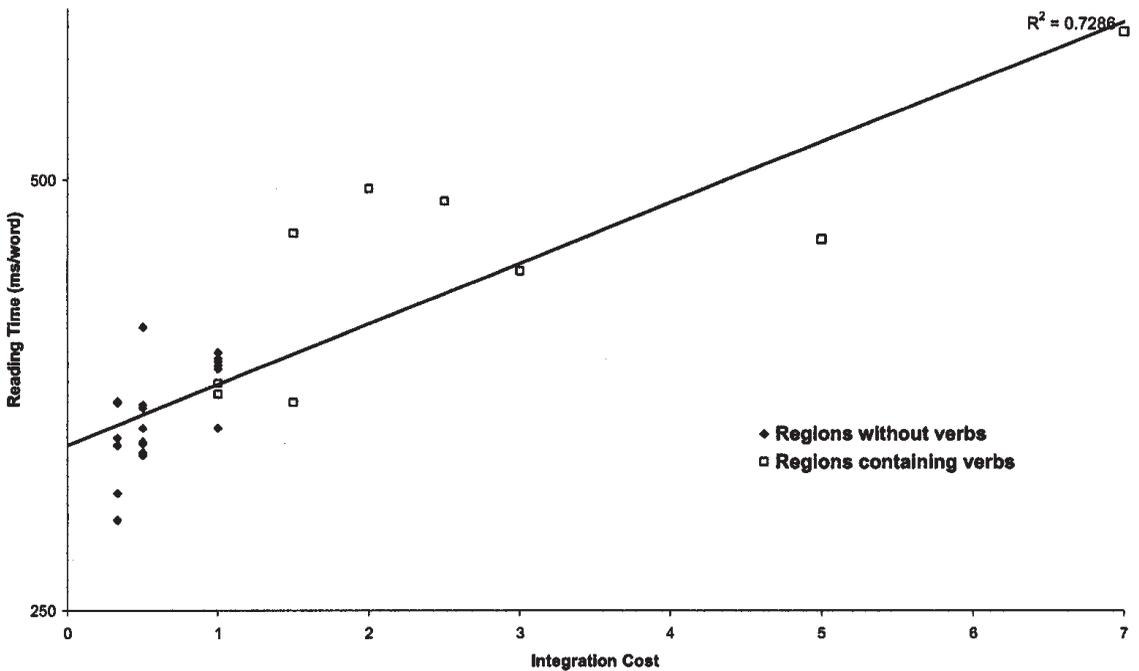

Fig. 6. Integration cost as a linear predictor of reading times per region in Experiment 2.
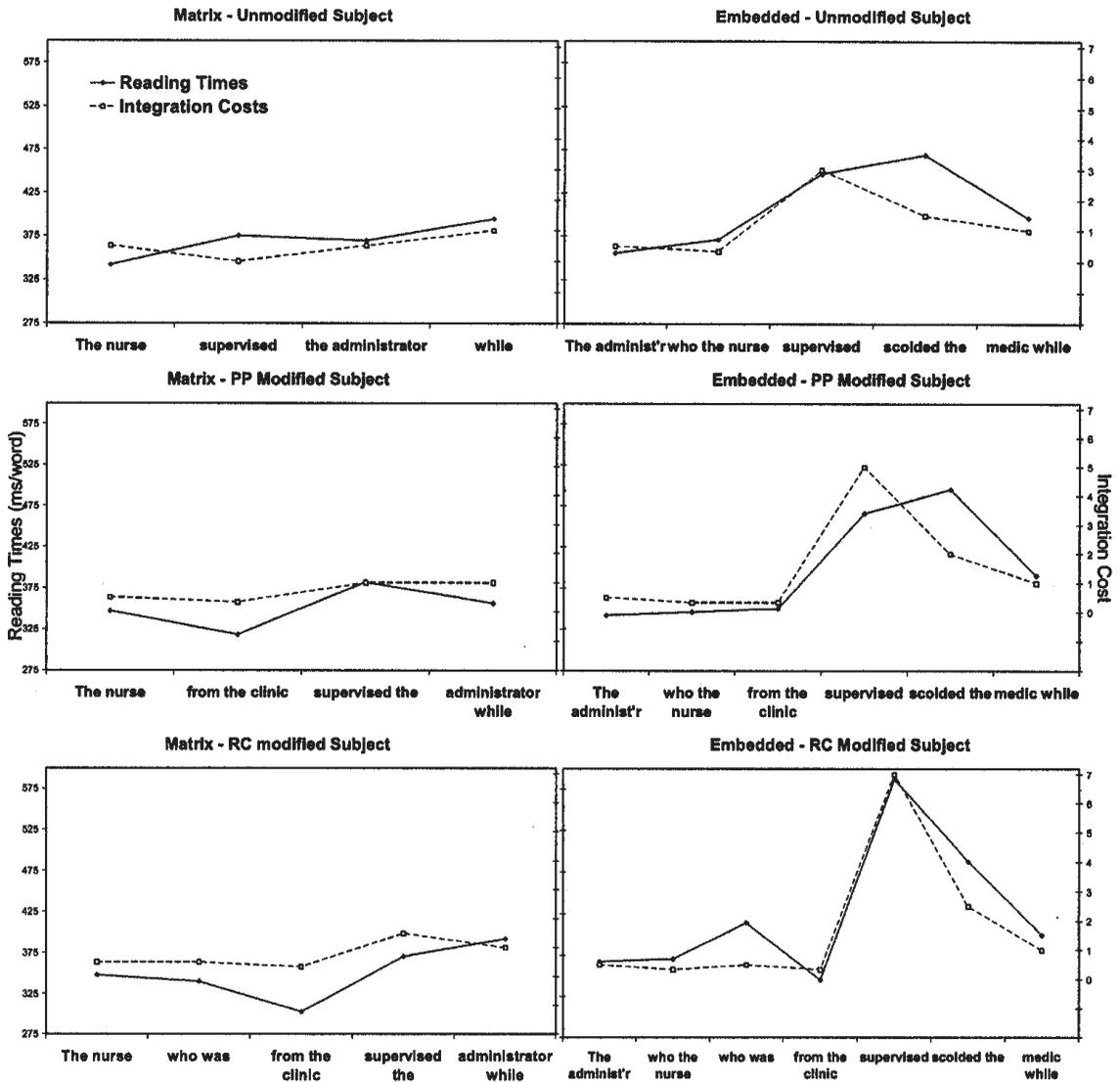
Fig. 7. Regional reading times plotted with predicted integration costs in Experiment 2.

3.2.3. Comparing integration costs and nonstructural factors

The preceding analyses included average reading times collapsed over items. This likely eliminated a substantial amount of variance due to lexical and plausibility factors and might, thereby, have exaggerated the influence of locality. To facilitate comparison between locality and nonstructural item-specific influences in these items, a series of correlational analyses were performed including interitem differences.

To evaluate the effects of plausibility, a survey was administered to 60 individuals from the MIT community who did not participate in the main experiment. For each item, participants

were directed to rate the plausibility on a scale of 1 (*plausible*) to 7 (*implausible*) of one of the conditions given in (11):

(11a) Clause 1—unmodified nonspecific object
    The nurse supervised someone.
(11b) Clause 1—unmodified object
    The nurse supervised the administrator.
(11c) Clause 1—PP nonspecific object
    The nurse from the clinic supervised someone.
(11d) Clause 1—PP object
    The nurse from the clinic supervised the administrator.
(11e) Clause 2—nonspecific object
    The administrator scolded someone.
(11f) Clause 2—object
    The administrator scolded the medic.

Previous evidence indicates that the products of thematic combination affect reading behavior as soon as the relevant role players and assigners are available (Garnsey, Pearlmutter, Myers, & Lotocky, 1997; Spivey-Knowlton & Sedivy, 1995; Trueswell et al., 1994). In accord with this, the ratings for each item were entered into regressions at verbs and objects where thematic role assignments first became available. Table 5 lists these positions by condition for the item in (9).

In addition to plausibility, two word-level attributes were also examined. The first of these was word length, which is directly related to reading time (e.g., Ferreira & Clifton, 1986). The second was log-normalized lexical frequency, which is inversely related to reading time (e.g., Plaut, McClelland, Seidenberg, & Patterson, 1996).

Table 5
Thematic combinations available at each region and condition for Item (9)

Experiment 2 Conditions	Verb 1 ("supervised")	Verb 2 ("scolded")	Object ("the medic")
Matrix —unmodified	The nurse supervised someone.	n/a	The nurse supervised the administrator.
The nurse supervised the administrator.			
Matri— RC/PP modified	The nurse from the clinic supervised someone.	n/a	The nurse from the clinic supervised the administrator.
The nurse (who was) from the clinic supervised the administrator.			
Embedded —unmodified	The nurse supervised the administrator.	The administrator scolded someone.	The administrator scolded the medic.
The administrator who the nurse supervised scolded the medic.			
Embedded —RC/PP modified	The nurse from the clinic supervised the administrator.	The administrator scolded someone.	The administrator scolded the medic.
The administrator who the nurse (who was) from the clinic supervised scolded the medic.			

Note. n/a = not applicable; RC = related phrase; PP = prepositional phrase.

Figure 8 depicts the relation between integration costs by region after interitem variance was included. This adds a great deal of noise to the data (indeed, many of the points plotted represent a single observation). Nevertheless, there was a significant linear correlation with integration costs across regions, $r = .40$, $F(1, 828) = 154.6$, $p < .001$; and words, $r = .34$, $F(1, 1711) = 227$, $p < .001$. Looking just at reading times at the first verb, 24.1% of the variance was captured, $r = .49$, $F(1, 159) = 50.5$, $p < .001$. Table 6 summarizes the regression analyses. The contribution of locality was greater than any other factor. This indicates that the robust correlations described previously were not an artifact of washing out the lexical and thematic variation across items.

Note that the inclusion of interitem variance could only have reduced the magnitude of any linear relation with integration costs, because integration costs were identical for each item. In contrast, each of the nonstructural factors varied across items. Thus they had the potential to explain a greater amount of variance when regressed with means across items. This did not happen. The predictive power of every factor decreased in the across-items analysis. Thus locality is an important determinant of reading times in the present sentences.

A stepwise regression was also performed to see the extent to which locality could explain the variance in reading times left unaccounted by the linear contributions of other factors. A multiple-regression equation was derived, using all of the nonstructural factors. The residual variance in reading times was then regressed against integration costs. Results of these analy-
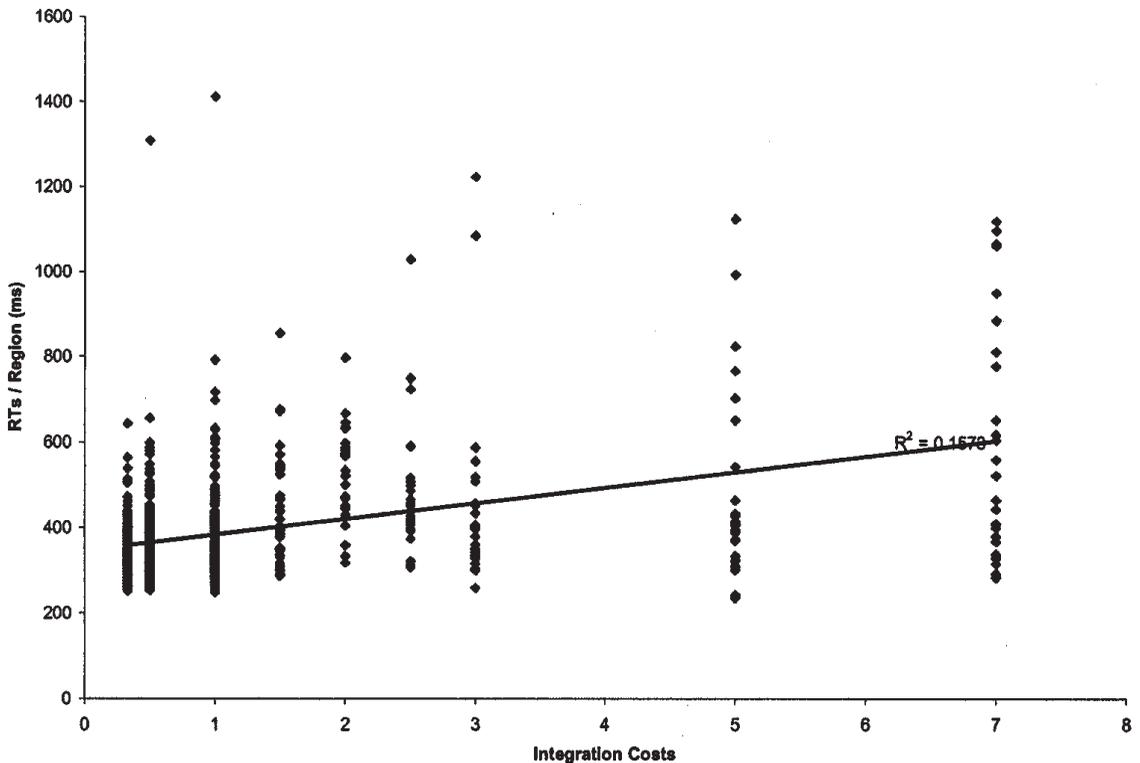

Fig. 8. Integration costs as a predictor of regional reading times for each item in Experiment 2.

Table 6

R^2s for regressions of several factors against reading times

Factor	Collapsing Across Items			Between Items		
	By Region	At Verbs & Objects	By Words	By Region	At Verbs & Objects	By Words
Integration cost	.729***	.706***	.61***	.157***	.11***	.117***
Log(lexical frequency)	.244**	.035	.17***	.044***	.003	.03***
Word length	.212**	.002	.208***	.053***	.012*	.053***
Propositional typicality[a]	n/a	.01	n/a	n/a	.022**	n/a
Locality after other factors have entered	.388***	.484**	.312***	.081***	.102***	.053***

Note. n/a = not applicable.

[a]Propositional typicality was reliably related to reading times at the points of thematic assignments ($p < .01$). This confirms that thematic plausibility has an effect in these items. However, this factor does not predict much variation in reading times. This is not surprising because items were not designed to vary along this dimension. In fact, items were constructed to be plausible and reflect circumstances of ordinary language comprehension.

*$p < .05$. **$p < .01$. ***$p < .001$.

Table 7

R^2s for log normalized conditional *n*-gram models collapsing across item regressions of several factors against reading times[a]

	By Region	At Verbs & Objects	By Words
Unigram	.216**	.318*	.029
Bigram	.296**	.442**	.144**
Trigram	.189*	.42**	.088*
4-gram	.2*	.496**	.068*

[a]The unigram model is based on category frequency, because the conditional model is impossible. There was one 4-gram category sequence with zero frequency. For the region and word analyses, this value was set to 0.1 to avoid division by zero.

*$p < .05$. **$p < .01$.

ses are reported on line 5 of Table 6. There was still a significant relation between locality and the residual reading times by words, verbs, and regions (Fs > 10, $p < .001$).

To obtain a reasonable baseline for a frequency-based prediction model, the performance of various *N*-gram models was evaluated in predicting this data. Sequences of syntactic category tags were collected from the combined parsed Brown and *Wall Street Journal* corpora, which contains just over 1 million words (Marcus et al., 1993). Unigram, bigram, trigram, and 4-gram statistics were collected. Results for log-normalized conditional probability variants of these models are given in Table 7. Although the models were strongly correlated with one another (r^2s > .45, ps < .001), the bigram model performed the best across regions and words. Nevertheless this model was outperformed by the predictions of distance-based integration costs.

3.3. Discussion

A linear model of distance-based integration costs makes excellent quantitative predictions of reading times by words and regions across the sort of subject modifications examined in this experiment. This was true even when item-specific variance was included in analyses. By inserting additional material within subject postmodifiers, reading time difficulty at the verbs increased. These results do not support the notion that complexity is driven by structural rarity or level of embedding. If it was, difficulty should have been most prominent at the onsets of modifying phrases. The majority of variation in reading times here was observed over regions containing verbs.

It should be emphasized that reading is a complex behavior, requiring the coordination of diverse information types and the recruitment of multiple mental operations. Given that many factors can contribute to reading difficulty, the extent to which this simple metric of complexity predicts the direction and magnitude of the observed differences among conditions is extremely impressive.

4. General Discussion

The evidence suggests that the difficulty associated with integrating a new input item in English reading is heavily determined by the length of the material intervening between the input item and the site of its target dependents. Locality is not a specialized heuristic for adjudicating between analyses of an ambiguous string or a by-product of tendencies specific to production, but is a pervasive resource constraint in language processing.

These data undermine a broad class of experience-based models of comprehension difficulty, namely, those that claim complexity is primarily a function of encountering, identifying, or building rare syntactic configurations.[6] These models predict difficulty when infrequent structures are introduced. Instead, readers exhibited difficulty at sites requiring long-distance integrations, which occurred *after* regions containing uncommon configurations. This is consistent with the distance-based model of linguistic resource use given in (5). However, it is not possible to rule out the entire class of statistical models of complexity. In principle, there are an infinite number of statistics (or combinations thereof) that might influence the processing of a given structure. And, to be sure, there are certain cases where frequency of a given form has been shown to be inversely related to comprehension difficulty (e.g., Gennari & MacDonald, 2004; Gibson & Pearlmutter, 1994; Spivey-Knowlton & Sedivy, 1995; Tabor, Tanenhaus, & Juliano, 1997). Even in the previously mentioned studies, the structures that caused the greatest difficulty are probably the least common. For instance, less than 15% of RCs in the Brown and *Wall Street Journal* corpora are object extracted (Rohde, 2002). However, these correlations in and of themselves do not directly support the notion that rare structures are more difficult to process. There are at least two resource-based explanations for why structural distributions in production could parallel comprehension difficulty in this way. First, production processes may have similar capacity limits to perceptual processes. If it is also true that structures that are easier to produce are more common, then the structures that are easier to understand will be more frequent. Second, producers may take into account the needs of compre-

henders, either explicitly or implicitly, and adjust utterances to improve communication (Levelt, 1989; but see Ferreira & Dell, 1996, for evidence that speakers do not always consider listener needs in planning an utterance online).This point is especially germane to research that collects statistics from written texts such as newswire corpora, which are edited for readability. Unlike the comprehension-centered experience-based account, these resource-based accounts offer possible explanations for existing distributions of structures. Nevertheless, the predictions of the resource-based accounts are difficult to separate from an experience-based account solely on the basis of offline correlations between production phenomena and qualitative comprehension difficulty (cf. Gibson & Pearlmutter, 1994). By looking at online comprehension, the experiments here are able to distinguish the resource-based model from what is perhaps the most straightforward application of structural statistics to the problem of complexity. It remains to be seen whether an experience-based model of complexity can be formulated that can account for online comprehension data without appeal to memory-based resource limitations.

This work demonstrates that locality affects integration costs while reading sentences in English. However, it is still not clear whether locality affects all types of structural integrations or just particular ones. For the previously mentioned constructions, the length of both filler–gap dependencies and subject–verb dependencies were manipulated simultaneously. The representations and computations entailed by each are distinct. Thus, it is possible that the two types of representations may differentially be affected by locality. Indeed, Figure 3 suggests that adding PP and RC modifiers affected verb reading times more in the embedded conditions than in the unembedded conditions. The embedded conditions involve a filler–gap dependency, whereas the unembedded do not. Thus it is possible that the bulk of our observed effects are driven by long-distance difficulty in connecting fillers with their gaps. There is also evidence from Japanese that filler–gap dependencies are more sensitive to length manipulations than argument dependencies (such as the subject–verb dependencies investigated here). Miyamoto and Takahashi (2002) investigated locality in scrambled structures and found that increasing the distance between a scrambled constituent and the position in which it is canonically generated increased reading times at the gap site. A parallel manipulation of the length of an object–verb dependency did not reliably affect reading times. Like *wh* dependencies, scrambling is typically analyzed as a structural phenomenon that is not lexically governed. Still, it appears that argument-dependency integrations are affected by distance in English, although perhaps not as dramatically as nonlexical dependencies. For instance, there was a prominent peak in RTs at the main verb for subject-extracted RCs in Experiment 1. Further, Chen, Gibson, and Wolf (2002) showed that the difficulty of integrating a second argument with a ditransitive verb increases with the length of the intervening first object. Understanding how locality affects lexical and nonlexical dependencies is an interesting avenue of future investigation.

The work of Konieczny (2000) and Vasishth (2002) is also relevant to this discussion. These researchers found that reading times were actually *shorter* or unchanged when more material separated a head from its dependent. Such results are in direct contrast to the locality account and superficially seem to contradict the results presented here (see also Nakatani & Gibson, 2003, for similar results from Japanese). However, there were a number of properties of the materials that Konieczny and Vasishth investigated that made them distinct from our materials. Crucially, all of these studies involved verb-final languages—German (Konieczny, 2000),

Hindi (Vasishth, 2002), and Japanese (Nakatani & Gibson, 2003)—rather than English, a verb-initial language. In the relevant materials, the structure that was inserted was an additional dependent of an upcoming head, as in (12):

(12) Subject (optional adverbial dependent of the verb) (object) verb

As mentioned previously, Konieczny (2000) argued that more material between the endpoints of a predicted dependency constrains the properties of the predicted element, making it *easier,* rather than harder, to identify. Although this general account does not comport well with our results from English, it is possible that modifiers that are dependents of the predicted head do constrain its semantic and syntactic attributes. This might permit the processor to preactivate those features and facilitate integration. The effect of locality may then be obscured. This was not a confound in our experiments, because the length of the critical subject–verb dependencies was varied by manipulating subject modifiers. These additional constituents were not dependents of the verb to come, and they did not otherwise serve to narrow the attributes of the verb. As a result, they did not increase the predictability of verb, and locality effects were therefore robust.

Another result of Konieczny (2000) needs to be addressed, given the preceding view. He found that placing a relative clause between an NP and its role-assigning verb actually decreased RTs on the verb, in direct contrast to the results described here. The RC is not a dependent of the upcoming verb. However, it should be noted that intervening constitutents can make a verb easier to interpret without being dependent on that verb. For instance, the sentence in (13b) is more plausible than that in (13a), because the subject modifier provides a basis for the verbal event:

(13a) *The fisherman cried.*
(13b) *The fisherman who was cutting onions cried.*

In our experiment, we specifically controlled plausibility, so that adding a modifier did not increase (or decrease) the plausibility of the resultant NP–verb integration. Such a control was not performed in the Konieczny (2000) items. Thus it is possible that the decrease in RTs at the verb may have been due to increased plausibility of the sentence at the verb when the modifier was included.

Other results have also been used to speak against locality or configural explanations of complexity in general. For instance, several investigators have shown that the complexity difference between subject- and object-extracted RCs can be ameliorated, or even eliminated, by using distinct types of nouns in the matrix and embedded clauses (Gordon et al., 2001), or by manipulating the animacy of the matrix and embedded nouns (Mak et al., 2002; Traxler et al., 2002). These studies serve as important reminders that a variety of factors can affect how difficult it is to integrate a word into a partial structure. Although such factors may, in certain circumstances, mask the contributions of locality, none of these factors are capable of explaining these results, or for that matter, the wide array of processing phenomena that have been attributed to locality (see Gibson, 2000; and Hawkins, 1994, for overviews). On reflection, the fact that locality effects are so ubiquitous is not surprising. It is a natural consequence of extracting hierarchical structure from a serial input stream.

Notes

1. This assumes that the representatives degrade linearly with distance. A more psychologically motviated function might be sigmoidal or logarithmic to reflect human performance in ordered recall tasks.

2. Although this account predicts object-extracted RCs to be more difficult over the embedded clause, it actually seems to predict that the subject-extracted RCs should be more difficult at the main verb. In the latter, the matrix verb is directly preceded by an NP that is not its logical subject. This is a deviation from the SVO surface order of constituents. Thus it is not clear in their study why sentences containing object-extracted clauses should be harder overall. MacDonald & Christiansen need to elaborate on their surface order claim to explain why the object-extracted surface order in (i) a. diverges more overall from SVO than the subject-extracted surface order in (i) b.
   (ia.) S1 S2 V2 V1 O1
   (ib.) S1 V2 02 V1 O1

3. We are indebted to an anonymous reviewer for suggesting this possibility.

4. The complete set of materials employed in Experiments 1 and 2 can be obtained by contacting the first author.

5. One subject was excluded from the participant analysis, and one item was excluded from the items analysis, because they were missing means in the embedded RC condition.

6. This rests on the putative assumption that the parser builds connected structures incrementally when processing unambiguous input.

Acknowledgments

References

Altmann, G. T. M., van Nice, K. Y., Garnham, A., & Henstra, J.A. (1998). Late closure in context. *Journal of Memory and Language, 38,* 459–484.

Anderson, J. R. (1983). *The architecture of cognition.* Cambridge, MA: Harvard University Press.

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

Chen, E., Gibson, E., & Wolf, F. (2002). *Syntactic storage costs in sentence comprehension.* Poster session presented at the 15th Annual CUNY Conference on Human Sentence Processing, New York.

Dell, G. S., Burger, L. K., & Svec, W. R. (1997). Language production and serial order: A functional analysis and a model. *Psychological Review, 104,* 123–147.

Elman, J. L. (1991). Distributed representations, simple recurrent networks and grammatical structure. *Machine Learning, 7,* 195–225.

Ferreira, F., & Clifton, C., Jr. (1986). The independence of syntactic processing. *Journal of Memory and Language, 25,* 348–368.

Ferreira, V. S., & Dell, G. S. (1996, November). *Do speakers choose their words cooperatively? Investigating the production–comprehension interface.* Poster session presented at the 37th Annual Meeting of the Psychonomics Society, Chicago.

Ford, M. (1983). A method for obtaining measures of local parsing complexity throughout sentences. *Journal of Verbal Learning and Verbal Behavior, 22,* 203–218.

Ford, M., Bresnan, J., & Kaplan, R. (1982). A competence-based theory of syntactic closure. In J. Bresnan (Ed.), *The mental representation of grammatical relations* (pp. 727–796). Cambridge, MA: MIT Press.

Frazier, L. (1979). *On comprehending sentences: Syntactic parsing strategies.* Unpublished doctoral dissertation, University of Connecticut. Bloomington, IN: Available from Indiana University Linguistics Club.

Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance: Vol. 12. The psychology of reading* (pp. 559–585). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Frazier, L., & Clifton, C., Jr. (1996). *Construal.* Cambridge, MA: MIT Press.

Frazier, L., & Clifton, C., Jr., (1998). Sentence reanalysis and visibility. In J. D. Fodor & F. Ferreira (Eds.), *Reanalysis in sentence processing* (pp. 143–176). Dordrecht, The Netherlands: Kluwer Academic.

Frazier, L., & Flores d'Arcais, G. (1989). Filler driven parsing: A study of gap filling in Dutch. *Journal of Memory and Language, 28,* 331–344.

Garnsey, S. M., Pearlmutter, N. J., Myers, E., & Lotocky, B. (1997). The relative contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language, 37,* 58–93.

Gennari, S., & MacDonald, M. (2004, March). *Relating production and comprehension of relative clauses.* Paper presented at the 17th Annual CUNY Conference on Human Sentence Processing. College Park, MD.

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). The dependency locality theory: A distance-based theory of linguistic complexity. In Y. Miyashita, A. Marantz, & W. O'Neil (Eds.), *Image, language, brain* (pp. 95–126). Cambridge, MA: MIT Press.

Gibson, E., & Hickock, G. (1993). Sentence processing with empty categories. *Language and Cognitive Processes, 8,* 147–161.

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

Gibson, E., Pearlmutter, N., & Torrens, V. (1999). Recency and lexical preferences in Spanish. *Memory and Cognition, 27,* 603–611.

Gordon, P. C., Hendrick, R., & Johnson, M. (2001). Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 1411–1423.

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

Hakes, D. T., Evans, J. S., & Brannon, L. L. (1976). Understanding sentences with relative clauses. *Memory and Cognition, 4,* 283–290.

Hale, J. (2003). The information conveyed by words in sentences. *Journal of Psycholinguistic Research, 32,* 101–124.

Hawkins, J. A. (1994). *A performance theory of order and constituency.* Cambridge, England: Cambridge University Press

Holmes, V. M., & O'Regan, J. K. (1981). Eye fixation patterns during the reading of relative-clause sentences. *Journal of Verbal Learning and Verbal Behavior, 20,* 417–430.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehensions. *Psychological Review, 87,* 329–354.

Just, M. A., Carpenter, P. A., & Keller, T. A. (1996). The capacity theory of comprehension: New frontiers of evidence and arguments. *Psychological Review, 103,* 773–780.

D. Grodner, E. Gibson/Cognitive Science 29 (2005)

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General, 111,* 228–238.

Kimball, J. (1973). Seven principles of surface structure parsing in natural language. *Cognition, 2,* 15–47.

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language, 30,* 580–602.

Konieczny, L. (2000). Locality and parsing complexity. *Journal of Psycholinguistic Research, 29,* 627–645.

Lashley, K. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), Cerebral mechanisms in behavior. NY: Wiley.

Levelt, W. (1989). *Speaking: From intention to articulation.* Cambridge, MA: MIT Press.

Lewis, R. L. (1993). *An architecturally-based theory of human sentence comprehension.* Unpublished doctoral dissertation, Carnegie Mellon University, Pittsburgh, PA.

Lewis, R. L. (1996). Interference in short-term memory: The magical number two (or three) in sentence processing. *Journal of Psycholiguistic Research, 25,* 93–115.

MacDonald, M., Pearlmutter, N., & Seidenberg, M. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review, 101,* 676–703.

MacDonald, M. C. (1993). The interaction of lexical and syntactic ambiguity. *Journal of Memory and Language, 32,* 692–715.

MacDonald, M. C. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes, 9,* 157–201.

MacDonald, M. C. (1999). Distributional information in language comprehension, production, and acquisition: Three puzzles and a moral. In B. MacWhinney (Ed.), *The emergence of language* (pp. 177–196). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: A comment on Just & Carpenter (1992) and Waters & Caplan (1996). *Psychological Review, 109,* 35–54.

MacDonald, M. C., Just, M. A., & Carpenter, P. A. (1992). Working memory constraints on the processing of syntactic ambiguity. *Cognitive Psychology, 24,* 56–98.

Mak, W. M., Vonk, W., & Schriefers, H. (2002). The influence of animacy on relative clause processing. *Journal of Memory and Language, 30,* 580–602.

Marcus, M. P., Santorini, B., & Marcinkiewicz, M. A. (1993). Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics, 19,* 313–330.

Marslen-Wilson, W., & Tyler, L. K. (1980). Central processes in speech understanding. *Philosophical Transactions of the Royal Society, Series B, 295,* 317–332.

Marslen-Wilson, W., & Tyler, L. K. (1987). Against modularity. In J. L. Garfield (Ed.), *Modularity in knowledge representation and natural-language understanding* (pp. XX–XX). Cambridge, MA: MIT Press.

Miyamoto, E. T., Gibson, E., Pearlmutter, N. J., Aikawa, T., & Miyagawa, S. (1999). A U-shaped relative clause attachment preference in Japanese. *Language and Cognitive Processes, 14,* 663–686.

Miyamoto, E. T., & Takahashi, S. (2002). Sources of difficulty in processing scrambling in Japanese. In M. Nakayama (Ed.), *Sentence processing in East Asian languages* (pp. 167–188). Stanford, CA: Center for the Study of Language and Information.

Nakatani, K., & Gibson, E. (2003, March). *An on-line study of Japanese nesting complexity.* Poster session presented at the 16th Annual CUNY Conference on Human Sentence Processing, Cambridge, MA.

Pearlmutter, N. J., & Gibson, E. (2001). Recency in verb phrase attachment. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 574–590.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. E. (1996). Understanding normal and impaired word reading: Computational principles in quasiregular domains. *Psychological Review, 103,* 56–115.

Rayner, K., & Pollatsek, A. (1995). *The psychology of reading.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Rohde, D. L. T. (2002). *A connectionist model of sentence comprehension and production.* Unpublished doctoral dissertation, Carnegie Mellon University, Pittsburgh, PA.

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

Spivey-Knowlton, M. J., & Sedivy, J. (1995). Resolving attachment ambiguities with multiple constraints. *Cognition, 55,* 227–267.

Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1994). Referential context and syntactic ambiguity resolution. In C. Clifton Jr., L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 415–439). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Spivey-Knowlton, M. J., Trueswell, J. C., & Tanenhaus, M. K. (1993). Context effects in syntactic ambiguity resolution: Discourse and semantic influences in parsing reduced relative clauses. *Canadian Journal of Experimental Psychology, 47,* 276–309.

Stallings, L. M., MacDonald, M. C., & O'Seaghdha, P. G. (1998). Phrasal ordering constraints in sentence production: Phrase length and verb disposition in heavy-NP shift. *Journal of Memory and Language, 39,* 392–417.

Stevenson, S. (1994). Competition and recency in a hybrid network model of syntactic disambiguation. *Journal of Psycholinguistic Research, 23,* 295–322.

Stromswold, K., Caplan, D., Alpert, N., & Rauch, S. (1996). Localization of syntactic comprehension by positron emission tomography. *Brain and Language, 52,* 452–473.

Sturt, P., Scheepers, C., & Pickering, M. (2002). Syntactic ambiguity resolution after initial misanalysis. *Journal of Memory and Language, 46,* 371–390.

Tabor, W., Juliano, C. and Tanenhaus, M. K. (1997). Parsing in a dynamical system: An attractor-based account of the interaction of lexical and structural constraints in sentence processing. *Language and Cognitive Processes. 12*(2/3), 211–271.

Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268,* 1632–1634.

Tanenhaus, M. K., & Trueswell, J. C. (1995). Sentence comprehension. In J. L. Miller & P. D. Eimas (Eds.), *Speech, language and communication* (pp. 217–262). San Diego, CA: Academic.

Thornton, R., & MacDonald, M. C. (1997). *The role of phrase length in modification ambiguities.* Manuscript in preparation.

Townsend, D. J., & Bever, T. G. (2001). *Sentence comprehension.* Cambridge, MA: MIT Press.

Traxler, M. J., Morris, R. K., & Seely, R. E. (2002). Processing subject and object relative clauses: Evidence from eye movements. *Journal of Memory and Language, 47,* 69–90.

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic disambiguation. *Journal of Memory and Language, 33,* 285–318.

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

Vasishth, S. (2002, September). *Distance effects or similarity-based interference? A model comparison perspective.* Poster session presented at the 8th Annual Conference on Architectures and Mechanisms for Language Processing, Tenerife, Canary Islands, Spain.

Vosse, T., & Kempen, G. (2000). Syntactic structure assembly in human parsing: A computational model based on competitive inhibition and a lexicalist grammar. *Cognition, 75,* 105–143.

Wanner, E., & Maratsos, M. (1978). An ATN approach in comprehension. In M. Halle, J. Bresnan, & G. Miller (Eds.), *Linguistic theory and psychological reality* (pp. 119–161). Cambridge, MA: MIT Press.

Warren, T., & Gibson, E. (2002). The influence of referential processing on sentence complexity. *Cognition, 85,* 79–112.

Wasow, T. (1997). End-weight from the speaker's perspective. *Journal of Psycholinguistic Research, 26,* 347–361.

Waters, G. S., Caplan, D., & Hildebrandt, N. (1987). Working memory and written sentence comprehension. In M. Coltheart (Ed.), *Attention and performance: Vol. 12. The psychology of reading* (pp. 531–555). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Appendix A. Reading Times in Experiment 1

Table A1
Reading times from Experiment 1 in msec

Word	Object Extracted	Subject Extracted
1. The	307.2 (12)	309.6 (11)
2. reporter	373.1 (19)	360.2. (17)
3. who	343 (14)	349.8 (15)
4. the\sent	348.1 (17)	354.8 (16)
5. photographer\the	357.6 (17)	334.3 (13)
6. sent\photographer	422.1 (27)	384 (26)
7. to	375.8 (16)	346.5 (14)
8. the	338.6 (14)	318.4 (12)
9. editor	482.9 (42)	403.6 (30)
10. hoped	401.1 (22)	404.6 (126)
11. for	366 (18)	356.9 (17)
12. a	349.31 (15)	352.7 (16)

    Note. Standard errors given in parentheses.

Appendix B. Reading Times in Experiment 2

Table B1
Reading times per word for Experiment 2

Word Category	Matrix			Embedded		
	Bare	PP	RC	Bare	PP	RC
The	319 (9.4)	318 (7.0)	325 (9.9)	323 (9.5)	322 (8.9)	322 (8.9)
noun	364 (15.2)	375 (13.0)	371 (11.8)	389 (23.3)	362 (11.3)	414 (20.3)
who				396 (34.8)	339 (11.1)	352 (10.1)
the				345 (12.6)	349 (15.8)	356 (23.3)
noun				373 (13.8)	350 (13.6)	404 (23.5)
who			365 (14.3)			370 (17.5)
was			315 (7.0)			459 (29.8)
prep		340 (9.8)	311 (9.1)		377 (16.0)	366 (17.3)
det		293 (6.6)	295 (6.7)		313 (8.2)	319 (12.7)
noun		323 (12.0)	301 (7.8)		361 (34.7)	352 (18.7)
verb				448 (37.3)	466 (40.8)	588 (64.0)
verb	375 (15.2)	393 (19.6)	389 (17.6)	500 (39.7)	558 (43.8)	538 (50.3)
det	339 (9.6)	370 (16.8)	352 (10.5)	439 (30.4)	432 (23.5)	438 (29.1)
noun	400 (33.2)	335 (12.6)	405 (29.8)	386 (19.2)	385 (20.6)	400 (19.8)
connective	395 (30.4)	377 (13.4)	379 (17.5)	407 (22.9)	396 (27.2)	399 (15.6)

    Note. PP = prepositional phrase; RC = relative clause. Standard errors given in parentheses.