Journal of Experimental Psychology:
Learning, Memory, and Cognition
2001, Vol. 27, No. 2, 574-590

0278-7393/01/$5.00 DOI: 10.1037//0278-7393.27.2.574

Recency in Verb Phrase Attachment

Neal J. Pearlmutter
Northeastern University

Edward Gibson
Massachusetts Institute of Technology

Four experiments investigated attachment preferences in constructions involving 3 verb phrases (VPs) followed by an attaching modifier. Readers preferred attachment to the most recent (lowest) VP site overall and preferred to attach the modifier to the middle VP over the highest VP, demonstrating a monotonic recency-based preference ordering. This pattern could not be attributed to lexical or plausibility-based preferences. The results contrast with the pattern for relative clause attachment into 3 potential noun phrase sites, where the preference ordering is nonmonotonic (e.g., E. Gibson, N. J. Pearlmutter, E. Canseco-Gonzalez, & G. Hickok, 1996), and support the multiple-constraint theory described by E. Gibson and N. J. Pearlmutter (1998), which proposes that recency/locality and a secondary factor, predicate proximity, combine with lexical, grammatical, prosodic, and contextual constraints to determine attachment preferences.

Probably the most fundamental property of all human languages from the perspective of comprehension and production processes is *seriality*: Linguistic utterances are generated and must be comprehended as a sequence of elements ordered in time. This property forces the comprehension system to maintain partial representations of an utterance in memory as the utterance is being processed. Given the assumption that memory representations tend to weaken over time (e.g., Anderson, 1983; Just & Carpenter, 1992), more recently processed parts of an utterance will tend to be more easily accessed, and one result of this will be a preference to connect new input to more recently rather than less recently processed material.

Such a recency effect can be seen in speakers' intuitions for some kinds of ambiguous phrase attachments. In Examples 1a and 1b, an incoming word can attach to either a recent or nonrecent verb phrase (VP) attachment site. In Example 1a, the adverb *yesterday* can modify either the more or less recent verb (*explained* or *said,* respectively); and in Example 1b, the particle *up* can attach to either *shook* or *smashed.* In both examples, speakers typically prefer the more recent attachment.

1a. Rosencrantz said that Guildenstern explained Tom's play yesterday.

1b. William shook the friend who had smashed his new cart up.

These intuitions have been captured in a variety of theories (e.g., late closure, Frazier, 1979; recency preference, Gibson, Pearlmutter, Canseco-Gonzalez, & Hickok, 1996; locality, Gibson, 1998; right association, Kimball, 1973; see also Fodor, 1998; Frazier & Fodor, 1978; Lewis, 1996; MacDonald, 1999; Stevenson, 1994), and Altmann, van Nice, Garnham, and Henstra (1998) demonstrated a recency effect on-line using examples similar to Example 2 (see also Igoa, Carreiras, & Meseguer, 1998, for related evidence from Spanish).

2. Glenn and Gary will sell the real estate Ross owned yesterday/tomorrow.

As in Example 1a, the temporal adverbs *yesterday* and *tomorrow* must attach to one of the two VPs in Example 2. However, the VP sites differ in tense marking (future for *will sell,* past for *owned*), and each adverb can only grammatically attach to the site with compatible tense marking. Altmann et al. (1998) showed that in a null context readers had more difficulty reading the adverb that had to attach nonrecently (*tomorrow*) than the one that attached recently (*yesterday*). A strongly biasing preceding context could mitigate the difficulty of the nonrecent attachment, but it could not reverse the recency preference.

Another ambiguity claimed to demonstrate a recency preference is shown in Example 3, where comprehenders prefer to treat *the dialogue* as the direct object of *writes,* rather than as the subject of the upcoming matrix clause.

3. Whenever David writes the dialogue comes out strikingly odd.

Neal J. Pearlmutter, Psychology Department, Northeastern University; Edward Gibson, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology.

Parts of this research were supported by National Science Foundation Grants BCS-9729007 and BCS-9729037. Portions of this work were presented at the 1996 Psychonomic Society Conference, Chicago, and at the 1997 City University of New York Sentence Processing Conference, Santa Monica, California.

We greatly appreciate the comments and advice of Marc Brysbaert, Barbara Hemforth, Don Mitchell, and two anonymous reviewers on earlier versions of this article. Thanks are also due to Lyn Frazier, who originally noted the prediction on which this work is based. We thank Collin Chang, Ian Kelly, Joanne Malinka, Aurora Mendelsohn, Kimberly Mitchell, Joey Pasquino, Patricia Pimentel, Shelagh Scollin, Jennifer Turco, and San Tunstall for aid in constructing materials, conducting the experiments, and analyzing data. We also appreciate the work of Tom, Dave, Will, and Art in providing the framework for some examples.

Correspondence concerning this article should be addressed to Neal J. Pearlmutter, Psychology Department, 125 NI, Northeastern University, Boston, Massachusetts 02115. Electronic mail may be sent to pearlmutter@neu.edu.

This ambiguity, like others, is sensitive to numerous factors including lexical preferences, prosody, and discourse biases, but when these are controlled there is a strong preference to attach to the immediately preceding VP (Clifton, 1993; Ferreira & Henderson, 1991; Frazier & Rayner, 1982; Kennedy & Murray, 1984; Kjelgaard & Speer, 1999; Mitchell, 1987; Mitchell & Holmes, 1985; Speer, Kjelgaard, & Dobroth, 1996; Stowe, 1989; Sturt, Pickering, & Crocker, 1999; Warner & Glass, 1987). This is a recency preference under the assumption that the ambiguity involves a choice between attachment to the embedded VP just created for the preceding verb (*writes*) and attachment into the matrix clause created several words earlier, when the subordinating conjunction *whenever* was processed (e.g., Frazier, 1979; cf. Gorrell, 1995; Pritchett, 1988; for alternative interpretations).

Whereas attachment ambiguities involving VP attachment sites are associated with strong recency biases, the same is not always true of attachment ambiguities involving noun phrase (NP) attachment sites. For example, less recent attachment is preferred in Spanish constructions like Example 4, where the relative clause (RC) *que se observó desde el satélite* (*that was observed from the satellite*) can either attach nonrecently (high in the syntactic structure) to *órbita* (*orbit*) or recently (low) to *planeta* (*planet*) (e.g., Carreiras & Clifton, 1993; Cuetos & Mitchell, 1988; Gibson, Pearlmutter, & Torrens, 1999; Mitchell & Cuetos, 1991). Gibson et al. (1999) found that Spanish comprehenders read the early part of the RC in Example 4 more quickly when it was disambiguated toward high attachment (*órbita*) than when it was disambiguated toward low attachment (*planeta*).

4. El astrónomo predijo la órbita del planeta que se observó desde el satélite. (The astronomer predicted the orbit of the planet that was observed from the satellite.)

However, the attachment of RCs to NP sites is more complex than just an antirecency preference. When a third potential NP attachment site is added to sentences like Example 4, as in Example 5, the overall attachment preference shifts to the most recent NP site *planeta* (Gibson, Pearlmutter, et al., 1996; Gibson et al., 1999). Furthermore, the next easiest attachment is to the highest NP site *cambio* (*change*), and attachment to the intermediate site *órbita* is the most difficult of the three possibilities.

5. El astrónomo predijo el cambio de la órbita del planeta que se observó desde el satélite. (The astronomer predicted the change of the orbit of the planet that was observed from the satellite.)

One additional fact adds more complexity to these results: Although the pattern of RC attachment preferences varies cross-linguistically in the 2NP construction (see Mitchell & Brysbaert, 1998, for a recent overview of evidence from English, Spanish, and a variety of other languages), the on-line pattern for RC attachments in the 3NP construction appears to be low/high/middle for all the languages that have been tested thus far: English (Gibson, Pearlmutter, et al., 1996), Spanish (Gibson, Pearlmutter, et al., 1996; Gibson et al., 1999), German (Walter & Hemforth, 1998), and Japanese (Miyamoto, Gibson, Pearlmutter, Aikawa, & Miyagawa, 1999).[1]

To account for the differing preferences in 2NP and 3NP constructions, Gibson, Pearlmutter, et al. (1996) proposed that a second factor, predicate proximity, interacts with recency. According to their proposal, recency specifies an increasing cost associated with increasingly distant attachments, whereas predicate proximity specifies a fixed cost associated with attachments that are not as close as possible to a predicate phrase (typically a VP). In Spanish, predicate proximity is strong enough to dominate in 2NP RC attachment ambiguities, but recency dominates for more extended ambiguities, involving three or more sites. In English, predicate proximity is weaker than recency even in 2NP RC attachment ambiguities, so recent attachment is preferred even with only two sites. However, predicate proximity applies in both languages, as evidenced by the low/high/middle preference for the 3NP cases in each language.[2]

Alternatives to predicate proximity have been proposed for the factor interacting with recency to create the pattern of 2NP and 3NP RC attachment preferences. One proposal, which predates the evidence from 3NP constructions, is relativized relevance, which states that comprehenders, "other things being equal, preferentially construe a phrase as being relevant to the main assertion of the sentence" (Frazier, 1990, p. 321; see Frazier & Clifton, 1996, for a full presentation of the system within which this factor applies). In the sequences of right-branching NPs in the 2NP and 3NP constructions, the first NP is structurally closest to the main assertion of the sentence, which is the main predicate. Thus, relativized relevance can predict the same pattern as predicate proximity in RC attachments. A related alternative relies on discourse prominence, such that attaching phrases are attracted to the most prominent elements in the discourse (Hemforth, Konieczny, & Scheepers, 1999; Hemforth, Konieczny, Scheepers, & Strube, 1998; cf. Frazier & Clifton, 1996; Gilboy, Sopena, Clifton, & Frazier, 1995; Schafer, Carter, Clifton, & Frazier, 1996). According to this proposal, the arguments of a verb are usually prominent in the discourse (more so than modifiers of the arguments), causing a preference for the first attachment site. Under Hemforth et al.'s (1998, 1999) version of this proposal (the anaphoric binding hypothesis), this factor applies only to pronominal elements, such as the relative pronoun initiating an RC, which needs to be coindexed with a preceding NP. A third alternative, discussed by Gibson, Pearlmutter, et al. (1996), is primacy preference, a preference to attach to the first available site in a sequence, which might be justified from general properties of short-term memory (in particular, the serial position effect in recall; e.g., Murdock, 1962; Tulving, 1968). Although Gibson, Pearlmutter, et al. (1996) dis-

[1] Wijnen (1998, 1999) provides evidence suggesting that the preference pattern in Dutch may be high/low/middle, but, critically, the middle site is still clearly least preferred.

[2] Preferences in RC attachment ambiguities have also been shown to be influenced by a variety of other factors independent from recency and predicate proximity, such as differences in the prepositions linking the sites (e.g., Gilboy, Sopena, Clifton, & Frazier, 1995) and the extent to which the sites have already been modified (e.g., Thornton, MacDonald, & Gil, 1999). Gibson (1998) describes a discourse-entity–based formulation of recency/locality that can account for the effects of different site-linking prepositions. Modifiability effects on our view simply arise from an additional independent referential context constraint. The multiple-constraint theory described in Gibson and Pearlmutter (1998), which includes recency and predicate proximity as two constraints among a larger set, is compatible with this view. Our approach here is to attempt to isolate and examine the influences of recency and predicate proximity in particular.

missed primacy preference as an unlikely explanation for theoretical reasons—essentially, potential attachment site sequences are unlikely to be identifiable prior to encountering the attaching category—there is no empirical evidence against it.[3]

This article investigates an interesting prediction of the two-factor theories in preferences involving VP attachment sites. Whereas all of the two-factor theories predict a nonmonotonic preference ordering of sites for NP attachments (low/high/middle), these theories differ in their predictions with respect to attachment preferences into three preceding VP sites, as in Example 6, where the adverbial phrase *last week* must modify the VP headed by one of *noted, confirmed,* or *confessed.*

6. The judge noted that the guard confirmed that the prisoner confessed the crime last week when testimony continued in the trial.

All of the theories predict that the most recent and lowest site (*confessed*) will be preferred, but the theories differ on their predictions with respect to the relative preference for the high and middle sites (*noted* and *confirmed,* respectively). If the factor interacting with recency is predicate proximity, then middle attachment should be favored over high attachment, resulting in the monotonic preference ordering (low/middle/high). Predicate proximity applies equally to all three VPs (each is a predicate phrase); as a result, preferences are determined by recency. In contrast, relativized relevance predicts that high attachment should be preferred over middle attachment, because the high VP is the main assertion in the sentence. Thus, relativized relevance predicts the same nonmonotonic preference ordering for VPs as for NPs (low/high/middle). The same nonmonotonic ordering is predicted if the interacting factor is the short-term memory-based primacy preference.

Like predicate proximity, Hemforth et al.'s (1998) anaphoric binding factor combines with recency to predict the monotonic preference ordering (low/middle/high), although for a different reason. Because there is no anaphor to be bound in an adverbial phrase, anaphoric binding does not apply, and recency determines the preferences. If there were an anaphor to be bound in the adverbial phrase, then the preference ordering would be determined by recency combined with the more prominent VP in the discourse (Hemforth et al., 1998, 1999). However, it is unclear which VP attachment site would be favored in such a case, because a full specification of discourse prominence remains to be worked out. If the first clause in a sentence is more prominent in the discourse than the second, then a nonmonotonic preference ordering (high over middle) is predicted. Otherwise, the same monotonic ordering is predicted as with predicate proximity (low/middle/high).

The relative preference for high versus middle attachment in the construction in Example 6 has not been examined experimentally in the literature. Frazier and Fodor (1978, p. 301) observed that the intuitive preference ranking of the two sites is not clear. Gordon (1982) reported an intuition that high attachment is preferred to middle attachment in constructions like Example 6 but provided no experimental data to support this claim. Corpus frequencies gathered by Gibson and Loomis (1994) demonstrated that middle attachment is more frequent than high attachment in adverbial attachments to one of three preceding VPs. But as observed by Gibson, Schütze, and Salomon (1996) and Gibson and Schütze (1999), corpus frequencies do not always correlate with comprehension preferences, so such evidence is at most suggestive.

Our experiments evaluate the relative preference for high versus middle attachment in constructions like Example 6, with three preceding VPs. Experiment 1 uses an off-line questionnaire, and Experiments 2–4 use self-paced reading to determine the attachment preferences.

Experiment 1

Experiment 1 examined whether readers' preferred attachment of an adverbial phrase is to the highest or to the middle VP site when the lowest (and most recent) site is not available. The critical difference among the two-factor theories in a construction involving three potential VP attachment sites is the relative preference for the high versus the middle site. If the factor interacting with recency is predicate proximity (Gibson, Pearlmutter, et al., 1996) or anaphoric binding (Hemforth et al., 1998, 1999), then middle attachment should be preferred. Alternatively, if the factor interacting with recency is relativized relevance (Frazier, 1990), primacy (Gibson, Pearlmutter, et al., 1996), or possibly discourse prominence, then high attachment should be preferred over middle attachment.

Method

Participants. Fourteen Massachusetts Institute of Technology (MIT) undergraduates and 18 Northeastern University undergraduates participated for $5. All were native English speakers.

Materials and design. Sixteen stimulus pairs like that shown in Example 7 were constructed. The stimuli always began with a subject NP (e.g., *the judge* in Example 7), followed by either a future auxiliary (always *will* or *would*) or a past or present perfective auxiliary (*had, has,* or *have*). A verb (e.g., *note*) that took a sentential complement was next. The sentential complement always began with *that* followed by a subject NP (e.g., *the guard*), another future or perfective auxiliary, a different complement-taking verb (e.g., *confirm*), and another sentential complement. The second sentential complement had the same structure as the first, except that the auxiliary was always past or present perfective and the verb was followed by an object NP rather than another sentential complement. A two-word future adverbial phrase (e.g., *next week*) occurred after the object NP, and the rest of the sentence further modified the adverbial (e.g., *when testimony continues in the trial* in Example 7). All 16 sentence pairs are shown in Appendix A.

7a. The judge will note that the guard has confirmed that the prisoner has confessed the crime next week when testimony continues in the trial.

[3] A different approach to explaining the attachment preferences is the *tuning* or exposure-based proposal of Mitchell and colleagues (see Mitchell & Brysbaert, 1998; Mitchell, Cuetos, Corley, & Brysbaert, 1995). On this proposal, people tabulate the resolutions of ambiguities when they are encountered, with the result that people prefer to resolve ambiguities with the resolution most frequently encountered in the past. However, although some component of attachment preferences may depend on exposure to similar structures, attachment preferences and corresponding corpus frequencies fail to correlate for some 3NP attachment ambiguities (Gibson & Schütze, 1999; Gibson, Schütze, & Salomon, 1996) as well as for some 2NP cases (Mitchell & Brysbaert, 1998). Thus, other principles such as recency and a factor like predicate proximity are probably also involved.

7b. The judge has noted that the guard will confirm that the prisoner has confessed the crime next week when testimony continues in the trial.

The two versions of each pair differed in whether the high or the middle VP was in the future tense and was therefore the appropriate site for attachment of the future adverbial. In Example 7a, because the highest VP (*will note*) is in future tense and the other two (*has confirmed* and *has confessed*) must be interpreted as occurring in the past, the adverbial *next week* must attach to the highest site. In Example 7b, on the other hand, the highest and lowest sites (*has noted* and *has confessed*) must refer to past events, and thus the only site available for attachment of *next week* is the middle one (*will confirm*).

Thirty filler items were also constructed. They were similar in overall length and number of clauses (four to five on average) to the 16 experimental stimuli, and they contained a variety of structures, including right- and center-embedded relative clauses, complement clauses, and prepositional phrases (e.g., *Ella's comments did not imply that the fact that the programmer will be fired would be announced; The new puppy frolicked with the cat that chased the toy that was dragged across the floor by the boy*). None of them contained the same ambiguity as the experimental stimuli. The 30 fillers were combined with the 16 experimental items to form two randomly ordered five-page presentation lists. Each list contained exactly one version of each experimental item. The pages of each list were separately randomized for each participant.

Procedure. Participants completed the lists by hand. They were instructed to rate each item for how difficult it was to understand, by circling a number on a scale printed below each item in the list. The scale ranged from 1 (labeled *good/easy*) to 5 (labeled *bad/hard*). The instructions stressed that participants should rate each item on the basis of their first impression and gave several examples.

Results and Discussion

The mean difficulty rating for the high attachment condition was 3.04, and the mean rating for the middle attachment condition was 2.73. These values differed reliably, $t_1(31) = 3.52$, $p < .01$; $t_2(15) = 2.27$, $p < .05$, indicating that readers preferred to attach to the middle site over the high site. This result suggests that the influence of recency is not mitigated by a secondary factor in attachment ambiguities involving preceding VP sites, in contrast to cases involving attachment to NP sites. This pattern of results is predicted by the interaction of recency with either predicate proximity or anaphoric binding, but not by other secondary factors that have been proposed to interact with recency. In particular, the preference for middle over high attachment contradicts the predictions made by relativized relevance, primacy, and discourse prominence. These alternatives predict a preference to attach to the high site (the main predicate, the first predicate, and the most prominent discourse event) over the middle one, just as they predict a preference to attach to the first NP in a series of NP attachment sites.

Although the result from Experiment 1 provides some evidence, the questionnaire task was off-line, so the results might not reflect initial preferences. Experiment 2, therefore, made use of a word-by-word reading time measure to examine attachment preferences on-line. In addition, Experiment 1 was not designed to test the preference for the most recent (low) attachment site relative to the other two. According to all of the theories considered, attachment to the most recent site should be preferred, but this should be confirmed empirically. Furthermore, the construction used in Experiment 1 involved switching the tense of embedded clauses from past to future or from future to past. These sentences thus required a complex discourse event structure, and intuitions suggest that they are quite difficult to understand. Examining attachment preferences in constructions that are easier to process overall would therefore provide useful additional data. Finally, as noted earlier, Hemforth et al.'s (1998, 1999) anaphoric binding hypothesis made the same predictions as predicate proximity in Experiment 1, because the attaching adverbial phrase contained no anaphor that needed to be bound. Investigating a case involving anaphor binding would therefore provide further evidence about the anaphoric binding hypothesis, allowing consideration of its use of discourse prominence and possibly differentiating its predictions from those of predicate proximity. Experiment 2 was designed with all of these goals in mind.

Experiment 2

Experiment 2 continued to make use of ambiguous modifier attachments into one of three possible VPs. However, unlike in Experiment 1, the ambiguously attached modifier was an infinitival purpose clause rather than an adverbial phrase, as shown in Example 8. The purpose clause *to ensure himself a place on the advisory board* must modify one of the three VPs (headed by *reported, announced,* or *patented*), and the attachment was disambiguated by the number marking and gender marking on *himself,* which must agree with the subject NP of the modified VP. In Example 8, because only *salesman* is singular and masculine, the purpose clause must attach to the VP headed by *patented.* Because the disambiguation did not depend on tense marking, the VP sites of Experiment 2 were all in simple past tense, making the sentences overall easier to understand than in Experiment 1.

8. The anchorwoman reported that the investors announced that the salesman patented the device to ensure himself a place on the advisory board.

In addition, infinitival purpose clauses are typically assumed to be initiated by a phonologically null pronominal (preceding the inflection marker *to;* e.g., Chomsky, 1981), which needs to be coindexed with the subject NP of one of the preceding VPs as part of the attachment process. Thus, Experiment 2 will also permit examination of Hemforth et al.'s (1998, 1999) anaphoric binding factor and its reliance on discourse prominence in some detail: Unlike in Experiment 1, anaphoric binding should apply, and the attachment preference pattern should be influenced by discourse prominence, because the pronominal will preferentially be bound by the most prominent available referent. Alternatively, the null pronominal in infinitival purpose clauses might turn out not to be bound by the same process as that which applies to relative pronouns in RC attachment. In this case, under an anaphor binding approach like Hemforth et al.'s (1998, 1999), Experiment 2 can help to specify the behavior of the process that handles the pronominals in infinitival purpose clauses in particular. We will return to these possibilities in the Discussion section.

Method

Participants. Fifty-eight MIT students participated for $8 each. All were native English speakers, and none had participated in Experiment 1.

Materials and design. Eighteen sentence triples like the one shown in the top panel of Table 1 were constructed. These stimuli were similar to

Table 1
Example Stimulus Set for Experiments 2–4

Experiment	Attachment	Text and comprehension question
2–4	High	The anchorwoman reported that the investors announced that the salesman patented # the device / to ensure herself a place on the advisory board. Was it the anchorwoman who tried to get onto the advisory board?
2–4	Middle	The anchorwoman reported that the investors announced that the salesman patented # the device / to ensure themselves a place on the advisory board. Was it the investors who tried to get onto the advisory board?
2–4	Low	The anchorwoman reported that the investors announced that the salesman patented # the device / to ensure himself a place on the advisory board. Was it the salesman who tried to get onto the advisory board?
3	High	The investors announced that the anchorwoman reported that the salesman patented the device / to ensure themselves a place on the advisory board. Was it the investors who tried to get onto the advisory board?
3	Middle	The investors announced that the anchorwoman reported that the salesman patented the device / to ensure herself a place on the advisory board. Was it the anchorwoman who tried to get onto the advisory board?

Note. The slash indicates the position of the line break in Experiments 2 and 3; the pound sign (#) indicates the position of the line break in Experiment 4.

those in Experiment 1 in that they began with a matrix clause and two successively embedded sentential complements. However, instead of a following adverbial phrase, the Experiment 2 stimuli continued with an infinitival purpose clause (e.g., *to ensure herself a place on the advisory board* in the first example in Table 1), which had to modify one of the preceding VPs. The third word of the purpose clause was always a number- and/or gender-marked anaphor (e.g., *herself*), which had to agree with the subject NP of the VP to which the purpose clause attached. The number marking and gender marking on the anaphor were varied (*herself* vs. *themselves* vs. *himself* in Table 1) and the subject NPs were selected so that attachment was always forced to exactly one of the different VP sites depending on the anaphor. All of the potential attachment sites were in simple past tense. Appendix B provides a complete list of the Experiment 2 stimuli.

The stimuli were constructed and placed into three lists to balance the appearance of plural (*themselves*), neuter (*itself*), feminine (*herself*), and masculine (*himself*) anaphors across the three attachment conditions. The stimulus lists also contained 84 fillers of various structural types, including 15 fillers that were similar to the experimental items but without an attaching purpose clause (e.g., *The authorities concluded that the soldiers figured that the spy might warn the enemy before the infantry shipped out*) and 7 that included an anaphor (e.g., *The chemist realized that the formula described how to make the explosive but it contradicted itself in several places*). The remaining 62 fillers had a variety of different structures (e.g., *The orchard which the fence surrounded contained peach and apple trees of numerous varieties; The alley which the van stolen by the thief was found in has been the site of various crimes*). Each filler and experimental item also had a yes/no comprehension question. For the experimental items, the question always asked about the referent of the anaphor. The question is shown with each version for the example stimulus set in Table 1. Comprehension questions for the other experimental stimuli were similar.

Apparatus and procedure. An Apple Macintosh Centris personal computer presented the stimuli and collected all data. Participants read nine initial practice items followed by one of the 102-item lists in a random order. The experimental stimuli were displayed on two lines with the line break always just before the first word of the purpose clause (*to*), and the fillers were displayed similarly. The stimuli were presented using a noncumulative word-by-word self-paced moving-window procedure (Just, Carpenter, & Woolley, 1982). At the beginning of a trial, an item was displayed on the screen with all nonspace characters replaced by dashes. When the participant pressed the space bar, the first word of the item was displayed, replacing the corresponding dashes. When the participant pressed the space bar a second time, the first word reverted to dashes, and the second word was displayed in place of the appropriate dashes. Each subsequent press of the space bar revealed the next word and removed the previous word. Pressing the space bar on the last word of the item caused the item to be replaced by its yes/no comprehension question, which the participant answered by pressing one of two keys above the space bar on the keyboard. The computer recorded the time between each button-press as well as the comprehension question response and presented feedback about the participant's answer to the question. Most participants completed the experiment in approximately 30 min.

Results

Comprehension question performance. Comprehension performance was 68% correct for the high attachment condition, 71% correct for the middle attachment condition, and 82% correct for the low attachment condition. Each of these was reliably better than chance, all $t_1(57) > 7.2$, $SD < 22$, $ps < .001$; $t_2(17) > 4.5$, $SD < 21$, $ps < .001$.

One-way analyses of variance (ANOVAs) conducted separately for participants (F_1) and items (F_2; Clark, 1973) revealed an effect of attachment, although this was reliable only by participants, $F_1(2, 114) = 12.25$, $MSE = 262$, $p < .001$; $F_2(2, 34) = 3.13$, $MSE = 313$, $p < .10$. Paired comparisons indicated that comprehension was better in the low than in the high attachment condition, $F_1(1, 57) = 21.14$, $MSE = 272$, $p < .001$; $F_2(1, 17) = 5.31$, $MSE = 330$, $p < .05$, and comprehension was better in the low than in the middle attachment condition by participants only, $F_1(1, 57) = 14.83$, $MSE = 246$, $p < .001$; $F_2(1, 17) = 2.95$, $MSE = 376$, $p < .11$. The middle and high attachment conditions did not differ ($Fs < 1$).

Reading times. To adjust for differences in word length across conditions, for overall differences in participants' reading rates, and for differences in readers' sensitivity to word length, we constructed a regression equation predicting reading time from word length for each participant, using all filler and experimental items (Ferreira & Clifton, 1986; see Trueswell, Tanenhaus, & Garnsey, 1994, for discussion). At each word position, the reading

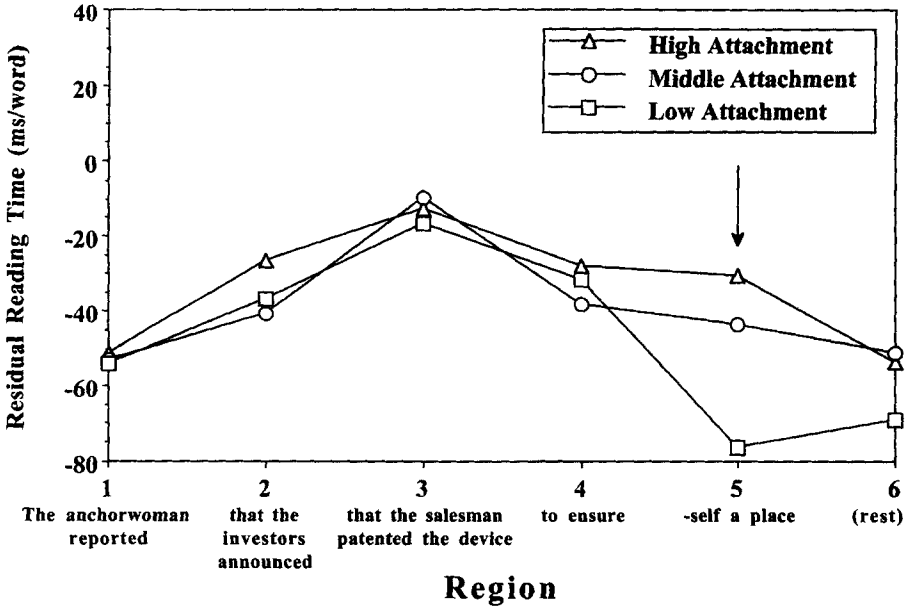

Figure 1. Experiment 2 trimmed residual reading time per word by region. Region 5 (marked with the arrow) is the three-word disambiguating region.

time predicted by the participant's regression equation was subtracted from the actual measured reading time to obtain a residual reading time. Residual reading times beyond 2 *SD* from the corresponding Condition × Position cell mean (i.e., collapsing across participants and items) were excluded, affecting less than 3.6% of the data, and all analyses were conducted on the resulting data set.[4] Appendix C reports the raw reading times trimmed at 2 *SD*. Patterns in the raw data were essentially numerically identical to those in the residual data set but were statistically weaker, particularly in analyses by items.

Figure 1 shows trimmed residual reading times by condition. For analysis purposes, the words were grouped into regions as shown in Figure 1. Region 1 contained the initial subject NP and the first verb; Region 2 contained the following complementizer, the second subject NP, and the second verb; and Region 3 contained the second complementizer, the third subject NP, and the third VP. Region 4 consisted of the first two words of the infinitival purpose clause; and Region 5, the disambiguation, contained the anaphor and the following two words.[5] The remainder of the sentence was Region 6, except for the final word, which is not shown in Figure 1.

None of the one-way ANOVAs in the four regions preceding the disambiguation revealed any reliable effects, although in Region 2 (the second subject NP and verb) the analysis by participants was marginal, $F_1(2, 114) = 2.63$, $MSE = 1132$, $p < .10$; $F_2(2, 34) = 2.23$, $MSE = 441$, $p = .12$; all other $F_1(2, 114) < 2$, $MSE > 994$, $ps > .20$; $F_2(2, 34) < 1.5$, $MSE > 330$, $ps > .25$. Because the stimuli were identical prior to the disambiguating region, this possible effect is likely to be spurious; it did not reappear in either Experiment 3 or 4, so we will not discuss it further.

At the disambiguation, the three conditions differed reliably, $F_1(2, 114) = 24.56$, $MSE = 1,352$, $p < .001$; $F_2(2, 34) = 24.55$, $MSE = 411$, $p < .001$. Individual mean comparisons revealed that the low attachment condition was faster than both the middle attachment condition, $F_1(1, 57) = 23.43$, $MSE = 1,358$, $p < .001$; $F_2(1, 17) = 27.24$, $MSE = 351$, $p < .001$, and the high attachment condition, $F_1(1, 57) = 41.10$, $MSE = 1,524$, $p < .001$; $F_2(1, 17) = 42.41$, $MSE = 450$, $p < .001$. High attachments were slower than middle attachments as well, although this difference was marginal by items, $F_1(1, 57) = 4.40$, $MSE = 1,173$, $p < .05$; $F_2(1, 17) = 3.78$, $MSE = 433$, $p < .10$.

Discussion

The pattern of difficulty at the disambiguating anaphor and the following two words indicates that readers had a clear preference to attach the purpose clause to the lowest (most recent) available verb site. This provides direct on-line evidence for a recency preference in VP attachment and fits with Altmann et al.'s (1998) results for adverbial attachment into two possible verb sites (see also Igoa et al., 1998). It also matches most speakers' intuitions

[4] In this and the following experiments, because of the relatively high comprehension error rates, trials for which the comprehension question was answered incorrectly were not excluded from the analyses. However, the (numerical) pattern of results did not change if such trials were excluded.

[5] Multiword disambiguating regions are common in self-paced reading studies of ambiguity (e.g., Garnsey, Pearlmutter, Myers, & Lotocky, 1997; Gibson et al., 1999). We used a three-word region because Pearlmutter, Garnsey, and Bock (1999) showed that statistically reliable responses to subject–verb agreement violations, similar to the potential anaphor agreement violations in the current stimuli, did not appear until the word following the violation. In the current stimuli, this word was very often a short function word, and effects on such words often spill over onto the next one.

derived from three-site and simpler two-site VP attachment constructions.

In addition to the overall recency preference, readers also displayed a preference for the middle site over the high (least recent) one, providing further support for a strong influence of recency in VP attachment. This on-line pattern matches that measured by off-line ratings in Experiment 1, which used tense marking instead of number agreement and gender agreement to disambiguate and which examined attaching temporal adverbial phrases instead of purpose clauses. In Experiments 1 and 2, the results contrast sharply with preferences for RC attachment to one of three possible NP sites, where recency does have an influence but is partially counterbalanced by a secondary factor supporting attachment to the highest possible site. Experiments 1 and 2 provided no evidence suggesting that a secondary factor was operative. Thus, these results argue against theories of attachment ambiguity in which recency competes with a second factor regardless of the category of the potential attachment sites (e.g., relativized relevance, primacy in short-term memory), indicating that a factor such as predicate proximity, which does differentiate among NP and VP attachment sites, provides a better description.

Experiment 2 also allows a more detailed examination of the predictions of Hemforth et al.'s (1998, 1999) anaphoric binding factor. There are three possibilities to consider: (a) Infinitival purpose clauses are not initiated by an anaphoric element; (b) such clauses are initiated by a null pronominal, and it behaves like the relative pronoun initiating an RC; or (c) such clauses are initiated by a null pronominal, but it is processed differently from anaphors handled by Hemforth et al.'s anaphoric binding factor.

The first possibility seems unlikely: While linguistic theories do vary in terms of how they implement the relationship between infinitival clauses and the clauses to which they attach (e.g., Chomsky, 1981, vs. Pollard & Sag, 1994), all theories must posit some kind of anaphoric process for infinitival clauses, because the verb of the purpose clause has an understood subject that must be coreferential with some other entity in the discourse (see Jones, 1991, for discussion).

On the second possibility—that the anaphoric binding factor should apply to infinitival purpose clauses just as to RCs—anaphoric binding should contribute to attachment preferences in Experiment 2 (unlike in Experiment 1) along with recency. Unfortunately, as noted earlier, the site supported by the anaphoric binding factor depends on discourse prominence, for which there is not currently an explicit theory. However, the Experiment 2 results delimit the alternatives: If discourse prominence applies such that later clauses are more prominent than earlier ones, then the correct monotonic ordering (low/middle/high) will result, because anaphoric binding simply supports recency. Alternatively, if discourse prominence applies such that the matrix clause is most prominent (similarly to relativized relevance, Frazier, 1990), then the incorrect nonmonotonic preference ordering (low/high/middle) will result. Thus, if infinitival purpose clauses engage anaphoric binding, the Experiment 2 data require a discourse prominence theory in which prominence is accorded to the most recent clause in a series of right-branching clauses. At the same time, the data from RC attachments into NP sites require that discourse prominence remains with the least recent NP in a series of right-branching NPs within a clause. These two requirements need not be contradictory, but the theory of discourse prominence relevant for anaphoric binding will need to be further specified.

The third possibility—that purpose clauses contain a null pronominal whose coindexation process does not engage anaphoric binding—is also viable, because the null element to be coindexed for infinitival purpose clauses is not necessarily identical to that for tensed RCs in many syntactic theories. Of course, the process that does apply to the null pronominal in infinitivals would still have to be specified, and, as for discourse prominence discussed earlier, the results of Experiment 2 would require that it create a preference that follows recency for coindexing with the available subject NPs. Whether such a process exists remains an open question.

Experiment 2 thus provides evidence against a variety of possible secondary factors (e.g., relativized relevance, primacy in short-term memory) that might interact with recency to determine preferences, and it is compatible with predicate proximity being the secondary factor interacting with recency. It also allows further specification of the nature of anaphoric binding within Hemforth et al.'s (1998, 1999) theory.

Of course, these conclusions assume that candidate accounts of attachment preferences involve recency interacting with a secondary factor. However, an alternative is available that might potentially predict the Experiment 1 and 2 patterns without appeal to either recency or a competing factor. Work in lexicalist constraint-based theories (e.g., MacDonald, Pearlmutter, & Seidenberg, 1994; Trueswell & Tanenhaus, 1994) has argued that lexically based preferences (e.g., argument structure biases, Garnsey, Pearlmutter, Myers, and Lotocky, 1997; tense morphology biases, Trueswell, 1996) as well as contextual factors (e.g., plausibility, Garnsey et al., 1997; pragmatic reference, Spivey-Knowlton & Sedivy, 1995) control ambiguity resolution for at least some kinds of structural ambiguity. For the constructions examined in Experiments 1 and 2, lexical factors of the sort considered for other ambiguities are not directly relevant, but for both temporal adverbial and purpose clause attachment, differences in the lexical semantics of the verb attachment sites might be relevant. For example, both stative and eventive VPs can be modified by temporal adverbials, as in *Ari hated/slapped Cal last night during the windstorm,* but eventive verbs are probably more likely to be modified. Similarly, purpose clause modification requires that the modified action be volitional (e.g., Mauner, Tanenhaus, & Carlson, 1995; *Henry listened to/?overheard the soldiers to get a sense of their morale*), and thus some verbs, by virtue of their meaning, are probably less likely to be potential sites for purpose clause attachment.

Contextual factors might also be relevant for attachment into VP sites. The range of contextual biases is in principle unlimited, but Altmann et al. (1998) showed that referential pragmatic constraints were not sufficient to explain VP attachment preferences (into two possible sites). We therefore focus on plausibility, which is the other major contextual factor generally considered in ambiguity resolution (e.g., Garnsey et al., 1997; MacDonald et al., 1994 and the references therein). Plausibility can certainly be expected to influence the eventual attachment of a purpose clause, although its use will often depend on the comprehender having processed much of the clause. For example, the purpose clause *to end the feud* is a much more plausible modifier in *Mercutio apologized to Tybalt to end the feud* than in *Mercutio antagonized Tybalt to end the feud.* This will not be apparent until the end of the purpose clause, but

in general, the plausibility associated with attachment of a purpose clause to a particular site should be a function both of the content of the purpose clause and (in our stimuli) the content of the subject NP and verb head of the VP to which the purpose clause attaches. Thus, although our stimuli were constructed to be plausible in each of their different attachment versions, there may nevertheless have been small plausibility differences that resulted in differences in attachment difficulty.

In Experiments 1 and 2, these lexical and contextual possibilities were uncontrolled, and thus it is possible that recency is irrelevant and that the pattern we found resulted from lexical or broader contextual biases. Experiment 3 examined this possibility, focusing on the middle versus high attachment comparison, by exchanging the lexical items instantiating the middle and high sites in Experiment 2. If lexical factors or plausibility were responsible for the relative preference for middle attachment, then the preference should switch to the high site when the lexical items switch. Experiment 3 included both the original Experiment 2 versions and the exchanged versions to gain more power. Thus, when the versions are considered together, any lexically and plausibility based preferences should disappear. If middle attachment is still preferred to high attachment, lexical preferences and plausibility cannot be responsible, although of course they still might have an influence as well.

Experiment 3

Method

Participants. Forty-three Northeastern University undergraduates participated for class credit or $5. All were native English speakers, and none had participated in Experiments 1 or 2.

Materials and design. The Experiment 2 materials were used again in Experiment 3, but two additional versions of each item were created in which the subject NPs, verbs, and complementizers of the first two clauses were exchanged. Table 1 shows all five versions for one item. As in Experiment 2, the number marking or gender marking on the anaphor in the purpose clause disambiguated attachment, so *herself* in the first version in Table 1 forces high attachment to *reported, themselves* in the second version forces middle attachment to *announced,* and *himself* in the third version forces low attachment to *patented.* In the new versions shown in the bottom panel of Table 1, the subject NPs of the first two clauses were exchanged, and therefore *themselves* forces high attachment, even though the purpose clause is still attaching to the VP headed by *announced,* as in the second version. Similarly, *herself* in the bottom panel forces middle attachment to *reported.*

Two additional items were constructed, so that a total of 20 experimental items were used. In addition, because the first two clauses had to be able to exchange, two of the original Experiment 2 items were changed: One complete item was rewritten, and one of the verbs in one of the other items was replaced. Appendix B shows the complete set of stimuli.

A new set of 50 fillers was created, with a variety of structures, that were similar in length and complexity to the experimental items (e.g., *The blanket on the couch against the wall was splattered with paint, and Jake wondered where his annoying roommate had disappeared to; The admissions office knew that the advisor had written that the student had done a good job, but they were having trouble locating the letter).* The fillers and experimental items were combined to form five lists, and yes/no comprehension questions like those in Experiment 2 were written for all stimuli.

Apparatus and procedure. An Apple Macintosh Quadra personal computer was used to present the stimuli and collect data. The same procedure was used as in Experiment 2. Participants saw 10 practice items followed by one of the 70-item lists in a random order and completed the experiment in approximately 25 min.

Results

For analyses, the five versions shown in Table 1 were collapsed into the same three conditions as in Experiment 2: high attachment, middle attachment, and low attachment.

Comprehension question performance. Comprehension performance was 60% correct for the high attachment condition, 65% correct for the middle attachment condition, and 77% correct for the low attachment condition. As in Experiment 2, these values were all different from chance, all $t_1(42) > 3.5$, $SD < 20$, $ps < .001$; $t_2(19) > 2.8$, $SD < 21$, $ps < .05$.

One-way ANOVAs showed a reliable main effect of attachment, $F_1(2, 84) = 9.83$, $MSE = 310$, $p < .001$; $F_2(2, 38) = 6.16$, $MSE = 234$, $p < .01$. Individual mean comparisons revealed a pattern like that in Experiment 2, but somewhat clearer: The high and middle attachment conditions did not differ, $F_1(1, 42) = 1.57$, $MSE = 260$, $p > .20$; $F_2(1, 19) = 1.79$, $MSE = 129$, $p > .15$, but comprehension was reliably better in the low attachment condition than in the other two; low versus high: $F_1(1, 42) = 14.93$, $MSE = 381$, $p < .001$; $F_2(1, 19) = 8.92$, $MSE = 306$, $p < .01$; low versus middle: $F_1(1, 42) = 10.53$, $MSE = 290$, $p < .01$; $F_2(1, 19) = 5.11$, $MSE = 268$, $p < .05$.

Reading times. Residual reading times were computed and trimmed at 2 *SD* as in Experiment 2, affecting less than 3.5% of the data. Figure 2 shows trimmed residual reading times by condition for each region, and Appendix C shows the correspondingly trimmed raw reading times.

As in Experiment 2, there were no reliable effects prior to the disambiguating region, all $F_1(2, 84) < 1.5$, $MSE > 641$, $ps > .30$; $F_2(2, 38) < 1.5$, $MSE > 387$, $ps > .30$. The marginal (by participants) effect at Region 2, from Experiment 2, did not reappear. At the disambiguation (Region 5), the three conditions differed reliably, $F_1(2, 84) = 9.08$, $MSE = 1{,}085$, $p < .001$; $F_2(2, 38) = 12.87$, $MSE = 324$, $p < .001$, and individual mean comparisons revealed that the high attachment condition was read more slowly than the middle attachment condition, $F_1(1, 42) = 9.08$, $MSE = 762$, $p < .01$; $F_2(1, 19) = 11.16$, $MSE = 270$, $p < .01$. Low attachments were faster than both high attachments and middle attachments, although by participants the latter comparison was only marginal; low versus high: $F_1(1, 42) = 13.08$, $MSE = 1{,}488$, $p < .01$; $F_2(1, 19) = 18.46$, $MSE = 445$, $p < .001$; low versus middle: $F_1(1, 42) = 3.15$, $MSE = 1{,}006$, $p < .10$; $F_2(1, 19) = 4.97$, $MSE = 256$, $p < .05$.

Discussion

Like Experiment 2, Experiment 3 indicated that readers had the most difficulty with high attachments, followed by middle attachments, and then low attachments. This was the case despite the control for lexical and plausibility biases created by exchanging the subject NPs and verbs of the high and middle attachment sites. If such biases had been responsible for the Experiment 2 results, then the preference for middle over high attachment should have disappeared in Experiment 3. These results therefore provide further support for theories involving recency and a secondary factor like predicate proximity (e.g., Gibson, Pearlmutter, et al., 1996)

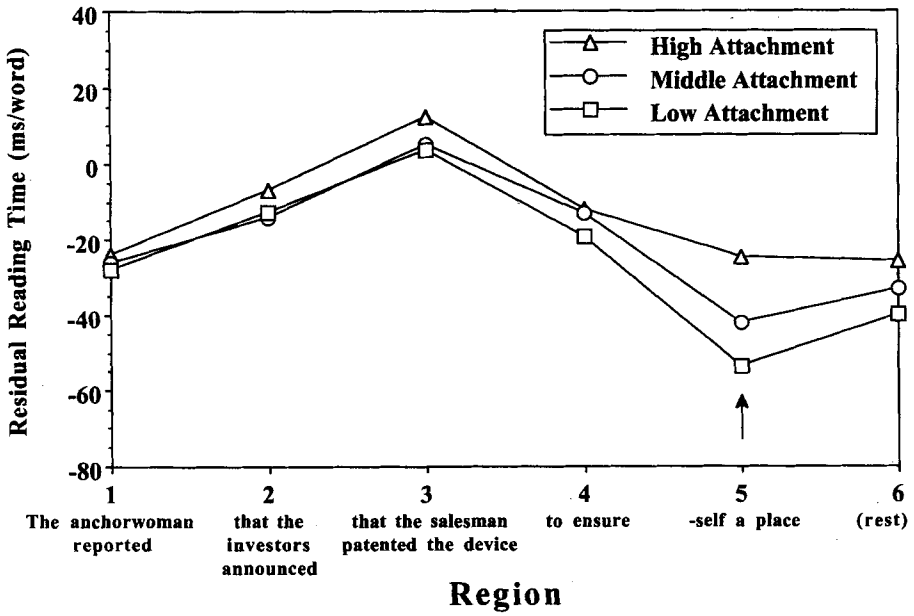

Figure 2. Experiment 3 trimmed residual reading time per word by region. Region 5 (marked with the arrow) is the three-word disambiguating region.

and suggest that secondary factors such as relativized relevance and primacy in short-term memory cannot adequately account for the pattern of preferences across both attachments to NP sites and attachments to VP sites.

Either lexical preferences or contextual biases might still be involved in determining biases for attachments into VP sites; the current data only indicate that some other factor like recency is involved as well. In addition, because Experiment 3 only controlled for lexical biases and plausibility on the high and middle attachment sites, the overall preference for low attachment still might be determined by such factors. This seems unlikely given the wide range of intuitive and experimental support for a preference to attach to the most recently available site (e.g., Altmann et al., 1998; Frazier & Fodor, 1978; Gibson, 1998; Igoa et al., 1998), but these experiments do not rule out the possibility.

Although Experiment 3 provided clear evidence about the preference for middle over high attachment, the low versus middle attachment comparison was somewhat weaker than in Experiment 2, and the size of the difference between the low and middle attachment conditions was surprisingly smaller (at least numerically) than that between the middle and high conditions. This was the reverse of the pattern in Experiment 2. This is important from the perspective of recency as described by Gibson, Pearlmutter, et al. (1996), because on that approach, difficulty should increase with increasing attachment distance, but at a negatively accelerated rate. Thus, the increment in difficulty between low and middle attachment is predicted to be at least as large or larger than that between middle and high attachment, and if recency is the prime determinant of preferences in our stimuli, this pattern should appear in the reading times.

We cannot determine definitely whether either the Experiment 2 pattern or the Experiment 3 pattern more accurately reflects underlying preferences. One possibility is that the decreased low versus middle difference in Experiment 3 arose from the small proportion of low attachment resolutions relative to high and middle resolutions. Each participant read 10 high and 10 middle attachment cases but only 5 low attachment resolutions in Experiment 3, whereas each resolution appeared 6 times in Experiment 2. This possibility receives some support from the fact that the low attachment condition mean was slower in Experiment 3 than in Experiment 2, whereas the high and middle attachment condition means were faster in Experiment 3 than in Experiment 2. To examine which of the Experiment 2 or Experiment 3 patterns would replicate, and to be sure that the low and middle attachment conditions do in fact differ, we conducted an additional experiment following the design of Experiment 2, with an equal number of high, middle, and low attachment resolutions.

Experiment 4 also examines four other potential concerns about Experiments 2 and 3.[6] First, it provides an additional check for the existence of differences prior to disambiguation. Although Experiment 3 revealed no such effects, it would be useful to be sure that they do not reappear with the same design as Experiment 2. A second concern is that in the first two on-line experiments the stimuli were always presented with a line break just preceding the beginning of the purpose clause, and the line break might have created an artificial boundary between the purpose clause and the potential attachment sites. Although the effect of such a boundary is unknown (but see Gibson & Schütze, 1999), it creates a potential confound. In Experiment 4, the line break was therefore shifted earlier, so that it always immediately followed the third verb. A third concern about Experiments 2 and 3 is that in some of the stimuli readers might not always have immediately recognized *to*

[6] Thanks to Marc Brysbaert, Don Mitchell, and an anonymous reviewer for pointing out these issues.

as an infinitival marker, because it could also be a preposition in those items, prior to the presentation of the next word (e.g., *faked the injury to get attention* vs. *faked the injury to his ankle*). At a minimum, this would likely add noise to the results, and from the perspective of a theory like Hemforth et al.'s (1998, 1999), it could result in a delay in the operation of anaphoric binding, supposing that infinitivals do contain an anaphor to be bound, whereas prepositional phrases do not. Thus, the stimuli in Experiment 3 were altered for Experiment 4 to remove these ambiguities (e.g., *faked the accident to get attention*). The stimuli were also adjusted for one other potential ambiguity. In four of the Experiment 2 and 3 items, the end of the attaching purpose clause contained a phrase that could potentially attach to an earlier site rather than to one within the purpose clause (e.g., *last week* in *set the fire to get herself into the news last week*). This would not necessarily have affected the attachment decisions for these specific items, but it might potentially have had an effect on the overall preference for recent versus nonrecent attachment across an experiment. For each of the four items from Experiment 3 containing this ambiguity, we changed the end of the purpose clause in Experiment 4 to remove the ambiguity (e.g., *set the fire to get herself into the news reports*).

Experiment 4

Method

Participants. Sixty-four Northeastern University undergraduates participated for class credit. All were native English speakers, and none had participated in the earlier experiments.

Materials and design. Twenty-one sentence triples like those used for Experiment 2 (see the top panel of Table 1) were constructed. Eleven of these were identical to items from Experiment 3 (without swapping high and middle sites). One new triple was constructed, and the remaining nine were very similar to items from Experiment 3 but with changes to the third VP or the attaching purpose clause to eliminate potential concerns about, first, the infinitival head (*to*) being treated as a preposition and, second, additional ambiguously attaching phrases near the end of the purpose clause. Appendix B shows the complete set of stimuli.

As in Experiment 2, the stimuli were placed into three lists to balance the appearance of the different anaphors across conditions. The same set of 50 fillers was used as in Experiment 3, except that the line break in many was shifted so that line lengths would remain similar to those in the experimental items. As in Experiments 2 and 3, yes/no comprehension questions were constructed for each item.

Apparatus and procedure. Three Apple Macintosh Quadra personal computers were used to present the stimuli and collect data. Procedures were the same as in Experiments 2 and 3, except that the line break for each experimental item was always placed immediately after the third verb rather than just before the infinitival purpose clause. Participants saw 10 practice items followed by one of the 71-item lists in a random order, and they completed the experiment in approximately 25 min.

Results

Comprehension question performance. Comprehension performance was 55% correct for the high attachment condition, 60% correct for the middle attachment condition, and 76% correct for the low attachment condition. These values were reliably different from chance, with the exception of the high attachment condition in the analysis by items, $t_1(63) = 2.01$, $SD = 20$, $p < .05$; $t_2(20) = 1.61$, $SD = 15$, $p < .12$; all other $t_1(63) > 3.7$, $SD < 22$, $ps < .001$; $t_2(20) > 2.09$, $SD < 26$, $ps < .05$.

The attachment conditions differed reliably from each other, $F_1(2, 126) = 20.79$, $MSE = 354$, $p < .001$; $F_2(2, 40) = 6.35$, $MSE = 383$, $p < .01$, and individual mean comparisons revealed a pattern identical to that in Experiment 3. The high and middle attachment conditions did not differ, $F_1(1, 63) = 2.40$, $MSE = 321$, $p > .10$; $F_2 < 1$, but the low attachment condition was better than the other two; low versus middle: $F_1(1, 63) = 18.45$, $MSE = 423$, $p < .001$; $F_2(1, 20) = 5.58$, $MSE = 469$, $p < .05$; low versus high: $F_1(1, 63) = 42.56$, $MSE = 317$, $p < .001$; $F_2(1, 20) = 10.20$, $MSE = 435$, $p < .01$.

Reading times. Reading times were trimmed at 4 *SD* (affecting less than 1.2% of the data) and were analyzed as in Experiment 2. Figure 3 shows trimmed residual reading times by condition for each region, and Appendix C shows the correspondingly trimmed raw reading times.

As in Experiments 2 and 3, there were no reliable effects of attachment in any of the regions preceding the disambiguation, all $F_1(2, 126) < 1.1$, $MSE > 1{,}101$, $ps > .35$; $F_2(2, 40) < 1.2$, $MSE > 417$, $ps > .30$. At the disambiguation (Region 5), reading times differed reliably, $F_1(2, 126) = 8.25$, $MSE = 1{,}273$, $p < .001$; $F_2(2, 40) = 10.31$, $MSE = 332$, $p < .001$. All three conditions differed, although for the high versus middle attachment comparison, this was reliable only by items: high versus middle, $F_1(1, 63) = 2.41$, $MSE = 1{,}614$, $p < .13$; $F_2(1, 20) = 4.88$, $MSE = 237$, $p < .05$; middle versus low, $F_1(1, 63) = 10.61$, $MSE = 635$, $p < .01$; $F_2(1, 20) = 8.07$, $MSE = 289$, $p < .05$; high versus low, $F_1(1, 63) = 13.30$, $MSE = 1{,}570$, $p < .001$; $F_2(1, 20) = 14.44$, $MSE = 470$, $p < .01$.

Discussion

Like the preceding experiments, Experiment 4 indicated that readers preferred middle attachment to high attachment, and, like Experiments 2 and 3, it indicated that readers preferred low attachment to middle attachment. This was the case despite changes to several stimuli to correct two incidental ambiguities and despite the different location of the line break. Furthermore, as in Experiment 3, there was no hint of any difference among the conditions prior to disambiguation.

Unlike Experiment 3, but matching Experiment 2, the size of the difference between the low and middle attachment conditions in Experiment 4 was numerically larger than that between the middle and high conditions. This fits with the pattern predicted by recency as described by Gibson, Pearlmutter, et al. (1996). It suggests that the reason the low versus middle difference was relatively small in Experiment 3 was that middle and high attachment resolutions appeared twice as often as low attachment resolutions in that experiment, thus artificially increasing the difficulty of the low attachment cases.

General Discussion

Our results provide clear evidence supporting a monotonic preference ordering for attachment of modifiers into three potential VP sites in English. Comprehenders judged middle attachment to be easier than high attachment off-line in Experiment 1, and middle attachment conditions were read more quickly than high attachment conditions in Experiments 2, 3, and 4. Low attachment was preferred to both middle and high attachment as well. Thus, in

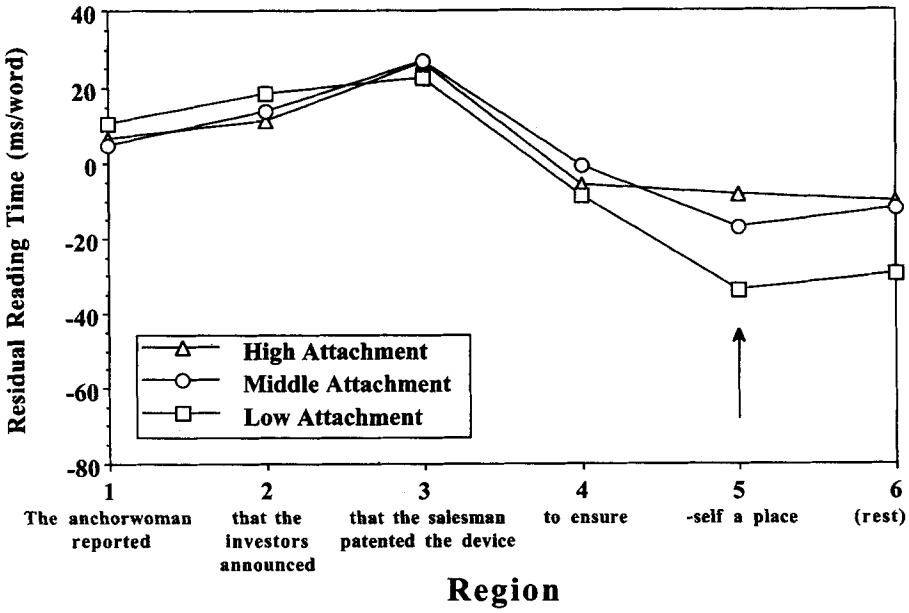

Figure 3. Experiment 4 trimmed residual reading time per word by region. Region 5 (marked with the arrow) is the three-word disambiguating region.

both off-line and on-line measures, using different attaching phrases (temporal adverbials vs. purpose clauses), different kinds of disambiguation (tense marking vs. anaphoric number agreement), different line breaks, and different fillers, readers displayed a clear recency preference for attachments into VPs. The preference ordering was consistently low/middle/high.

This pattern matches the pattern for attachment of modifiers into two potential VP sites (Altmann et al., 1998; Igoa et al., 1998), but it contrasts sharply with the pattern of attachment of RCs into three potential NP sites, which is nonmonotonic in English and in other languages (Spanish, German, and Japanese). Readers prefer low attachment followed by high attachment; middle attachment is reliably most difficult (Gibson, Pearlmutter, et al., 1996; Gibson et al., 1999; Miyamoto et al., 1999; Walter & Hemforth, 1998; cf. Wijnen, 1998, 1999, for Dutch). This contrast between modifier attachments into VP and NP sites is predicted by a combination of recency and predicate proximity, because the latter will create a preference to attach to the highest NP in a series of preposition-linked NP sites (opposing recency), but it will not differentiate among multiple VP sites, yielding strong recency effects. On the other hand, the contrast between attachments into VP and NP sites is not compatible with several other potential secondary factors proposed to combine with recency and explain preferences. Both Frazier's (1990) relativized relevance factor and a primacy factor based on short-term memory (e.g., as considered by Gibson, Pearlmutter, et al., 1996) take no account of the category of site into which the modifier is attaching, and thus they cannot distinguish between attachments into multiple NP and multiple VP sites. In the right-branching constructions considered here and in the literature, both of these factors support attachment to the first site (structurally highest or earliest) in opposition to recency, and they therefore predict a nonmonotonic preference ordering. This is compatible with the results for NP attachment but not with the current results.

Approaches relying on discourse-prominence–based secondary factors also have difficulty with the monotonic preference ordering in the current results, assuming that the matrix clause is the most prominent one in the Experiment 1–4 stimuli. Alternatively, these results can be viewed as further specifying the nature of discourse prominence: They require that prominence be accorded to the most recent clause in our stimuli, but that it be accorded to NPs as close as possible to the matrix predicate for the purposes of RC attachment.

The modifier attachment theory of Hemforth et al. (1998, 1999) may be able to account for the current results, depending on what assumptions are made about the nature of anaphoric binding. The theory is compatible with Experiment 1 regardless of the nature of anaphoric binding, because the adverbial phrases did not include an anaphor. The theory is also compatible with Experiments 2–4 under either of two assumptions, as discussed earlier: Either anaphoric binding applies to attaching purpose clauses and discourse prominence favors the most recent clause; or some factor other than anaphoric binding handles the null pronominal of purpose clauses, and this other factor prefers to coindex the pronominal with the most recent available subject NP.

Experiment 3 allowed elimination of two other potential explanations for the attachment preference pattern, lexical preferences, and plausibility biases, both closely associated with constraint-based lexicalist approaches (e.g., MacDonald et al., 1994; Trueswell & Tanenhaus, 1994). Experiment 3 controlled for these potential biases by exchanging the lexical content of the high and middle attachment sites while holding constant the content of the attaching purpose clauses. If either of lexical biases or plausibility were responsible for middle attachment being preferred to high attachment in the first two experiments, these effects should have disappeared in Experiment 3, but Experiment 3 showed the same pattern as Experiments 1, 2, and 4. Nevertheless, these results

certainly do not rule out the influence of lexical and other biases, either for other constructions or for the current one. In fact, the view described in Gibson and Pearlmutter (1998) predicts that both lexical and contextual constraints have a broad influence in processing. The point here, though, is that such effects are isolable from those of recency.

Another possible factor considered in recent constraint-based work on attachment into multiple possible NP sites is modifiability, which can be seen as a more general version of Crain and Steedman's (1985) principle of parsimony. Thornton, Gil, and MacDonald (1998; Thornton, MacDonald, & Gil, 1999) argue that NPs will be more likely targets for modifier attachment to the extent that they are not already otherwise modified, where modification refers specifically to nonargument attachment (see, e.g., Schütze & Gibson, 1999, for discussion of the argument/modifier distinction). Miyamoto et al. (1999) propose that a similar factor must interact with recency and predicate proximity to explain RC attachment preferences in Japanese. Although Thornton et al. (1999) suggest that attachment distance (recency) is a relevant factor in determining preferences, they note that recency has been confounded with modifiability in prior work, such that, at least for cases involving attachment to two NP sites, the less recent site is also less modifiable. This is because the less recent site will generally already have been modified by the prepositional phrase containing the more recent site when the ambiguously attached RC is encountered. The current studies, however, control modifiability in that none of the VP sites in the Experiment 1–4 constructions have been modified prior to encountering the temporal adverbial or purpose clause. These results therefore support the idea that recency is an important general factor in attachment decisions across a wide variety of constructions (ambiguous and unambiguous; Gibson, 1998), irrespective of other interacting biases.

Supposing that recency is a factor in attachment decisions, we can also consider some possible explanations for its origins. One class of explanations directly relates preferences in comprehension to the statistical biases present in the language to which a comprehender has been exposed (e.g., Cuetos & Mitchell, 1988; Juliano & Tanenhaus, 1994; MacDonald, 1999; Mitchell, Cuetos, Corley, & Brysbaert, 1995; Tabor, Juliano, & Tanenhaus, 1997), such that comprehenders' preferences essentially mirror the biases present in their input. Varying the grain size over which such models collapse input patterns (e.g., Mitchell et al., 1995) or using an algorithm that allows the model to determine its own grain size (e.g., Tabor et al., 1997), allow these models some leeway in the tightness of the connection between biases in the input and predicted preferences. As noted earlier, however, comprehension preferences and biases in text (as measured by corpus frequency counts) do not always agree (e.g., Gibson & Schütze, 1999; Gibson, Schütze, & Salomon, 1996; Mitchell & Brysbaert, 1998), and it therefore remains unclear whether directly mapping input biases to parsing preferences can completely account for the range of attachment preference effects.

Even if some kind of direct mapping model can account for the range of attachment preferences, the question about the source of recency effects still remains open. This is because the direct mapping on such models cannot itself explain the original source of preferences; some additional reason for the biases in the input must be specified. In the case of recency, one suggestion comes from MacDonald (1999; Thornton et al., 1999), who argues that the source of recency effects in modifier attachment to VPs during comprehension is a constraint on language production. In particular, she proposes that shorter phrases, because they can be prepared more easily, will tend to be produced earlier than later ones. Because of this heaviness constraint on production (e.g., Hawkins, 1994), the distribution of adverbial phrase attachments into two possible VP sites will tend to match the predictions of recency. This can be seen by consideration of Example 9, in which *yesterday* can modify either *realized* or *was happy.*

9. Willy realized that Biff was happy yesterday.

If the speaker intended the adverbial *yesterday* to modify *realized,* then the adverbial would most likely have been produced earlier (e.g., *Willy realized yesterday that . . .*), because it is shorter than the other constituent associated with *realized,* the clause *that Biff was happy.* As a result, MacDonald (1999) argues, the distribution of attachments by adverbials to the more recent VP site in such constructions is likely to favor recent attachment. If comprehenders are sensitive to such distributional properties (as in a direct mapping framework), they should prefer to attach *yesterday* to the more recent VP, but they will do so without relying on any independent recency factor. Thus, one source of recency effects might be the distributional properties of a language created by constraints applied during production.

As noted by MacDonald (1999), this account makes an additional important prediction, which is that recency effects should not appear when the heaviness of the attaching element is the same or larger than the heaviness of the material across which the nonrecent attachment occurs. This can be seen in Sentence 10, in which *very very much* modifies *likes.*

10. Art likes milling around very very much.

Very very much is heavier than *milling around* in Example 10, and thus there is no production pressure to generate it early, even though it is intended to modify the earlier site *likes.* As a result, there will be no distributional bias favoring either recent or nonrecent attachment in such cases, and the comprehension system should not display any relative difficulty associated with nonrecent attachment.

The Experiment 2–4 comparison between middle and low attachment provides a test of this prediction, because the attaching purpose clause modifiers were relatively long, as shown in Example 11 (repeated from Table 1). The relevant comparison is between the purpose clause (*to ensure themselves a place on the advisory board*) and the clause preceding it, which contained the low attachment site (*that the salesman patented the device*). The mean purpose clause length in Experiment 4 (Experiment 2 and 3 values were nearly identical) was 8.2 words (13.5 syllables), whereas the mean low site clause length was 6.3 words (9.4 syllables; see Appendix B).

11. The anchorwoman reported that the investors announced that the salesman patented the device to ensure themselves a place on the advisory board.

Because the purpose clauses were if anything longer than the lowest site material, a distributional account based on relative heaviness predicts that middle attachments should have been no more difficult than low attachments. Thus, this account cannot

explain the reliable recency preference across the middle and low VP sites in Experiments 2–4, and a relative heaviness account cannot provide a full account of recency effects, even considering just those cases involving attachments of modifiers into multiple potential VP sites.

Two further points about the operation of a relative heaviness factor also require some attention. First, because the computation of relative heaviness requires knowledge of the length of the attaching phrase, its application to initial attachment preferences is not clear. This might be resolved in one of three ways: (a) Relative heaviness might not apply in determining initial attachment preferences; in this case, though, it obviously cannot be the source of recency effects. (b) Relative heaviness might be computed at the beginning of the purpose clause on the basis of an estimate of the clause's likely length. But regardless of whether purpose clauses are treated as a separate category for estimation or pooled with similar attaching modifiers (see later discussion), the best guess is likely to be that the purpose clause will be about as long as the preceding clause. This guess would in fact be correct for the Experiment 2–4 stimuli (the purpose clauses were actually longer than the preceding clauses), and as noted earlier, this yields the wrong prediction. (c) Initial preferences might be based on initially available relative heaviness, so that, essentially, all modifiers initially attach as if they are short, but preferences might change toward less recent attachment as more and more of the attaching modifier is processed. This would account for the effects at disambiguation in Experiments 2–4, but it does not predict the lack of difficulty in Example 10. *Very very much* should initially attach recently, but this would have to be revised when *much* is processed. Furthermore, the pattern of comprehension performance in Experiments 2–4, measured after the end of each sentence, seemed to match the initial preference pattern measured at the disambiguation, particularly for the low versus middle attachment comparison, which suggests that the initial recency preference was not later altered. Thus in each of these cases, relative heaviness is unable to account for the Experiment 2–4 results.

The second concern about the operation of relative heaviness and, in fact, for direct mapping accounts in general is that such accounts require a grain-size parameter that determines the nature of the categories for which the direct mapping is defined (e.g., individual lexical items, VP-VP-adverbial sequences, VP-VP-modifier sequences, [multiple-VP]-modifier sequences, [multiple-XP]-modifier sequences). Without specifying this parameter, it is often not possible to determine precisely what such models predict. For purpose clause attachments, the relevant issue is likely to be whether the direct mapping category determining the purpose clause preference also includes other modifiers like verb particles and temporal adverbials, which tend to be short and which will therefore create a bias for recent attachment. We cannot rule out the possibility that some grain size can be specified that will allow correct predictions, but this seems unlikely for several reasons. First, verb particles are relatively uncommon in English, appearing only with specific verbs, and they serve very different semantic and discourse functions from clausal modifiers. Second, even though temporal adverbials are quite common, they are not necessarily particularly short (e.g., most of the Experiment 1 stimuli contained an additional phrase or clause after the two-word adverbial phrase, which attached with the adverbial), and they too serve a somewhat different function from purpose clauses. Third, an additional category of attaching modifiers, tensed subordinate clauses, is probably relevant (e.g., *John confessed that he had proctored the test because Abigail promised him lots of goodies,* where *because Abigail promised him lots of goodies* can modify either the higher or lower clause). Such clauses are similar in semantic and discourse function to purpose clauses, and they are generally long. Thus, like other ambiguously attached clausal modifiers in multiple VP constructions, subordinate clauses will tend to be similar (or greater) in heaviness to the material over which they attach, and, consequently, they will tend to eliminate heaviness-induced recency effects. Despite these potential problems, a relative heaviness account might still be feasible, but further specification of the relevant grain size for direct mapping will ultimately be needed to evaluate this approach thoroughly.

An alternative source for recency arises from the nature of language as a sequential system and the manner in which the comprehension system must deal with this property (see also Lewis, 1996; Stevenson, 1994; for related approaches). This view relies on three very general assumptions: (a) Parts of an utterance are encountered in series rather than simultaneously; (b) memory is imperfect, or, more specifically, elements in working memory tend to become less available as additional elements are processed; and (c) more-available components of a memory representation are easier for the comprehension system to manipulate. Regardless of the nature of the representation involved (e.g., a syntactic tree, a mental model, a set of propositions, a series of words), the combination of these three assumptions will result in a comprehension system showing recency effects. This is because more recently encountered parts of an utterance will be associated with stronger representations in memory, and the system will therefore have less difficulty in connecting new input to such representations.

Although the recency metric we have adopted is essentially a structural one (locality; Gibson, 1998), it should be clear from the aforementioned assumptions that this is not critical; the choice of representation over which recency is defined will depend upon the specific processing theory in which recency is embedded. All the data of which we are aware are compatible with either a syntax- or a discourse-based recency factor and possibly with others as well. When considered in this manner, recency is notably different from metrics of complexity such as minimal attachment (Frazier, 1979, 1990). The latter relies critically on the content of syntactic representations, and thus its predictions will vary depending on a variety of theory-specific details of the internal representations constructed by the comprehension system. This is particularly relevant from the perspective of lexicalist constraint-based approaches, which have typically attempted to replace structure-based biases like minimal attachment with combinations of lexical and contextual biases. In the case of recency, because it can be justified on a much broader basis than can purely structural biases and because it appears to be needed to account for preferences even when lexical biases, plausibility, modifiability, and relative heaviness have all been controlled (Experiment 3; see also Gibson et al., 1999), it will likely be a necessary component of a constraint-based approach to sentence processing.

Gibson and Pearlmutter (1998) described a constraint-based approach of this sort, which included grammatical, lexical, prosodic, and contextual constraints, along with recency (locality) and predicate proximity. The interaction of these factors was claimed to be sufficient to account for sentence comprehension behavior.

Experiments 1–4 provide additional support for this view by isolating and further examining recency in particular, controlling for possible interactions with predicate proximity, grammatical constraints, lexical biases, and plausibility. These results also allowed us to rule out a variety of alternatives to predicate proximity and partial alternatives to recency as accounts of VP attachment preferences. In combination with evidence from Gibson (1998), Gibson, Pearlmutter, et al. (1996), and Gibson et al. (1999), these data argue for an independent recency factor that creates a preference to link an attaching element to a site over as short a distance as possible, with difficulty increasing (roughly) continuously with distance between the attaching element and the target site.

References

Altmann, G. T. M., van Nice, K. Y., Garnham, A., & Henstra, J. (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.

Carreiras, M., & Clifton, C., Jr. (1993). Relative clause interpretation preferences in Spanish and English. *Language and Speech, 36,* 353–372.

Chomsky, N. (1981). *Lectures on government and binding.* Dordrecht, The Netherlands: Foris.

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior, 12,* 335–359.

Clifton, C., Jr. (1993). Thematic roles in sentence parsing. *Canadian Journal of Psychology, 47,* 224–246.

Crain, S., & Steedman, M. (1985). On not being led up the garden path: The use of context by the psychological syntax processor. In D. R. Dowty, L. Karttunen, & A. M. Zwicky (Eds.), *Natural language parsing: Psychological, computational, and theoretical perspectives* (pp. 320–358). Cambridge, UK: Cambridge University Press.

Cuetos, F., & Mitchell, D. C. (1988). Cross-linguistic differences in parsing: Restrictions on the use of the Late Closure strategy in Spanish. *Cognition, 30,* 73–105.

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

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

Fodor, J. D. (1998). Learning to parse? *Journal of Psycholinguistic Research, 27,* 285–319.

Frazier, L. (1979). *On comprehending sentences: Syntactic parsing strategies.* Bloomington, IN: Indiana University Linguistics Club.

Frazier, L. (1990). Parsing modifiers: Special purpose routines in the human sentence processing mechanism? In D. A. Balota, G. B. Flores d'Arcais, & K. Rayner (Eds.), *Comprehension processes in reading* (pp. 303–330). Hillsdale, NJ: Lawrence Erlbaum.

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

Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition, 6,* 291–325.

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

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

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

Gibson, E., & Loomis, J. (1994). A corpus analysis of recency preference and predicate proximity. In R. A. Ram & K. Eiselt (Eds.), *Proceedings of the 16th Annual Conference of the Cognitive Science Society* (pp. 357–362). Hillsdale, NJ: Erlbaum.

Gibson, E., & Pearlmutter, N. J. (1998). Constraints on sentence comprehension. *Trends in Cognitive Sciences, 2,* 262–268.

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

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

Gibson, E., & Schütze, C. T. (1999). Disambiguation preferences in noun phrase conjunction do not mirror corpus frequency. *Journal of Memory and Language, 40,* 263–279.

Gibson, E., Schütze, C. T., & Salomon, A. (1996). The relationship between the frequency and the processing complexity of linguistic structure. *Journal of Psycholinguistic Research, 25,* 59–92.

Gilboy, E., Sopena, J., Clifton, C., & Frazier, L. (1995). Argument structure and association preferences in Spanish and English complex NPs. *Cognition, 54,* 131–167.

Gordon, R. B. (1982). *The listener resolves an ambiguity: Some principles governing the association of adverbials with clauses in English sentences.* Unpublished manuscript, Midwood High School, Brooklyn, NY.

Gorrell, P. (1995). *Syntax and perception.* Cambridge, UK: Cambridge University Press.

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

Hemforth, B., Konieczny, L., & Scheepers, C. (1999). Syntactic attachment and anaphor resolution: The two sides of relative clause attachment. In M. Crocker, M. Pickering, & C. Clifton, Jr. (Eds.), *Architectures and mechanisms for language processing* (pp. 259–281). Cambridge, UK: Cambridge University Press.

Hemforth, B., Konieczny, L., Scheepers, C., & Strube, G. (1998). Syntactic ambiguity resolution in German. In D. Hillert (Ed.), *Sentence processing: A crosslinguistic perspective* (pp. 293–312). San Diego, CA: Academic Press.

Igoa, J. M., Carreiras, M., & Meseguer, E. (1998). A study on late closure in Spanish: Principle-grounded vs. frequency-based accounts of attachment preferences. *Quarterly Journal of Experimental Psychology (A), 51,* 561–592.

Jones, C. (1991). *Purpose clauses: Syntax, thematics, and semantics of English purpose constructions.* Dordrecht, The Netherlands: Kluwer.

Juliano, C., & Tanenhaus, M. K. (1994). A constraint-based lexicalist account of the subject/object attachment preference. *Journal of Psycholinguistic Research, 23,* 459–471.

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

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

Kennedy, A., & Murray, W. S. (1984). Inspection times for words in syntactically ambiguous sentences under three presentation conditions. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 833–849.

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

Kjelgaard, M. M., & Speer, S. R. (1999). Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language, 40,* 153–194.

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

MacDonald, M. C. (1999). Distributional information in language compre-

hension, production, and acquisition: Three puzzles and a moral. In B. MacWhinney (Ed.), *The emergence of language* (pp. 177–196). Mahwah, NJ: Erlbaum.

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

Mauner, G., Tanenhaus, M. K., & Carlson, G. N. (1995). Implicit arguments in sentence processing. *Journal of Memory and Language, 34,* 357–382.

Mitchell, D. C. (1987). Lexical guidance in human parsing: Locus and processing characteristics. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 601–618). Hillsdale, NJ: Erlbaum.

Mitchell, D. C., & Brysbaert, M. (1998). Challenges to recent theories of cross-linguistic variation in parsing: Evidence from Dutch. In D. Hillert (Ed.), *Sentence processing: A cross-linguistic perspective* (pp. 313–335). San Diego, CA: Academic Press.

Mitchell, D. C., & Cuetos, F. (1991). The origins of parsing strategies. In C. Smith (Ed.), *Proceedings of the Current Issues in Natural Language Processing Conference* (pp. 1–12). Austin, TX: University of Texas.

Mitchell, D. C., Cuetos, F., Corley, M. M. B., & Brysbaert, M. (1995). Exposure-based models of human parsing: Evidence for the use of coarse-grained (nonlexical) statistical records. *Journal of Psycholinguistic Research, 24,* 469–488.

Mitchell, D. C., & Holmes, V. M. (1985). The role of specific information about the verb in parsing sentences with local structural ambiguity. *Journal of Memory and Language, 24,* 542–559.

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.

Murdock, B., Jr. (1962). The serial position effect of free recall. *Journal of Experimental Psychology, 64,* 482–488.

Pearlmutter, N. J., Garnsey, S. M., & Bock, K. (1999). Agreement processes in sentence comprehension. *Journal of Memory and Language, 41,* 427–456.

Pollard, C., & Sag, I. A. (1994). *Head-driven phrase structure grammar.* Chicago, IL: University of Chicago Press.

Pritchett, B. L. (1988). Garden path phenomena and the grammatical basis of language processing. *Language, 64,* 539–576.

Schafer, A., Carter, J., Clifton, C., Jr., & Frazier, L. (1996). Focus in relative clause construal. *Language and Cognitive Processes, 11,* 135–163.

Schütze, C. T., & Gibson, E. (1999). Argumenthood and English prepositional phrase attachment. *Journal of Memory and Language, 40,* 409–431.

Speer, S. R., Kjelgaard, M. M., & Dobroth, K. M. (1996). The influence of prosodic structure on the resolution of temporary syntactic closure ambiguities. *Journal of Psycholinguistic Research, 25,* 247–268.

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

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

Stowe, L. A. (1989). Thematic structures and sentence comprehension. In G. N. Carlson & M. K. Tanenhaus (Eds.), *Linguistic structure in language processing* (pp. 319–357). Dordrecht, The Netherlands: Kluwer.

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

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

Thornton, R., Gil, M., & MacDonald, M. C. (1998). Accounting for crosslinguistic variation: A constraint-based perspective. In D. Hillert (Ed.), *Sentence processing: A crosslinguistic perspective* (pp. 211–225). San Diego, CA: Academic Press.

Thornton, R., MacDonald, M. C., & Gil, M. (1999). Pragmatic constraint on the interpretation of complex noun phrases in Spanish and English. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 1347–1365.

Trueswell, J. C. (1996). The role of lexical frequency in syntactic ambiguity resolution. *Journal of Memory and Language, 35,* 566–585.

Trueswell, J. C., & Tanenhaus, M. K. (1994). Towards a lexicalist framework of constraint-based syntactic ambiguity resolution. In C. Clifton, L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 155–179). Hillsdale, NJ: Erlbaum.

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.

Tulving, E. (1968). Theoretical issues in free recall. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory* (pp. 2–36). Englewood Cliffs, NJ: Prentice-Hall.

Walter, M., & Hemforth, B. (1998, March). *Attachment preferences of extraposed and adjacent relative clauses following three-site NPs in German.* Poster presented at the 11th Annual CUNY Human Sentence Processing Conference, New Brunswick, NJ.

Warner, J., & Glass, A. L. (1987). Context and distance-to-disambiguation effects in ambiguity resolution: Evidence from grammaticality judgments of garden path sentences. *Journal of Memory and Language, 26,* 714–738.

Wijnen, F. (1998). *Dutch relative clause attachment in two- and three-site contexts.* Poster presented at the 11th Annual CUNY Human Sentence Processing Conference, New Brunswick, NJ.

Wijnen, F. (1999). *Prosodic phrasing and relative clause attachment in a three-site context.* Poster presented at the 12th Annual CUNY Human Sentence Processing Conference, New York, NY.

Appendix A

Experiment 1 Stimuli

The two conditions (high and middle attachment) differed in the auxiliaries in the first two clauses. The auxiliaries preceding the slashes appeared in the high attachment version, whereas the auxiliaries following the slashes appeared in the middle attachment version. Tense marking on the verb after each auxiliary was also changed as necessary to match the auxiliary. The disambiguating adverbial phrase is in boldface.

1. The president will/has reply/ied that the manager has/will object(ed) that the employees have hated the schedule **next week** when the layoffs are announced.
2. The principal will/has report(ed) that the teacher has/will argue(d) that the children have learned the story **next month** when the school reopens.
3. The priest will/has recognize(d) that the children have/will understand/stood that the parents have worried a lot **next year** when they leave home.
4. The judge will/has note(d) that the guard has/will confirm(ed) that the prisoner has confessed the crime **next week** when testimony continues in the trial.
5. The ambassador will/has complain(ed) that the tryant has/will assume(d) that the US has dictated the policies **next Tuesday** at the United Nations.
6. The network will/has announce(d) that the team has/will reveal(ed) that the athlete has disputed the contract **tomorrow night** at 8PM on national TV.
7. The students would/had regret(ted) that the faculty had/would concede(d) that the committee had recommended no one **next week** after the job interviews.
8. The executive will/has boast(ed) that the analyst has/will indicate(d) that the business has shown high profits **next month** at the shareholders' meeting.
9. The mother would/has suspect(ed) that the father had/would forget/gotten that the boy had repeated a grade **next year** when enrollment begins.
10. The editor would/had hint(ed) that the reviewer had/would confide(d) that the scientist had criticized the article **tomorrow night** when the panel meets privately.
11. The secretary would/had answer(ed) that the mayor had/would imply/ied that the lawyer had seen the documents **next Friday** when the city council assembles.
12. The attorney will/has affirm(ed) that the chairman has/will acknowledge(d) that the bank has guaranteed the loans **next month** when the regulatory board meets.
13. The detective would/had suggest(ed) that the gangster had/would admit(ted) that the police had found the hideout **next month** after everything has been moved.
14. The spokesman would/had mention(ed) that the CEO had/would deny/ied that the buyers had noticed the change **tomorrow morning** when the convention begins.
15. The expert would/had claim(ed) that the jury had/would consider(ed) that the convict had explained the motive **next June** when the hearings take place.
16. The newspaper would/had print(ed) that the director had/would appreciate(d) that the author had written the screenplay **next May** before any casting begins.

Appendix B

Experiment 2–4 Stimuli

The conditions differed in which of the three anaphors separated by slashes appeared in the purpose clause: They are shown in the order high/middle/low attachment. All stimuli were used in all three experiments, except for Items 19 and 20, which were used only in Experiments 3 and 4, and Item 21, which was used only in Experiment 4. Material in parentheses marks changes for Experiments 3 and 4: One complete item was replaced (Item 8), and a verb in Item 2 was changed. Material in square brackets separated by a slash differed between Experiments 2 and 3 and Experiment 4: Text preceding the slash was used only in Experiments 2 and 3, and text following the slash was used only in Experiment 4. Lexical preferences for high versus middle attachments in Experiment 3 were controlled by swapping the subject NP, verb, and complementizer in the first two clauses. See the *Materials and design* subsections of each experiment for details.

1. The grandmother claimed that the fireman said that the arsonists set the fire to get herself/himself/themselves into the news [last week/reports].
2. The chairwoman admitted that the professors decided (argued) that the hockey player should take a second exam to clear herself/themselves/himself of any appearance of impropriety.
3. The hostess argued that Rush Limbaugh proclaimed that the fans listened to the show to make herself/himself/themselves feel important [during the crisis/and sophisticated].
4. The witnesses testified that the princess disclosed that the king remained silent to make themselves/herself/himself seem impartial [during the investigation/or at least not too prejudiced].
5. The president hinted that the senators stated that the woman arranged the [payoff/meeting] to get himself/themselves/herself more time on the air.

(Appendixes continue)

6. The prosecutors questioned whether the suspect suggested that the maid hid the evidence to allow themselves/himself/herself a little extra time.
7. The landlord asserted that the neighbors complained that the girl faked the [injury/accident] to guarantee himself/themselves/herself a reasonable settlement in the case.
8. The girl announced that the boy responded that the teachers asked hard questions to establish herself/himself/themselves as the smartest in the history class. (The baroness confirmed that the detective disclosed that the thieves hid the painting to assure herself/himself/themselves a generous reward from the government.)
9. The anchorwoman reported that the investors announced that the salesman patented the device to ensure herself/themselves/himself a place on the advisory board.
10. The company declared that the architects maintained that the secretary photocopied the blueprints to prevent itself/themselves/herself from being sued by the client.
11. The family members implied that the heiress insisted that the son fled the country to shield themselves/herself/himself from public scrutiny [last April/and embarrassment].
12. Father O'Reilly explained that the nun revealed that the church bought the orphanage to build himself/herself/itself up in the public eye.
13. The lawyers established that the judge ruled that the institute must publicize the findings to prevent themselves/himself/itself from appearing biased in any way.
14. The boy tattled that his sister contended that the other kids [took/kept] the vase to clear himself/herself/themselves of any hint of suspicion.
15. The editor printed that the producers confirmed that the actress wrote the screenplay to earn himself/themselves/herself some recognition in Hollywood.
16. The baseball player said that the network reported that the owners signed the contract to promote himself/itself/themselves in the new season.
17. The cooks gossiped that the trucker commented that the waitress flirted with everyone to amuse themselves/himself/herself [at the diner on Wednesday/and the customers around the counter].
18. The hospital acknowledged that the nurse indicated that the gentleman demanded a second opinion to protect itself/herself/himself against a possible lawsuit.
19. The bodyguards insisted that the mistress stated that the butler forged the [letter/will] to keep themselves/herself/himself from being suspected in the murder.
20. The actress claimed that the studio revealed that the producers cancelled the controversial show to save herself/itself/themselves from being publicly ridiculed.
21. The corporation proved that the union organizer bragged that the workers protested the policy to secure itself/himself/themselves some positive articles in the newspapers.

Appendix C

Trimmed Raw Reading Times

Table C1
Experiment 2: Trimmed Raw Reading Time by Condition (ms/word)

	Region					
Attachment	1	2	3	4	5	6
High	370	390	400	366	382	350
Middle	368	376	400	357	370	354
Low	368	377	393	361	335	329

Note. Reading times were trimmed at 2 *SD*. Region 5 is the disambiguating region.

Table C2
Experiment 3: Trimmed Raw Reading Time by Condition (ms/word)

	Region					
Attachment	1	2	3	4	5	6
High	384	396	411	371	375	361
Middle	381	388	401	367	357	358
Low	376	393	394	362	345	349

Note. Reading times were trimmed at 2 *SD*. Region 5 is the disambiguating region.

Table C3
Experiment 4: Trimmed Raw Reading Time by Condition (ms/word)

	Region					
Attachment	1	2	3	4	5	6
High	400	396	405	363	375	364
Middle	396	399	406	366	366	362
Low	404	406	402	359	351	340

Note. Reading times were trimmed at 4 *SD*. Region 5 is the disambiguating region.

Received June 10, 1999
Revision received June 14, 2000
Accepted July 3, 2000 ■