Syntax 7:1, April 2004, 55–78

READING-TIME EVIDENCE FOR INTERMEDIATE LINGUISTIC STRUCTURE IN LONG-DISTANCE DEPENDENCIES

Edward Gibson and Tessa Warren

Abstract. Most linguistic theories since Chomsky (1973) have hypothesized that long-distance dependencies crossing multiple clauses are mediated by intermediate structures. This paper provides a new source of evidence for the existence of such intermediate structures: reading times during online sentence comprehension. The experiment presented here compares reading times for two structures involving the long-distance extraction of a *wh*-filler: (a) a structure in which a clause intervenes between the endpoints of the extraction, and (b) a structure in which a nominalization of the clause intervenes. The logic of the experiment relies on two hypotheses: first, that intermediate structures mediate the relationship between a *wh*-filler and its θ-role-assigning verb when a clause intervenes between them but not when a nominalization intervenes; and second, that reading times for a word increase as linear distance increases between the word and the position on which it is dependent in the partial structure for the input (Gibson 1998, 2000; Grodner, Watson & Gibson 2000). In combination, these hypotheses predict that reading times at the region in which the verb assigns a θ-role to the *wh*-filler will be faster in the clausal conditions than in the nominalized conditions, because in the clausal conditions intermediate structure mediate the *wh*-filler verb dependency and cause it to be more local. This prediction was confirmed.

1. Introduction

To account for the grammaticality of certain long-distance dependencies and the ungrammaticality of others, linguists have proposed the existence of intermediate structures in long-distance dependencies out of right-branching clauses (Chomsky 1973, 1977, 1981, 1986; Gazdar, Klein, Pullum & Sag 1985). We will refer to the hypothesis that there is some linguistic structure mediating long-distance dependencies through clauses as the *intermediate structure hypothesis*. For example, in Chomsky's (1973) transformational framework, a phrase can cross at most one bounding node at each step in a derivation from D-structure to S-structure, where the categories NP and IP (S) are the bounding nodes in English. Derivations that cross two or more bounding nodes in a single step violate a grammatical principle called *Subjacency*. To allow unbounded dependencies through clauses, Chomsky proposed that movement through an intervening clause can take place in two

* We would like to thank Susanne Tunstall for her help in running the experiment reported here. We would like to thank the following people for useful discussions of earlier presentations of this work: Dan Grodner, Robert Kluender, Alec Marantz, Edson Miyamoto, Neal Pearlmutter, Colin Phillips, Ivan Sag, Whitney Tabor, Matthew Walenski, Tom Wasow, the MIT/Northeastern sentence processing lab group, and the audience at the 11th CUNY Sentence Processing Conference at New Brunswick, N.J.

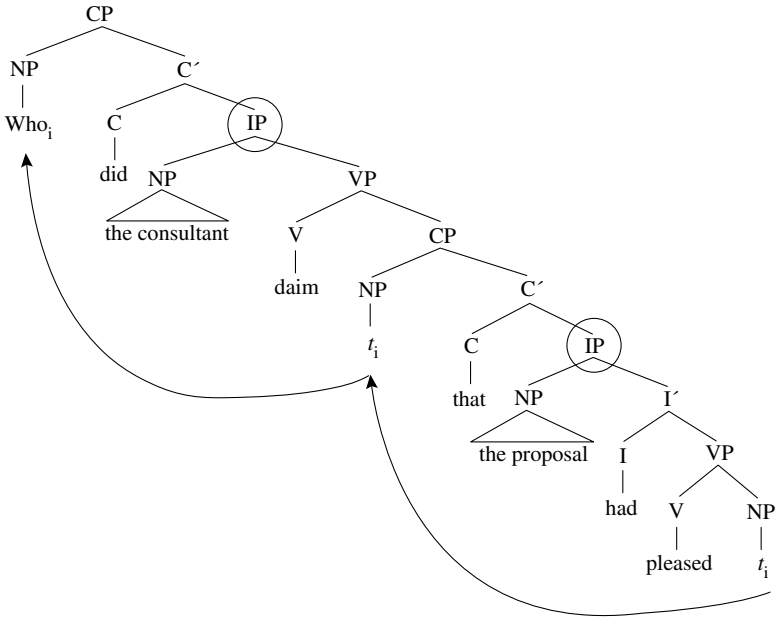
steps, leaving an empty category in the specifier position of the intervening CP. Thus one long ungrammatical movement step can take place as two shorter grammatical movements. For example, consider how this framework applied to account for the grammaticality contrast between (1) and (2).[1]

(1) Who$_i$ did the consultant claim t_i that the proposal had pleased t_i?

(2) *Who$_i$ did the consultant wonder which proposal had pleased t_i?

In (1) the *wh*-filler *who* is interpreted as the object of the verb *pleased*, as indicated by the coindexed empty category in this position. If the movement from object position of *pleased* to the matrix CP specifier position were to take place in one derivational step, this movement would cross two bounding nodes: the IP headed by *pleased* and the IP headed by *did claim*. The specifier position of the embedded CP offers an intermediate landing site for the long-distance movement, so that the movement from the object position of *pleased* to this position crosses only one bounding node, and the subsequent movement from this position to the matrix Spec,CP position crosses one further bounding node as shown in (3).

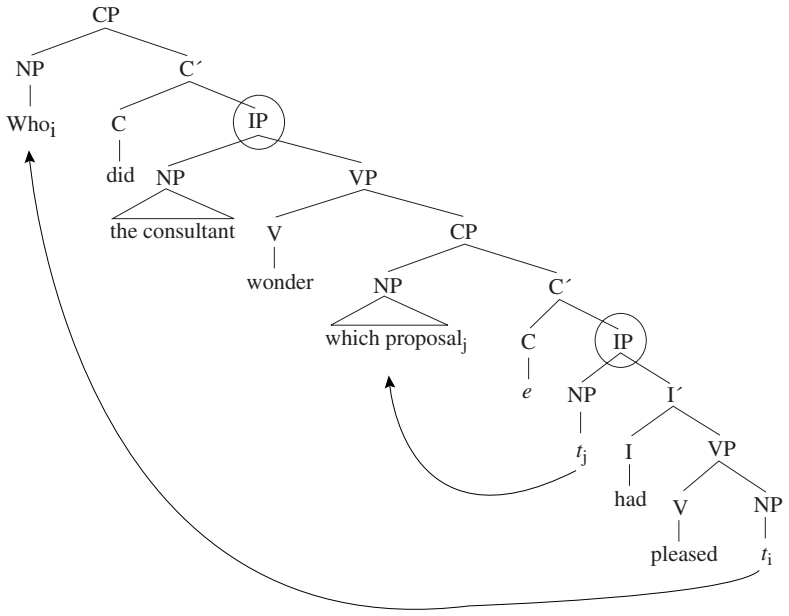
(3)



[1] Following standard practice in the psycholinguistic community, we adopt the sentence-level phrase-structure assumptions from Chomsky (1986), according to which a sentence is an I(nflection) phrase (IP) which is dominated by a C(omplementizer) phrase (CP). None of our results depend on these assumptions.

In (2) the movement from the object position of the verb *pleased* to the Spec,CP position in the matrix clause crosses two bounding nodes: the embedded and matrix IPs (see (4)). In contrast to the previous example, the intervening Spec,CP position is already filled, so there is no position in which an intermediate empty category can be posited. Hence, this extraction violates Subjacency.

(4)



In nontransformational theories, features of a fronted element (such as a *wh*-phrase) are passed from a head to its dependents to allow the existence of grammatical long-distance dependencies. To block ungrammatical long-distance dependencies in such theories, there are constraints on the feature passing between heads and dependents. For example, under Generalized Phrase Structure Grammar (GPSG; Gazdar et al. 1985) and Head-driven Phrase Structure Grammar (HPSG; Pollard & Sag, 1994) the feature that is passed from head to dependent in *wh*-extractions is called a slash feature. Slash features can be passed down most right-branching dependents as in (1), but there are restrictions on passing slash features through CPs with further extractions, as in (2). See also Kaplan and Bresnan 1982 and Zaenen 1983 for a related approach within the Lexical Functional Grammar (LFG) framework, and Steedman 1996 for an approach within Combinatory Categorial Grammar.

Perhaps the most convincing evidence for the existence of intermediate structures in long-distance dependencies is the fact that in some languages there is a morphological change in the complementizer and/or verb in an extraction domain. See Chung 1982, 1994 for evidence from Chamorro (cf. Dukes 1992);

Georgeopoulos 1985, 1991 for evidence from Palauan; Torrego 1984 for evidence from Spanish; and McCloskey 2000 for evidence from Irish English.

In this paper, we use a method from the sentence-processing literature—self-paced reading—to attempt to evaluate the existence of intermediate linguistic structures in long-distance dependencies in English, a language without a morphological reflex of such an intermediate structure. It is important to obtain converging evidence for any strong theoretical claim, such as the intermediate structure hypothesis. Reaction-time methods provide one such converging data point. A number of studies have attempted to find reaction-time evidence for the existence of empty categories such as traces of *wh*-movement (Swinney, Ford, Frauenfelder & Bresnan 1988; Nicol & Swinney 1989; Nicol, Fodor & Swinney 1994; cf. McKoon, Ratcliff & Albritton 1996; Chen, Wolf & Gibson 2002) or NP-movement (Bever & McElree 1988; MacDonald 1989; Bever & Sanz 1997),[2] but there is little psycholinguistic evidence relevant to the existence of intermediate empty categories (or corresponding slash categories, etc.) in long-distance dependencies. In the only previous attempt that we are aware of to study such structures psycholinguistically, Frazier and Clifton (1989) compared extractions across a single clause to extractions across two clauses and found that extractions across two clauses were more complex, even when compared with nonextracted versions. They interpreted their results as suggestive evidence in favor of intermediate structures in long-distance dependencies ("successive cyclicity"). However, there are a number of difficulties in interpreting their results. First, the example items were not independently controlled for lexical and plausibility factors. Second, the items that they tested were all locally ambiguous, with the consequence that the effects that they observed can be independently accounted for in terms of ambiguity-resolution preferences, such as the Minimal Chain Principle (De Vincenzi 1991) or the dependency locality theory (DLT; Gibson 1998, 2000).

The evidence reported here relies on the observation that reading times for a word increase as the linear distance increases between the word and the position to which the word connects in the structure for the input so far, as proposed by the DLT (Gibson 1998, 2000; Grodner, Watson & Gibson 2000). According to the DLT, integrating a newly input syntactic head *h* into the current structure for the input requires reactivation/retrieval of the element(s) in the current structure to which *h* connects syntactically (e.g., is part of a head-dependent relationship or is part of the same syntactic chain). The difficulty of reactivating previous elements in the structure depends on how far back in the input they have occurred, and what kind of elements

[2] See Pickering and Barry 1991 and Sag and Fodor 1994 for accounts of most of these studies without relying on empty categories. See also Pickering and Barry 1991 and Traxler and Pickering 1996 for evidence against the empty category accounts, but see Gibson and Hickok 1993 for an empty-category treatment of these observations.

have occurred in the interim, and how much they interfere with each other and the head to be connected (cf. Lewis 1996, Gordon et al. 2001). For example, Grodner et al. (2000) found a high correlation between reading times at the word positions in items like (5) and (6) and integration cost as defined in (7) and (8).

(5) The reporter who sent the photographer to the editor hoped for a good story.

(6) The reporter who the photographer sent to the editor hoped for a good story.

(7) Integration cost of a syntactic head: The cost of syntactically integrating a newly input syntactic head h_2 to a syntactic head h_1 in the current structure is proportional to the sum the number of new referents in the discourse (e.g., nouns and verbs whose referents have not yet been mentioned) that have been processed since h_1 was last activated (adapted from Gibson 2000).

(8) Integration cost at a word *w*: The integration cost at a word is the sum of all integrations that take place at *w*.

For example, people read the embedded verb *sent* faster when it is in a subject-extracted relative clause (RC) as in (5) than when it is in an object-extracted RC as in (6). In (5), a local integration takes place at *sent*: the *wh*-pronoun *who* is linearly adjacent to the verb *sent*, to which it integrates as subject and agent of *sent*. There are two potential reasons why reading the verb *sent* in (6) is more difficult. First, two integrations take place at the word *sent* in (6), as opposed to only one in (5): (a) the NP *the photographer* is interpreted as the subject and agent of *sent* (a local integration in which the elements being integrated are linearly adjacent), and (b) the *wh*-filler *who* is interpreted as the object and patient of *sent* (possibly mediated through an empty object). Furthermore, the second of these integrations is a nonlocal integration: the empty object of *sent* must be integrated back to the *wh*-filler *who* across the NP *the photographer* and the verb *sent*. The combination of these two integrations leads to greater reading times at *sent* in (6) than in (5). Additionally, people read the matrix verb *hoped* slower than the embedded verb *sent* in the subject-extracted RC (5), presumably because the integration is longer distance at the matrix verb (between *the reporter* and *hoped*) than at the embedded verb in (5).

Gibson (1998, 2000) hypothesizes that integration cost is probably sensitive to a number of factors—anything that might cause a decrease in activation of earlier elements in the input. Thus, having additional words between two elements to be integrated might cause increased integration difficulty. He proposes that new discourse structure in terms of discourse referents causes a

substantial integration cost increment and that is why integration cost is initially defined in terms of discourse structure. Part of the empirical motivation for quantifying distance in terms of new discourse structure comes from experiments that compare materials with pronouns in embedded positions (which refer to current elements in the discourse) to materials with NPs that introduce new discourse elements in these positions. For example, if the embedded subject *the photographer* in (6) is replaced with the pronoun *you*, then reading times at the embedded verb and main verb are much faster (Warren & Gibson 2002, Gordon, Hendrick & Johnson 2001).[3]

Integration costs as defined in (7) and (8) account for many other psycholinguistic phenomena, including preferences in ambiguity resolution (Gibson, Pearlmutter, Canseco-Gonzales & Hickok 1996; Altmann, van Nice, Garnham & Henstra 1998; Pearlmutter & Gibson 2001), the complexity of nested structures (Gibson 1998) and heaviness effects (Bever 1970, Hawkins 1994).

2. Experiment

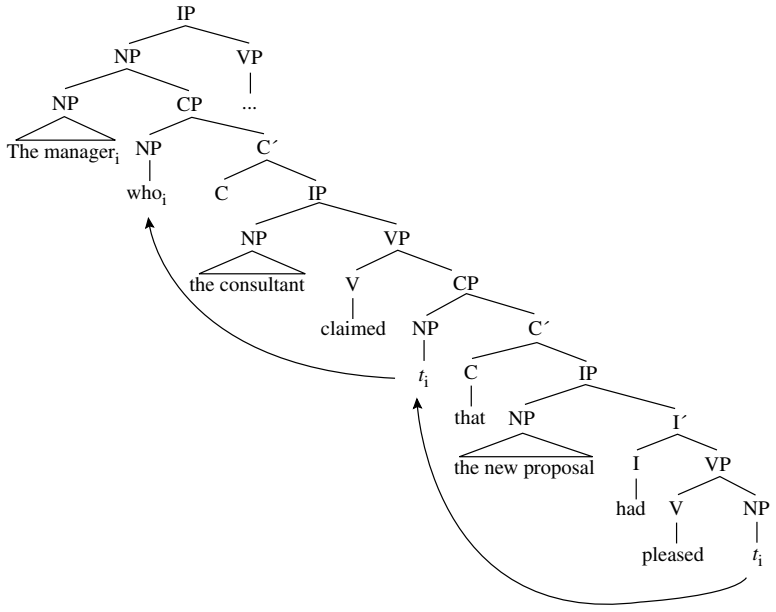
Given that longer distance integrations take longer to process, the intermediate structure hypothesis predicts that processing times at a site where a *wh*-filler gets its interpretation—the most embedded gap site—should be faster when there is an intermediate structure to be integrated with, compared to a case in which there is no intermediate structure to be integrated with, because such a connection will be longer if there is no intermediate structure. For example, consider (9), an RC version of the matrix-sentence long-distance dependency in (1). For expository purposes, we adopt the transformational empty-category theory for our discussion of the relevant examples (see Frazier 1993 for some suggestive evidence in favor of the transformational analysis over the slash-passing analysis).

(9) Extraction across a VP (+intermediate structure):
The manager who$_i$ the consultant claimed that the new proposal had pleased t_i will hire five workers tomorrow.

Under all current syntactic theories, there is some coindexed linguistic structure mediating the connection between the filler *who* and the embedded object position following *pleased*, as in the intermediate trace theory representation pictured in (10).

[3] Gordon et al. (2001) interpret the faster reading times for object-extracted relative clauses with pronouns to be because there is less interference between pronouns and full definite noun phrases, like the head noun in (6) *the reporter*. This debate is not relevant for the purposes here: either of these proposals (or some combination of the two) provides a definition of integration distance that suffices. See Warren, Gibson, Jameson, and Hirsch 2003 for further data relevant to this question.

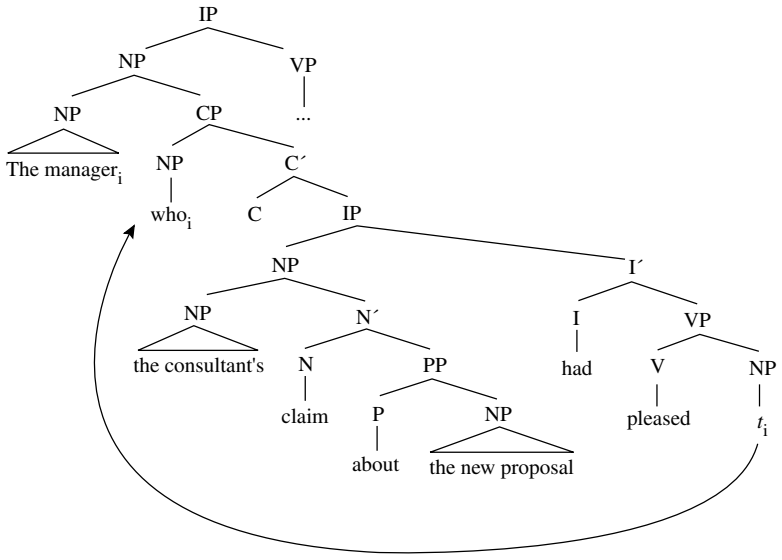
(10)



If the clause *the consultant claimed* is nominalized as in (11) and (12), there is no structure mediating the *wh*-filler *who* and the object position of the verb *pleased*.

(11) Extraction across an NP (–intermediate structure):
The manager who$_i$ the consultant's claim about the new proposal had pleased t_i will hire five workers tomorrow.

(12)



Because there is no structure mediating the relationship between the filler *who* and the verb *pleased* in (11), the integration of the verb *pleased* with the preceding filler crosses a longer distance than the integration of the verb *pleased* with its preceding intermediate structure in (9) (e.g., under the empty category theory, the integration with the empty category following *claimed*). Thus, the intermediate structure hypothesis predicts that processing the verbal region *had pleased* in (9) should be faster than processing the same region in (11). The experiment reported here tests this prediction.

There is a confounding influence in the comparison between (9) and (11) that also predicts that processing the verbal region will be faster in (9) than in (11). In particular, the distance between the verb *pleased* and the head noun *proposal* of its subject NP is shorter in (9) than the distance between the verb *pleased* and the head noun *claim* of its subject NP in (11). The longer subject-verb distance in (11) should also cause longer reading times at the verb in (11) than in (9). To control for this confound in the comparison between (9) and (11), two nonextracted control conditions were also constructed:

(13) No extraction, local subject-verb integration (VP):
The consultant claimed that the new proposal had pleased the manager who will hire five workers tomorrow.

(14) No extraction, nonlocal subject-verb integration (NP):
The consultant's claim about the new proposal had pleased the manager who will hire five workers tomorrow.

Examples (13) and (14) are formed from (9) and (11), respectively, by making the RC the main clause in each, and by placing the subject NP of (9) and (11) *the manager* into its extracted object position. The resulting sentences have the same subject-verb integration distances at the verb *pleased* as their extracted counterparts, but without integrations involving the verb and an extracted RC pronoun. Thus the intermediate structure hypothesis predicts that the difference in reading times during the region *had pleased* should be larger for the extracted conditions than for the nonextracted conditions. In other words, an interaction between extraction (+,−) and intervening phrasal type (NP, VP) is predicted during this region.

The intermediate structure hypothesis also predicts reading-time delays at the points where intermediate structures are posited. In particular, if there is an empty category in the specifier of the CP mediating the long-distance dependency between the RC pronoun *who* and its role-assigning verb *pleased*, then reading times should be slower at the point of processing this intermediate empty category as compared with the control condition in which no intermediate empty category is posited. Unfortunately, it is difficult to compare the regions of interest in this case because of substantial lexical differences among the conditions in the target regions. In the VP condition the region of interest is the word *that* preceded by the NP *the consultant* and the

verb *claimed*, whereas in the NP condition the region of interest is the word *about* preceded by the genitive NP *the consultant's claim*. The intermediate structure hypothesis predicts that the difference between the reading times between *that* and *about* in the extracted condition (where an intermediate structure would be posited) should be larger than the difference between the reading times for the same region in the nonextracted versions. Thus an interaction between extraction (+,−) and intervening phrasal type (NP, VP) is predicted at this word but in the reverse direction of the predicted interaction at the VP region *had pleased*. However, the fact that the lexical differences are so substantial in and before this region will make effects in this region hard to interpret.

The design of the experiment allows one additional prediction to be tested. In the +extraction conditions (9) and (11), a *wh*-filler *who* is being held awaiting its θ-role-assigning verb while the region *the consultant claimed that the new proposal…* is being processed. The same integrations take place in the same region in the –extraction conditions in (13) and (14), but without the *wh*-filler pending. If online reading times are sensitive to syntactic storage of this kind (cf. Wanner & Maratsos 1978; Gibson 1991, 1998; Grodner, Gibson & Tunstall 2002; Chen et al. 2002), then reading times should be slower in the +extraction conditions than in the –extraction conditions during the processing of this region.[4]

2.1 Method

2.1.1 *Participants*

The participants were 98 native English speakers, students and other affiliates of MIT, who were paid for their participation.

2.1.2 *Materials*

Twenty items were constructed, each with the four versions described above in a 2 × 2 design: extraction (+,−) crossed with intervening phrasal type (NP, VP). The nonextracted versions were constructed from the extracted versions by changing the RC of the extracted version into the main clause of the nonextracted version, and by placing the main-clause subject from the extracted version into its object position in the nonextracted version. The verb

[4] The discussion of online storage costs raises the possibility of an alternative explanation of a potential interaction during the critical region *had pleased*: If syntactic storage and the process of syntactic integration access the same pool of computational resources, then it is possible that such an interaction might occur when storage and integration are both being heavily taxed, as in the +extraction, NP condition, during the target region. However, such a possibility is unlikely in light of experimental evidence from Gibson et al., in press, in which syntactic storage and integration were both manipulated between low and high values, and no suggestion of such an interaction was observed.

phrase of the main clause in the extracted version was then added to the end of the nonextracted version, as an RC initiated by the relative pronoun *who*.

The +extraction, VP versions of the items were constructed with the goal that the *wh*-filler would be initially preferred to be interpreted as the object of the embedded verb. To allow the possibility of a long-distance dependency through the embedded clause, the intermediate verb was always a bridge verb, such as *claim* in (9).

Furthermore, to ensure that participants did not mistakenly posit an empty category in the direct-object position of the intermediate verb in the +extraction, VP version of an item (e.g., following *claim* in (9)), the verbs that were used in this position were strongly biased toward taking a sentence complement rather than an NP object, according to the norms from Garnsey, Pearlmutter, Myers, and Lotocky 1997. To the extent that any of these verbs could take an NP object, they could only take an inanimate NP object. To make the possibility more remote that an object empty category might be posited, the relativized NP was therefore always animate (e.g., *the manager* in (9)). A set of 11 verbs satisfied all of these requirements. None of these verbs was used in more than three items. Appendix A presents a complete list of the stimuli.

In addition to the 20 experimental items, 70 filler sentences of various types were constructed. These included 40 items from two other experiments with unrelated hypotheses. *Yes*/*no* comprehension questions were written for all stimuli as well. The experimental stimuli and fillers were combined to form four 90-item lists. The experimental stimuli were counterbalanced across lists, such that each version of an experimental stimulus item appeared in exactly one list, and each list contained the same number of items in each condition. Each participant saw exactly one list in a different pseudo-random order.

2.1.2.1 *Plausibility norming survey* It is well known that lexical frequencies and plausibilities have significant effects on comprehension difficulty as measured by reading times (e.g., MacDonald, Pearlmutter & Seidenberg 1994; Trueswell, Tanenhaus & Garnsey 1994; Garnsey et al. 1997; Gibson & Pearlmutter 1998). Lexical frequency is not an issue in the critical region of the current experiment because the same verb is used in extractions across both the NP and the VP intervening region (e.g., *pleased* in both (9) and (11)). There is a potential plausibility difference in the subject-verb integration taking place at this verb (e.g., in (9) the head noun of the subject NP to be integrated is *proposal*, whereas in (11) the head noun of the subject NP to be integrated is *claim*), but the same plausibility difference is also present in the nonextracted versions, so this plausibility difference is not an issue. There is a second potential plausibility difference at the point of processing the verb *pleased* that is not controlled in the comparison with the nonextracted versions: the object-verb integration. Although the verb and the NP interpreted as the object of this verb are the same (*pleased* and *the manager* in (9) and (11)), the fact that the subject NP for this verb is different will affect the

plausibilities of integrating the object into the semantic/discourse representation. For example, the plausibility of a proposal pleasing a manager (as in (9)) may be different from the plausibility of a claim about a proposal pleasing a manager (as in (11)), and this difference is independent of the difference between a claim's pleasing and a proposal's pleasing on their own.

A questionnaire was conducted to investigate this potential plausibility difference. Twenty-eight native English-speaking participants from the MIT area who did not take part in the self-paced reading experiment completed the survey. The items tested in this questionnaire were simplified versions of the nonextracted experimental items, without any material following the matrix verb direct object, as in (15) and (16).

(15) VP plausibility control:
The consultant claimed that the new proposal had pleased the manager.

(16) NP plausibility control:
The consultant's claim about the new proposal had pleased the manager.

Participants rated plausibility on a scale of 1 (natural) to 7 (unnatural). They were asked to judge the naturalness in the real world of the events described in the sentences—that is, how likely they were to occur.

The 20 stimulus items were interspersed with 70 filler sentences of various types. These included 25 items from another experiment with an unrelated hypothesis. Each participant rated exactly one version of each item. Each participant encountered the sentences in a different pseudo-random order.

The results of the survey were that five of the 20 items were found to be significantly more plausible ($p < .05$ by t-test) in one version than in the other (three in one direction, two in the other). These five items were therefore omitted from the reading-time analyses. The remaining 15 items were closely matched for plausibility across all versions. Collapsing across items, the overall means for the VP condition and the NP condition were very close, with a mean of 3.20 for the VP condition and 3.18 for the NP condition. The plausibility ratings for each item are presented along with the items in Appendix A.

2.1.3 *Procedure*

Participants were timed in a word-by-word self-paced noncumulative moving-window reading task (Just, Carpenter & Woolley 1982) controlled by custom software on a Macintosh Centris computer. Participants pressed the space bar to reveal each subsequent word and cause all other words to revert to dashes. At the end of each sentence, a *yes*/*no*-question appeared on the screen, which participants answered by pressing one of two keyboard keys. Participants were informed by a screen message when they answered incorrectly, to encourage

them to keep paying attention to the content of the sentences. The experimental trials were preceded by two screens of instructions and eight practice trials. The experiment took participants approximately 20 minutes. Some participants took part in an unrelated experiment either immediately before or after the current study. For these participants, the sequence of experimental sessions never took longer than an hour in total.

2.2 Analysis

For the purposes of analysis and presentation of the reading time data, items were separated into seven regions as illustrated in Table 1. The critical regions relevant to the intermediate structure hypothesis are regions 5 and 6, at the point of processing the most embedded verbal region (*had pleased*), and the words immediately following this region (*will hire* in the extracted conditions, *the manager* in the nonextracted conditions).

2.3 Results

The data for the five items whose plausibility scores were significantly different in the NP and VP conditions were omitted from the analyses. The data patterns are similar even when these items are left in the analyses. The data for six participants were also omitted from the analyses, because of comprehension-question response accuracies below 75% on the set of 30 filler items that were not part of any subexperiment. Because the target items in this experiment and in the unrelated experiment were more complex than the filler items, the response accuracy of this group of participants was even lower when all items were considered.

2.3.1 *Comprehension-question response accuracy*

The response accuracies for the four conditions, expressed as percentages, are presented in Table 2. Participants were correct in answering questions to the nonextracted conditions significantly more often than in answering questions to the extracted conditions, $F_1(1,91) = 41.0$, $MSe = 3.21$, $p < .001$; $F_2(1,14) = 16.4$, $MSe = 0.94$, $p = .001$. Participants were also better at answering questions about the VP conditions than about the NP conditions, but only in the participants analysis, $F_1(1,91) = 6.89$, $MSe = 2.46$, $p = .01$; $F_2\ (1,14) = 2.68$, $MSe = 1.06$, $p = .12$. There was no interaction between extraction type and intervening phrasal type (*ps* > .20).

2.3.2 *Reading times*

To adjust for differences in word length across conditions as well as overall differences in participants' reading rates, a regression equation predicting reading time from word length was constructed for each participant, using all

Table 1. Analysis regions for the experiment

Condition	**Region**						
	1	**2**	**3**	**4**	**5**	**6**	**7 (rest of sentence)**
+extract, VP	The manager who	the consultant claimed	that	the new proposal	had pleased	will hire	five workers tomorrow.
+extract, NP	The manager who	the consultant's claim	about	the new proposal	had pleased	will hire	five workers tomorrow.
–extract, VP	—	the consultant claimed	that	the new proposal	had pleased	the manager	who will hire five workers tomorrow.
–extract, NP	—	the consultant's claim	about	the new proposal	had pleased	the manager	who will hire five workers tomorrow.

Table 2. Percentages of correctly answered comprehension questions

Extraction	**Intervening phrasal type**	
	VP	**NP**
+extract	80 (2.1)	73 (2.6)
–extract	90 (1.8)	88 (1.7)

Note: Standard errors (computed over participant means) are presented in parentheses.

filler and experimental items (Ferreira & Clifton 1986; see Trueswell et al. 1994 for discussion). At each word position, the reading 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 greater than 3000 ms were excluded from analyses, affecting less than 0.5% of the data.

Mean word-by-word residual reading times computed across participants are plotted in Figure 1 for each condition in each of the seven regions. The data patterns are numerically the same as those in the reading times before the normalizing procedure, although not all of the critical tests reached significance in the prenormalized data. See Appendix B for a complete set of raw and residual reading time means by condition and position.

No tests are reported for region 1 (*the manager who*), as the conditions did not differ at this point. In region 2 (*the consultant claimed* in the VP conditions; *the consultant's claim* in the NP conditions) the extraction conditions were read significantly slower than the nonextraction conditions, $F_1(1,91) = 10.3$, $MSe = 11{,}298$, $p < .005$; $F_2(1,14) = 19.8$, $MSe = 1037$, $p < .001$, as predicted by syntactic storage-cost differences between the two conditions. There was also a suggestion that the NP conditions were slower

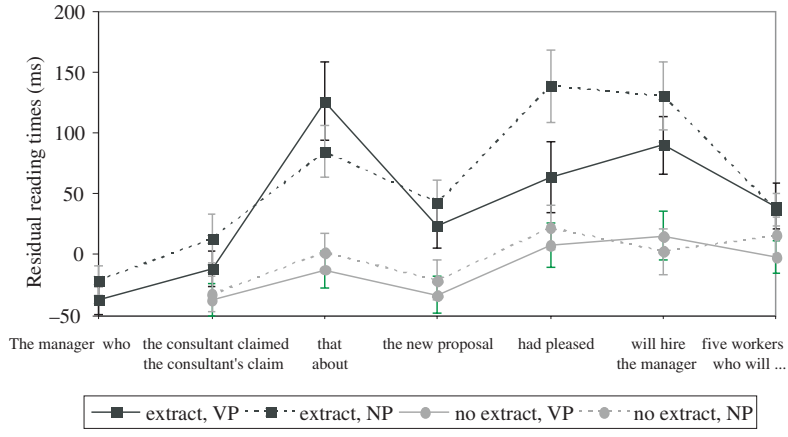


Figure 1. Residual reading times for the experiment

than the VP conditions in this region, but the effect only reached marginal significance in the participants analysis, $F_1(1,91) = 2.84$, $MSe = 7138$, $p = .10$, and was not significant in the items analysis, $F_2(1,14) = 1.71$, $MSe = 1379$, $p > .20$. There was no interaction between extraction type and intervening phrasal type in this region.

In region 3 (*that* in the VP conditions; *about* in the NP conditions) the extraction conditions were again read significantly slower than the nonextraction conditions, $F_1(1,91) = 29.1$, $MSe = 38{,}687$, $p < .001$; $F_2(1,14) = 34.7$, $MSe = 5232$, $p < .001$. There was no difference between the NP and VP conditions in this region ($Fs < 1$). However, there was a marginal interaction between extraction type and intervening phrasal type in the participants analysis, $F_1(1,91) = 3.13$, $MSe = 23{,}024$, $p = .08$, but this effect was not significant in the items analysis, $F_2(1,14) = 1.13$, $MSe = 3844$, $p = .31$.

In region 4 (*the new proposal*) the extraction conditions were read significantly slower than the nonextraction conditions, $F_1(1,91) = 24.0$, $MSe = 13{,}893$, $p < .001$; $F_2(1,14) = 32.5$, $MSe = 1867$, $p < .001$. There was also a suggestion that the NP conditions were slower than the VP conditions in this region, but the effect did not quite reach significance in either analysis, $F_1(1,91) = 2.61$, $MSe = 8619$, $p = .11$; $F_2(1,14) = 2.92$, $MSe = 1115$, $p = .11$. There was no interaction between extraction type and intervening phrasal type in this region.

In the critical region, region 5 (*had pleased*), the predicted interaction between extraction type and intervening phrasal type was significant, $F_1(1,91) = 4.36$, $MSe = 19{,}265$, $p < .05$; $F_2(1,14) = 5.17$, $MSe = 4405$, $p < .05$. Additionally, there were two main effects: (a) the extraction conditions were read significantly slower than the nonextraction conditions, $F_1(1,91) = 17.1$, $MSe = 39{,}731$, $p < .001$; $F_2(1,14) = 34.6$, $MSe = 3349$, $p < .001$; and (b) the NP conditions were read slower than the VP conditions, $F_1(1,91) = 12.5$, $MSe = 14{,}709$, $p < .001$; $F_2(1,14) = 14.0$, $MSe = 3020$, $p < .005$. The main effect of extraction was robust across the two individual comparisons (NP: $F_1(1,91) = 21.8$, $MSe = 28{,}510$, $p < .001$; $F_2(1,14) = 20.8$, $MSe = 5809$, $p < .001$; VP: $F_1(1,91) = 4.7$, $MSe = 30{,}485$, $p < .05$; $F_2(1,14) = 9.22$, $MSe = 1945$, $p < .01$), but the main effect of intervening phrasal type was significant only for the extraction conditions, $F_1(1,91) = 10.9$, $MSe = 23{,}687$, $p = .001$; $F_2(1,14) = 12.2$, $MSe = 5193$, $p < .005$, not for the nonextraction conditions ($Fs < 1$).

The pattern in region 6 (*will hire* in the extraction conditions; *the manager* in the nonextraction conditions) was similar to the pattern in region 5. The extraction conditions were read significantly slower than the nonextraction conditions, $F_1(1,91) = 37.8$, $MSe = 24{,}872$, $p < .001$; $F_2(1,14) = 41.2$, $MSe = 3963$, $p < .001$, but one should keep in mind that the lexical items in this comparison are different. There was no difference in reading times for the intervening phrasal type comparison ($Fs < 1$). There was a marginal interaction between extraction and intervening phrasal type in the items analysis, $F_2(1,14) = 3.81$, $MSe = 3123$, $p = .07$, but this effect was not

significant in the participants analysis, $F_1(1,91) = 2.17$, $MSe = 29{,}783$, $p = .14$. When regions 5 and 6 are treated as a single analysis region, the interaction between extraction and intervening phrasal type is significant, $F_1(1,91) = 6.09$, $MSe = 12{,}818$, $p < .02$; $F_2(1,14) = 7.92$, $MSe = 2135$, $p = .01$.

2.4 Discussion

The only significant effect that was observed in the regions prior to the critical region was that the +extraction conditions were read more slowly than the –extraction conditions in regions 2–4. This effect is as predicted by an online syntactic storage-cost theory, because an additional NP and a relative pronoun (*the manager* and *who*) need to be stored in working memory in these regions in the +extraction conditions (Wanner & Marastos 1978; Gibson 1991, 1998, 2000).

In region 5, the extracted NP condition was slower than the extracted VP condition, as predicted by the intermediate structure hypothesis. Furthermore, the interaction between extraction and intervening phrase type observed in region 5 was also predicted by the intermediate structure hypothesis. The main effect of extraction in this region reflects the additional integration cost of linking the object gap to the earlier *wh*-filler for the extraction conditions. Although there was also a main effect of intervening category in this region, such that the intervening NP conditions were processed more slowly than the intervening VP conditions, this effect was not present in the comparison involving the nonextracted conditions. The main effect is therefore probably a reflection of the statistical interaction between the two factors.

In region 6, the extracted conditions were significantly slower than the nonextracted conditions. This difference is probably due to the fact that there is a long integration between the matrix subject NP *the manager* and the matrix verbal region *will hire* in the extraction condition, whereas in the nonextraction condition there is only a local integration between the verb *pleased* and its direct object *the manager*. Thus, this difference is predicted by the distance-based integration hypothesis. The marginal interaction observed in this region is very similar to the effect observed in the previous region and may reflect a spill-over effect from that region, as can often occur in self-paced reading.

The suggestion of an interaction in the reverse direction in region 3, at the point when an intermediate structure would be posited under an empty-category-based theory, might be taken as suggestive support for the intermediate structure hypothesis. However, one must keep in mind that the lexical items are different in this comparison, making any interpretation difficult. Furthermore, the effect was not significant in the items analysis. Moreover, there are other possible explanations of the marginal interaction in region 3. An unlikely possibility is that participants may be initially postulating an object gap in this position, following the Active Filler Strategy

(Clifton & Frazier 1989), the Minimal Chain Principle (De Vincenzi 1991) or the DLT (Gibson 1998, 2000). Participants would then slow down when the lexical item *that* is encountered, disconfirming their empty-category hypothesis. However, this possibility is extremely unlikely, because (a) the bridge verbs preceding this position (e.g., *claimed*) were all highly biased toward taking sentence complements; and (b) to the extent that the verbs that were used can take an NP object, the only kind of object that they allow is an inanimate NP, but the object gap that would be posited here is coindexed with an animate NP (e.g., *the manager*).

Another potential explanation of the marginal effect in region 3 is that participants may be initially postulating an empty category in the subject position of an embedded sentence in the extracted VP condition, and the presence of the complementizer *that* informs them that their initial hypothesis was wrong, leading to longer reading times.[5] Evidence supporting this hypothesis comes from data suggesting that people prefer to posit a gap in subject position of an embedded clause in a long-distance extraction across clauses rather than in object position (Kluender & Cowles 1997). Thus the marginal interaction in region 3 does not necessarily support the intermediate structure hypothesis; it might be reflective of a small local ambiguity effect.

This observation raises the possibility that the crucial interaction at region 5 (*had pleased*) could indirectly be caused by people's tendency to *incorrectly* posit a gap in the subject position of the embedded clause. In particular, suppose that positing all empty elements associated with a head noun makes processing later referents to that noun easier, whether or not they are in the same syntactic chain. Then positing a gap at an incorrect early location might make positing the gap at the true site after the verb easier than in the NP condition, in which no such gap is posited. But this hypothesis is unlikely because of evidence from the processing of RC structures from King and Just (1991) and Grodner et al. (2000) on examples like (5) and (6), repeated here.

(5) The reporter$_i$ who$_i$ t_i sent the photographer to the editor hoped for a good story.

(6) The reporter$_i$ who$_i$ the photographer sent t_i to the editor hoped for a good story.

The gap associated with the *wh*-filler *who* occurs later in the object extraction in (6) than in the subject extraction in (5). If positing all empty elements associated with a head noun makes processing later referents to that noun easier, whether or not they are in the same syntactic chain, then readers should have less difficulty at the matrix verb *hoped* in the object extraction (6) than in the subject extraction (5), because (a) the referent of the filler *the reporter* needs to be accessed to process the verb *hoped*, and (b) the filler

[5] Thanks to Matthew Walenski for suggesting this possibility.

occurs more recently in the input in the object extraction than in the subject extraction. Contrary to prediction, there was a trend in the reverse direction in both experiments: reading times are numerically slower at the matrix verb *hoped* for the object extraction than for the subject extraction. Thus the speeded reading times for coreferent gaps and lexical NPs only seems to occur when the two are part of the same syntactic chain.

3. Concluding Remarks

The results of the experiment reported here demonstrate that reading times are faster at the θ-role-assigning verb for extractions through a clause, as in (9), than for extractions across a nominalization of the clause, as in (11). Current linguistic theories all propose the existence of structures mediating long-distance extractions through clauses, as in (9), but do not propose such mediating structures in extractions across a subject NP, as in (11). Given the observation that longer distance integrations take longer to perform, current linguistic theories therefore predict the observed reading-time difference, because the integration between the object empty category and its preceding coindexed NP is more local in the structure containing an intermediate coindexed structure. The subject-verb integration is also more local in the cross-clause condition than in the cross-NP condition, but reading times on two additional control conditions demonstrate that the reading-time difference at the verb is over and above any difference due to the difference in subject-verb integrations. A theory that did not propose intermediate structure in clauses, such as linguistic theories prior to 1973, would not predict the observed results. Thus the results provide experimental evidence for the existence of structures intermediating long-distance extractions through clauses.

References

ALTMANN, G. T. M., K. Y. VAN NICE, A. GARNHAM & J. A. HENSTRA. 1998. Late closure in context. *Journal of Memory and Language* 31:685–712.

BEVER, T. G. 1970. The cognitive basis for linguistic structures. In *Cognition and the development of language*, ed. J. R. Hayes, 279–362. New York: Wiley.

BEVER, T. G. & B. MCELREE. 1988. Empty categories access their antecedents during comprehension. *Linguistic Inquiry* 19:35–43.

BEVER, T. & M. SANZ. 1997. Empty categories access their antecedents during comprehension of Spanish unaccusatives. *Linguistic Inquiry* 28:69–91.

CHEN, E., F. WOLF & E. GIBSON. 2002. Syntactic storage costs in sentence comprehension. Poster presented at the 15th annual CUNY Conference on Human Sentence Processing, CUNY, New York.

CHOMSKY, N. 1973. Conditions on transformations. In *A Festschrift for Morris Halle*, ed. S. Anderson & P. Kiparsky, 232–286. New York: Holt, Rinehart & Winston.

CHOMSKY, N. 1977. On *wh*-movement. In *Formal syntax*, ed. P. W. Culicover, T. Wasow & A. Akmajian, 71–132. New York: Academic Press.

CHOMSKY, N. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.
CHOMSKY, N. 1986. *Barriers*. Cambridge, Mass.: MIT Press.
CHUNG, S. 1982. Unbounded dependencies in Chamorro grammar. *Linguistic Inquiry* 13:39–77.
CHUNG, S. 1994. *Wh*-agreement and "referentiality" in Chamorro. *Linguistic Inquiry* 25:1–44.
CLIFTON, C. JR., & L. FRAZIER. 1989. Comprehending sentences with long-distance dependencies. In *Linguistic structure in language processing*, ed. G. Carlson & M. Tanenhaus, 273–317. Dordrecht: Kluwer.
DE VINCENZI, M. 1991. Syntactic parsing strategies in Italian. Dordrecht: Kluwer.
DUKES, M. 1992. On the status of Chamorro *wh*-agreement. Paper presented at West Coast Conference on Formal Linguistics 11, University of California, Los Angeles.
FERREIRA, F. & C. CLIFTON, JR. 1986. The independence of syntactic processing. *Journal of Memory and Language* 25:348–368.
FRAZIER, L. 1993. Processing Dutch sentence structures. *Journal of Psycholinguistic Research* 22:85–108.
FRAZIER, L. & C. CLIFTON, JR. 1989. Successive cyclicity in the grammar and parser. *Language and Cognitive Processes* 4:93–126.
GARNSEY, S. M., N. J. PEARLMUTTER, E. MYERS & B. LOTOCKY. 1997. The relative contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language* 37:58–93.
GAZDAR, G., E. KLEIN, G. K. PULLUM & I. A. SAG. 1985. *Generalized phrase structure grammar*. Oxford: Blackwell.
GEORGEOPOULOS, C. 1985. Variables in Palauan syntax. *Natural Language & Linguistic Theory* 3:59–94.
GEORGEOPOULOS, C. 1991. *Syntactic variables, resumptive pronouns, and A′* binding in Palauan. Dordrecht: Kluwer.
GIBSON, E. 1991. A computational theory of human linguistic processing: Memory limitations and processing breakdown. Ph.D. thesis, Carnegie Mellon University, Pittsburgh, Penn.
GIBSON, E. 1998. Linguistic complexity: Locality of syntactic dependencies. *Cognition* 68:1–76.
GIBSON, E. 2000. The dependency locality theory: A distance-based theory of linguistic complexity. In *Image, language, brain*, ed. Y. Miyashita, A. Marantz & W. O'Neil, 95–126. Cambridge, Mass.: MIT Press.
GIBSON E., T. DESMET, D. GRODNER, D. WATSON & K. KO. In press. Reading relative clauses in English. *Cognitive Linguistics*.
GIBSON, E. & G. HICKOK. 1993. Sentence processing with empty categories. *Language and Cognitive Processes* 8:147–161.
GIBSON, E. & N. PEARLMUTTER. 1998. Constraints on sentence comprehension. *Trends in Cognitive Science* 2:262–268.
GIBSON, E., N. PEARLMUTTER, E. CANSECO-GONZALES & G. HICKOK. 1996. Recency preference in the human sentence processing mechanism. *Cognition* 59:23–59.
GORDON, P. C., R. HENDRICK & M. JOHNSON. 2001. Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 27:1411–1423.
GRODNER, D., E. GIBSON & S. TUNSTALL. 2002. Syntactic complexity in ambiguity resolution. *Journal of Memory and Language* 46:267–295.
GRODNER, D., D. WATSON & E. GIBSON. 2000. Locality effects on sentence processing. Paper presented at the 13th CUNY Sentence Processing Conference, University of California, San Diego.

HAWKINS, J. A. 1994. *A performance theory of order and constituency*. Cambridge: Cambridge University Press.
JUST, M. A., P. A. CARPENTER & J. D. WOOLLEY. 1982. Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General* 3:228–238.
KAPLAN, R. & J. BRESNAN. 1982. Lexical-Functional Grammar: A formal system for grammatical representation. In *The mental representation of grammatical relations*, ed. J. Bresnan, 173–281. Cambridge, Mass.: MIT Press.
KING, J. & M. A. JUST. 1991. Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language* 30:580–602.
KLUENDER, R. & W. COWLES. 1997. How acceptable is grammatical? Processing factors and the competence grammar. Poster presented at the 10th CUNY Sentence Processing Conference, Santa Monica, Calif.
LEWIS, R. 1996. A theory of grammatical but unacceptable embeddings. *Journal of Psycholinguistic Research* 25:93–116.
MACDONALD, M. 1989. Priming effects from gaps to antecedents. *Language and Cognitive Processes* 4:35–56.
MACDONALD, M., N. PEARLMUTTER & M. SEIDENBERG. 1994. The lexical nature of syntactic ambiguity resolution. *Psychological Review* 101:676–703.
MCCLOSKEY, J. 2000. Quantifier float and *wh*-movement in an Irish English. *Linguistic Inquiry* 31:57–84.
MCKOON, G., R. RATCLIFF & D. ALBRITTON. 1996. Sentential context effects on lexical decisions with a cross-modal instead of an all-visual procedure. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22:1494–1497.
NICOL, J. L., J. D. FODOR & D. SWINNEY. 1994. Using cross-modal lexical decisions tasks to investigate sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 20:1229–1238.
NICOL, J. L. & D. SWINNEY. 1989. The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research* 18:5–20.
PEARLMUTTER, N. J. & E. GIBSON. 2001. Recency in verb phrase attachment. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 27:574–590.
PICKERING, M. & G. BARRY. 1991. Sentence processing without empty categories. *Language and Cognitive Processes* 6:229–259.
POLLARD, C. & I. A. SAG. 1994. *Head-driven phrase structure grammar*. Chicago: University of Chicago Press.
SAG, I. A. & J. D. FODOR. 1994. Extraction without traces. *Proceedings of the 13th West Coast Conference on Formal Linguistics*, ed. S. Preuss, R. Aranovich, W. Byrne & M. Senturia, 365–384. Stanford, Calif.: CSLI Publications. [Distributed by Cambridge University Press.]
STEEDMAN, M. 1996. *Surface structure and interpretation*. Cambridge, Mass.: MIT Press.
SWINNEY, D., M. FORD, U. FRAUENFELDER & J. BRESNAN. 1988. On the temporal course of gap-filling and antecedent assignment during sentence comprehension. Ms., University of California, San Diego.
TORREGO, E. 1984. On inversion in Spanish and some of its effects. *Linguistic Inquiry* 15:103–129.
TRAXLER, M. J. & M. J. PICKERING. 1996. Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language* 35:454–476.
TRUESWELL, J. C., M. K. TANENHAUS & S. M. GARNSEY. 1994. Semantic influences on parsing: Use of thematic role information in syntactic disambiguation. *Journal of Memory and Language* 33:285–318.

WANNER, E. & M. MARATSOS. 1978. An ATN approach in comprehension. In *Linguistic theory and psychological reality*, ed. M. Halle, J. Bresnan & G. Miller, 119–161. Cambridge, Mass.: MIT Press.
WARREN, T. & E. GIBSON. 2002. The influence of referential processing on sentence complexity. *Cognition* 85:79–112.
WARREN, T., E. GIBSON, M. JAMESON & C. HIRSCH. 2003. Effects of NP type in reading cleft sentences in English. Poster presented at AMLaP 2003, Glasgow, Scotland.
ZAENEN, A. 1983. On syntactic binding. *Linguistic Inquiry* 14:469–504.

Edward Gibson
NE20-439, MIT
77 Massachusetts Ave.
Cambridge, MA 02139
USA

egibson@mit.edu

Tessa Warren
University of Pittsburgh

tessa+@pitt.edu

Appendix A: Experimental Items

The experimental stimuli are presented here. All four conditions are presented for the first item. Only the two extracted conditions are presented for the remaining items. The nonextracted versions are formed from the extracted versions by changing the relative clause of the extracted version into the main clause of the nonextracted version, and by placing the main-clause subject from the extracted version into its object position in the nonextracted version. The verb phrase of the main clause in the extracted version is then added to the end of the nonextracted version, as a relative clause initiated by the relative pronoun *who*.

The mean plausibility ratings for the simplified nonextracted versions are presented following the extracted versions for all of the items. These ratings were obtained in the norming survey described in the experiment. The plausibility scale spanned from 1 (natural) to 7 (unnatural). The five items that were excluded from analyses because of plausibility differences are marked with an asterisk.

*1. +extract, VP: The politician who the journalist predicted that the government report would bother is calling a press conference. (3.15)
+extract, NP: The politician who the journalist's prediction about the government report had bothered is calling a press conference. (5.29)
–extract, VP: The journalist predicted that the government report

would bother the politician who is calling a press conference.
–extract, NP: The journalist's prediction about the government report had bothered the politician who is calling a press conference.

2. +extract, VP: The manager who the consultant claimed that the new proposal had pleased will hire five workers tomorrow. (2.36)
+extract, NP: The manager who the consultant's claim about the new proposal had pleased will hire five workers tomorrow. (2.79)
3. +extract, VP: The general who the advisor thought that the sergeant's message had angered was attempting to appear calm. (2.57)
+extract, NP: The general who the advisor's thoughts about the sergeant's message had angered was attempting to appear calm. (2.71)
*4. +extract, VP: The girl who the policeman concluded that the nasty threat had frightened has stopped going to school. (3.00)
+extract, NP: The girl who the policeman's conclusion about the nasty threat had frightened has stopped going to school. (5.21)
5. +extract, VP: The student who the teacher predicted that the new idea would inspire is studying artificial intelligence. (3.14)
+extract, NP: The student who the teacher's prediction about the new idea had inspired is studying artificial intelligence. (4.29)
6. +extract, VP: The actress who the agent implied that the controversial rumor distressed had lobbied to play Evita. (2.71)
+extract, NP: The actress who the agent's implication about the controversial rumor distressed had lobbied to play Evita. (1.93)
7. +extract, VP: The woman who the man confirmed that the loving promise had thrilled was hoping to get married. (3.00)
+extract, NP: The woman who the man's confirmation of the loving promise had thrilled was hoping to get married. (4.00)
8. +extract, VP: The freshman who the sophomore realized that the drunken promise had surprised was new to the partying scene. (3.29)
+extract, NP: The freshman who the sophomore's realization about the drunken promise had surprised was new to the partying scene. (3.64)
*9. +extract, VP: The mathematician who the genius thought that the elusive proof had fascinated has retired from the department. (3.93)
+extract, NP: The mathematician who the genius' thoughts about the elusive proof had fascinated has retired from the department. (2.29)
10. +extract, VP: The victim who the counselor concluded that the critical comment had annoyed has switched to another psychoanalyst. (3.29)
+extract, NP: The victim who the counselor's conclusion about the critical comment had annoyed has switched to another psychoanalyst. (4.14)
11. +extract, VP: The patient who the doctor predicted that the test results would reassure had been afraid she had cancer. (2.43)
+extract, NP: The patient who the doctor's prediction about the test results had reassured had been afraid she had cancer. (3.14)

12. +extract, VP: The judge who the reporter implied that the controversial decision had embarrassed has decided to sue the paper. (4.07)
    +extract, NP: The judge who the reporter's implication about the controversial decision had embarrassed has decided to sue the paper. (2.50)

*13. +extract, VP: The orphan who the social worker denied that the suggestion had pleased was unhappy with her foster parents. (2.14)
    +extract, NP: The orphan who the social worker's denial of the suggestion had pleased was unhappy with her foster parents. (3.21)

14. +extract, VP: The schizophrenic who the psychologist hypothesized that the new theory could help has stopped taking his pills. (4.00)
    +extract, NP: The schizophrenic who the psychologist's hypothesis about the new theory has helped has stopped taking his pills. (3.21)

*15. +extract, VP: The decision which the professor assumed that the dean's recommendation had influenced was eventually made by the review board. (4.50)
    +extract, NP: The decision which the professor's assumption about the dean's recommendation had influenced was eventually made by the review board. (3.14)

16. +extract, VP: The defendant who the witness confirmed that the expert's testimony had implicated would admit to nothing. (3.14)
    +extract, NP: The defendant who the witness's confirmation of the expert's testimony had implicated would admit to nothing. (3.43)

17. +extract, VP: The client who the lawyer claimed that the media's accusation had hurt was fired from his job. (4.00)
    +extract, NP: The client who the lawyer's claim about the media's accusation had hurt was fired from his job. (3.36)

18. +extract, VP: The journalist who the editor stated that the updated information had contradicted was planning a series of articles. (4.07)
    +extract, NP: The journalist who the editor's statement about the updated information had contradicted was planning a series of articles. (2.71)

19. +extract, VP: The senator who the committee stated that the confession might redeem will remain on the ethics panel. (3.43)
    +extract, NP: The senator who the committee's statement about the confession might redeem will remain on the ethics panel. (2.36)

20. +extract, VP: The daughter who the mother thought that the difficult decision had strengthened has left home for good. (2.43)
    +extract, NP: The daughter who the mother's thoughts about the difficult decision had strengthened has left home for good. (3.43)

Appendix B

Residual (and raw) reading times per word in each region, by participant, in milliseconds, after trimming reading times greater than 3000 ms of residual reading time (affecting less than 0.5% of the data).

Condition	Region						
	1	2	3	4	5 (critical)	6	7 (rest)
+extract, VP	–37.7 (371)	–12.0 (419)	126.1 (529)	23.3 (451)	63.6 (489)	89.7 (502)	39.8 (461)
+extract, NP	–21.2 (388)	12.9 (460)	84.5 (497)	42.5 (471)	138.5 (564)	130.3 (542)	36.7 (457)
–extract, VP		–37.4 (394)	–12.5 (391)	–33.3 (394)	7.7 (434)	15.1 (434)	–1.9 (410)
–extract, NP		–32.7 (415)	1.8 (413)	–21.3 (408)	22.3 (449)	2.7 (422)	15.9 (427)