Cognition 122 (2012) 12–36

Contents lists available at SciVerse ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

The processing of extraposed structures in English

Roger Levy [a,*], Evelina Fedorenko [b], Mara Breen [c], Edward Gibson [b]

[a] Department of Linguistics, UC San Diego, 9500 Gilman Drive #0108, La Jolla, CA 92093-0108, USA
[b] Department of Brain & Cognitive Sciences, Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, MA 02139, USA
[c] Department of Psychology, University of Massachusetts Amherst, Tobin Hall, 135 Hicks Way, Amherst, MA 01003, USA

ARTICLE INFO

Article history:
Received 8 December 2009
Revised 29 March 2011
Accepted 26 July 2011
Available online 27 October 2011

Keywords:
Sentence comprehension
Syntactic complexity
Parsing
Word order
Memory and language
Self-paced reading
Frequency
Prediction

ABSTRACT

In most languages, most of the syntactic dependency relations found in any given sentence are PROJECTIVE: the word–word dependencies in the sentence do not cross each other. Some syntactic dependency relations, however, are NON-PROJECTIVE: some of their word–word dependencies cross each other. Non-projective dependencies are both rarer and more computationally complex than projective dependencies; hence, it is of natural interest to investigate whether there are any processing costs specific to non-projective dependencies, and whether factors known to influence processing of projective dependencies also affect non-projective dependency processing. We report three self-paced reading studies, together with corpus and sentence completion studies, investigating the comprehension difficulty associated with the non-projective dependencies created by the extraposition of relative clauses in English. We find that extraposition over either verbs or prepositional phrases creates comprehension difficulty, and that this difficulty is consistent with probabilistic syntactic expectations estimated from corpora. Furthermore, we find that manipulating the expectation that a given noun will have a postmodifying relative clause can modulate and even neutralize the difficulty associated with extraposition. Our results rule out accounts based purely on derivational complexity and/or dependency locality in terms of linear positioning. Our results demonstrate that comprehenders maintain probabilistic syntactic expectations that persist beyond projective-dependency structures, and suggest that it may be possible to explain observed patterns of comprehension difficulty associated with extraposition entirely through probabilistic expectations.

© 2011 Published by Elsevier B.V.

1. Introduction

One of the central problems faced in the process of sentence comprehension is that the comprehender must infer HIERARCHICAL RELATIONS among the words of the sentence. For example, in the sentence

(1) Mary thought that John ate some toast with jam.

the comprehender must infer that *with jam* is dependent on (in this case, modifies) *toast*, which is part of the direct object of the verb *ate*, which in turn is the main verb of a sentential complement that is an argument of the verb *thought*. These hierarchical relations can be represented in terms of either constituent-structure trees or word–word dependency graphs (see Miller (2000) for a formal analysis demonstrating the intimate relationship between the two). Regardless of the formal apparatus with which these relationships are represented, they are a necessary part of computing sentence meaning.

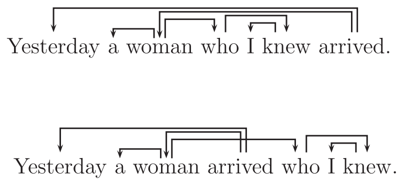
One of the striking regularities of natural-language syntax is that most such syntactic dependency relationships in most languages are PROJECTIVE. A set of word–word

* Corresponding author.
E-mail addresses: rlevy@ucsd.edu (R. Levy), evelina9@mit.edu (E. Fedorenko), mbreen@psych.umass.edu (M. Breen), egibson@mit.edu (T. Gibson).

0010-0277/$ - see front matter © 2011 Published by Elsevier B.V.
doi:10.1016/j.cognition.2011.07.012

dependency relationships is projective if no two dependencies in the set cross each other. The sentence in (2a), for example, has projective dependencies, illustrated by the dependency arrows drawn pointing to each dependent from its GOVERNOR in the style of Mel'cuk (1988; see also Hudson, 1984). The sentence in (2b), in contrast, is non-projective: the dependency between *Yesterday* and *arrived* crosses the dependency between *woman* and *who*. This non-projectivity arises from RIGHT-EXTRAPOSITION of the relative clause *who I knew* across the main verb of the sentence.

(2) a. b.



Formally, a crossing dependency is defined as follows. Let two words w_i, w_j be in a dependency relation with w_i preceding w_j, and two other words w_k, w_l be in another dependency relation with w_k preceding w_l. The two dependencies cross if the words are ordered in either of the two following ways:

$$w_i, w_k, w_j, w_l \quad \text{or} \quad w_k, w_i, w_l, w_j$$

In dependency graphs, the head word of a sentence is generally taken to be dependent on an invisible "root" word (assumed to be positioned either before the first or after the last word of the sentence), so that (2b) would be considered to have crossing dependencies even if *Yesterday* were omitted. In the languages for which it has been possible to quantify the frequency of non-projectivity, it has been clear that projective dependencies are far more common (Kruijff & Vasishth, 2003; Levy & Manning, 2004), and that non-projective depedencies are rarer than would otherwise be expected by chance (Ferrer i Cancho, 2006). Even for languages typically characterized as approaching complete word-order freedom (e.g., Walpiri; Austin & Bresnan, 1996; Hale, 1983; Simpson, 1983), the absence to date of quantitative corpus analysis means that it is not clear that non-projectivity is ever truly common in any language.

In phrase-structure terms, non-projectivity generally implies discontinuous constituency: some subset of the sentence constitutes a single phrase but is not a single continuous substring of the sentence. In (2), the phrase *a woman who I knew* is a continuous constituent in (2a) but a discontinuous constituent in (2b). The notion of discontinuity allows us to characterize the projectivity of individual dependencies: in our terminology, a dependency is non-projective if it causes to the DOMAIN of the governor—the set of words that the governor dominates in the syntactic tree—to be a discontinuous substring of the sentence (Nivre & Nilsson, 1999). Hence the sole non-projective dependency in (2b) is that of *who* on *woman* (since the domain of *arrived* is a continuous string). There are several formal means of representing non-projective dependency in phrase-structure trees; Fig. 1 illustrates several alternatives, including movement with traces as in Government and Binding (Chomsky, 1981); slash-passing as in Generalized Phrase Structure Grammar (Gazdar, Klein, Pullum, & Sag, 1985); discontinuous-constituency derivation trees as in Tree Adjoining Grammars (Joshi, Levy, & Takahashi, 1975), Head Grammars (Pollard, 1984), and Discontinuous Phrase Structure Grammar (Bunt, 1996); and base-generation of a right-extraposed constituent which is syntactically coindexed with the noun phrase it modifies (Culicover & Rochemont, 1990).

Parallel to their greater complexity of formal description, non-projective dependencies are also generally considered to pose greater computational challenges than projective dependencies. Efficient tabular parsing algorithms exist to exhaustively analyze any input sentence using projective-dependency or context-free phrase structure grammars in time cubic in the length of the sentence—denoted $O(n^3)$ in the algorithmic complexity literature (Cormen, Leiserson, Rivest, & Stein, 2001)—and quadratic in the size of the grammar (Earley, 1970; Younger, 1967).[1] In order to parse sentences using grammars allowing non-projective dependency or mildly context-sensitive phrase structure (which permits the representation of discontinuous constituency), more complex algorithms are necessary; for Tree-Adjoining Grammar (Joshi et al., 1975), for example, tabular parsing algorithms are $O(n^6)$ in sentence length (Nederhof, 1999; Vijay-Shanker & Joshi, 1985).[2,3]

Of the word–word dependency relationships that must be inferred during sentence comprehension, then, non-projective dependencies hold a place of special interest. This interest has to date been realized primarily within the theoretical syntax literature, but the rarity and computational complexity of non-projective dependencies makes them of considerable interest for sentence processing as well. This paper thus reports an investigation into the online processing of non-projective dependencies in the context of prominent contemporary theories of syntactic comprehension.

1.1. Online comprehension of non-projective structure

Perhaps the best-known previous study investigating whether non-projectivivity poses any special burden in human sentence comprehension is Bach, Brown, and Marslen-Wilson (1986)'s comparative study of crossing versus

[1] The parsing problem can in principle be recast as a problem of matrix multiplication, which permits sub-cubic asymptotic time complexity (Valiant, 1975), but this approach leads in practice to much slower parsing times for sentences of the lengths actually observed in natural language.

[2] Some types of non-projectivity can be represented directly in a context-free phrase-structure grammar, in the style of Generalized Phrase Structure Grammar (Gazdar et al., 1985), but this approach has the effect of radically increasing the size of the grammar and hence increasing the computational work associated with parsing.

[3] McDonald, Pereira, Ribarov, and Hajič (2005) introduced a minimum spanning tree algorithm for non-projective dependency parsing that is quadratic in sentence length. However, this algorithm does not allow *any* constraints on the allowable classes of non-projectivity, such that (for example) only limited cross-serial dependencies are admitted (Joshi, 1985). The consensus within mathematical linguistics, in contrast, is that natural languages only allow certain restricted types of non-projectivity, those characterized by the *mildly context-sensitive* class of formal languages (Culy, 1985; Joshi, Shanker, & Weir, 1991; Shieber, 1985)—those characterized by Tree-Adjoining Grammar (Joshi et al., 1975), Head Grammar (Pollard, 1984), Combinatory Categorial Grammar (Steedman, 2000)—though see Kobele (2006) and Kuhlmann and Nivre (2006) for more recent developments and refinements.

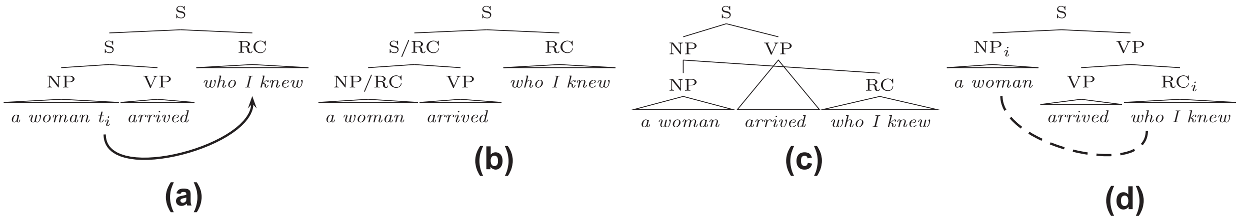


Fig. 1. Different phrase-structure representations for the non-projective dependency structure in (2b). In (a), the non-projective dependency is characterized by movement from an NP-internal trace. In (b), the non-projective dependency is represented by the missing-RC information transmitted between the NP and the top S categories. In (c), non-projectivity is directly represented as a discontinuous constituent. In (d), non-projectivity is represented as a special syntactic coindexing relation between a base-generated RC and the NP it modifies.

nested dependencies in Dutch versus German, which found that sentences with multiple levels of embedding were easier for Dutch speakers as crossing dependencies than for German speakers as nested dependencies. In addition to being a comparison across languages, this task did not use online measures of processing difficulty. For this reason, the results do not definitively address issues of processing cost.[4] More recently, considerable attention has been paid to the computation of unbounded filler-gap dependencies (e.g., Boland, Tanenhaus, Garnsey, & Carlson, 1995; Traxler & Pickering, 1996), which do involve non-projective dependency structures, but attention in these studies has focused primarily on how and when the gap site is inferred, and have not involved direct contrasts with corresponding projective structures. Turning to right-extraposition, Francis (2010) studied English relative clause (RC) extraposition and grammatical weight effects, finding an interaction between the two such that longer RCs favored extraposition; however, this study used whole-sentence reading times and thus does not localize processing difficulty. Finally, Konieczny (2000) studied German RC extraposition and found that reading times on relative pronouns in extraposed RCs were greater than on relative pronouns in in situ RCs.[5] This existing work thus leaves two important questions regarding the processing of non-projective dependency structures unaddressed:

1. Under what circumstances are non-projective dependency structures easier or harder to comprehend than corresponding projective-dependency structures?
2. How can these differences in comprehension difficulty be understood with respect to existing theories of online comprehension?

In this paper we attempt to address these questions through the study of right extraposition of English RCs. RC extraposition turns out to have several advantages from the perspective of studying crossing-dependency processing. Unlike the situation with many filler-gap dependencies in English, RC extraposition maintains the same order of dependent and governor (*who* and *woman* respectively in (2)) as the in situ case, facilitating direct comparisons of online processing difficulty. Although it is an uncommon structure, RC extraposition is by no means unheard of, and as will be seen in the experiments reported in this paper, native speakers are perfectly able to comprehend sentences involving the simpler varieties of extraposition. Finally, right-extraposition is a widespread phenomenon cross-linguistically, so that results obtained for English may be compared relatively directly to future studies in other languages, which may be of considerable theoretical interest. To set the stage for our answers to the above questions, the next section outlines the predictions made by existing theories for the processing of extraposed relative clauses.

1.2. Right-extraposition: Predictions of existing theories

Broadly speaking, theories of syntactic comprehension have developed along two lines of inquiry: the problem of AMBIGUITY RESOLUTION and the problem of COMPLEXITY in (potentially unambiguous) structures. Theories of ambiguity resolution include the Sausage Machine and its descendant garden-path theories (Clifton & Frazier, 1989; Frazier & Fodor, 1978; Frazier, 1987), the Tuning Hypothesis (Mitchell, 1994; Mitchell, Cuetos, Corley, & Brysbaert, 1995), the constraint-based competition-integration model (Spivey-Knowlton, 1996; Spivey & Tanenhaus, 1998; McRae, Spivey-Knowlton, & Tanenhaus, 1998), and pruning/attention-shift models (Crocker & Brants, 2000; Jurafsky, 1996; Narayanan & Jurafsky, 1998; Narayanan & Jurafsky, 2002). Because right-extraposition does not necessarily involve any local structural ambiguity, if locally unambiguous cases such as (2b) could be shown to induce processing difficulty, then theories that exclusively cover ambiguity resolution would not be sufficient to capture constraints involved in non-projective dependency constructions. The results of our first experiment indicate that this is indeed the case, hence we will not discuss ambiguity-resolution theories in any further detail.

There are several theories of processing complexity—as well as theories that attempt to explain both ambiguity-resolution and unambiguous-sentence complexity (e.g.,

[4] Kaan and Vasić (2004) presented a self-paced reading study of Dutch cross-serial dependency processing suggesting that sentences with three cross-serial verbs are more difficult to comprehend than similar sentences with only two cross-serial verbs. However, this pattern of difficulty emerged—and, indeed, was strongest—among the NP arguments of the sentence, before the verbs were encountered and any cross-serial dependencies completed. Hence it is not clear that these results bear directly on the question of crossing-dependency processing complexity.

[5] Since acceptance of this paper, we have also learned of unpublished data by Konieczny and Bormann (presented as Konieczny and Bormann, 2001) which in some ways anticipated our Experiment 3, but whose results were inconclusive.

Gibson, 1991)—that make clear predictions regarding the processing of extraposed structures, based on differing principles. We cover each of these theories, and their basic predictions in some detail here, and revisit each of these theories later in the paper.

1.2.1. Derivational complexity/inherent difficulty of non-projective dependencies

The Derivational Theory of Complexity (DTC) was among the first applications of syntactic theory to psycholinguistics. Its origins lie in a hypothesis articulated by Miller (1962) that the complete comprehension of a sentence by a speaker involves detransforming the sentence into a "kernel" form, together with annotations indicating the transformations relating the kernel to the surface (perceived) version. Although the DTC has since fallen out of favor (Slobin, 1966; see also Fodor, Bever, & Garrett, 1974; Bever, 1988; Tanenhaus & Trueswell, 1995), it makes a simple prediction about extraposition on the assumption that the in situ variety is the "kernel" form—that is, that extraposed RCs have the structure as in Fig. 1a: namely, that right-extraposed RCs (e.g., (2b)) will be more difficult to process than their corresponding unextraposed variants (e.g., (2a)). Alternatively, such a pattern of results could be interpreted in slightly different, more limited terms: that all sentences with non-projective dependency structures—rather than all non-kernel sentences, as in the original DTC—are simply more costly in comprehension than projective dependencies. Within the scope of this paper, such an interpretation would be indistinguishable from the more general DTC; hence we lump coverage of the two together in the remainder of the paper, though of course the two interpretations would have different consequences for comprehension of other syntactic constructions.

1.2.2. Decay and/or interference in memory retrieval

Two prominent theories posit that in the online construction of word–word syntactic dependencies, the retrieval of the earlier element in a dependency is a comprehension bottleneck: the Dependency Locality Theory (DLT; Gibson, 1998, 2000; Grodner & Gibson, 2005) and Similarity-Based Interference (SBI; Gordon, Hendrick, & Johnson, 2001, 2004; Lewis & Vasishth, 2005; Lewis, Vasishth, & Van Dyke, 2006). According to these theories, greater linear distance between a governor and its dependent can give rise to greater processing difficulty for at least one of two reasons: (1) the activation level of the earlier item has DECAYED, making it harder to retrieve; (2) the material that intervenes between the two items may possess features sought by the retrieval cue (the later item) and thus INTERFERE with retrieval of the target (the earlier item). On the assumption that a retrieval of the governing noun is necessary to construct the syntactic relationship between the governing noun and the extraposed RC, these theories predict that an extraposed RC will be more difficult to process than an in situ, adjacent RC. However, these theories as constructed thus far do not distinguish an extraposed RC from an RC that is in a projective dependency relationship with the head noun, but is not linearly adjacent to it; this point will become prominent in Experiment 2.

1.2.3. Probabilistic expectations

There are several theories that predict differential difficulty in unambiguous contexts on the basis of probabilistic expectations, including surprisal (Boston, Hale, Kliegl, Patil, & Vasishth, 2008; Demberg & Keller, 2008; Hale, 2001; Levy, 2008; Roark, Bachrach, Cardenas, & Pallier, 2009; Smith & Levy, 2008; see also MacDonald & Christiansen, 2002), entropy reduction (Hale, 2003, 2006), and the top-down/bottom-up model of Gibson (2006). We will take surprisal as an exemplar for the type of predictions made by this class of models. Under surprisal, the difficulty of a word w in a sentence is determined by the log of the inverse conditional probability of the context in which it appears: $\log \frac{1}{P(w|\text{context})}$. Depending on how comprehenders formulate probabilistic expectations for upcoming events in a sentence, these conditional probabilities may reflect various structural features of the earlier part of the sentence. Here, we entertain two possible types of probabilistic expectations:

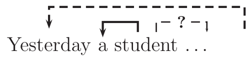
- *Collocational expectations.* Comprehenders may attend to a limited number of words immediately preceding a given word in order to formulate expectations. As an extreme case, in (2), the probability of the relative pronoun *who* when it appears might be conditioned on only the previous word or the previous word class. On the assumption that expectations are set RATIONALLY (Anderson, 1990; Marr, 1982; Shepard, 1987) on the basis of experience, we can estimate comprehenders' likely expectations through corpora. For example, in the Brown corpus (Kučera & Francis, 1967), the word *woman* appears 194 times and is followed by *who* 17 times; the word *arrived* appears 56 times and is never followed by *who*. Thus, the collocational surprisal of the RC onset is almost certainly higher in (2b) than in (2a). Alternatively, we could use the parsed version of the Brown corpus (Marcus, Santorini, & Marcinkiewicz, 1994), which has syntactic-category annotations, to estimate the probability of *who* conditioned on the fact that the previous word is a singular noun, as in *woman*, or a past-tense verb, as in *arrived*. Both these types of collocational probabilities are given in Table 1. Collocational expectations clearly predict a cost for RC extraposition, at least for examples of the type given in (2).
- *Structural expectations.* Alternatively, syntactic comprehension may be facilitated by expectations for upcoming words based on rich information that includes possible structural analyses of earlier parts of the sentence, as well as parts of the sentence that have not yet been seen. For example, in the unextraposed RC of

Table 1
Collocational and syntactic conditional probabilities of the relative clauses in Example (2). w_i denotes the word introducing the RC (*who* in these cases); c_{i-1} denotes the syntactic part of speech of the previous word. Probabilities are relative-frequency estimates from the parsed Brown corpus.

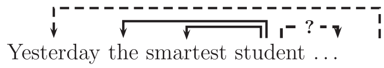
	$P(w_i\|w_{i-1})$	$P(w_i\|c_{i-1})$	$P(RC\|\text{syntactic context})$
(2a)	0.07	0.0044	0.00561
(2b)	0	0.00025	0.00004

(2a), the expectation for *who* should be identified with the probability that *woman* will be immediately postmodified by an RC, since the only way that *who* can appear at this point in the sentence is as the introduction of an RC. The probability *P*(RC|context), in turn, should in principle be conditioned not only on the fact that *woman* is the previous word, but on a variety of properties such as the fact that this word is the head of an indefinite noun phrase that is the subject of the sentence. In practice, not enough syntactically annotated corpus data are available to estimate the relevant probabilities conditioned on all these elements of the context, but we may approximate these probabilities by conditioning on a smaller number of structural properties. If, for example, we ignore the definiteness and specific head word and condition only on the status of the NP as a subject, then we find that an unextraposed relative clause immediately following the head of the subject NP is far more probable than a relative clause from the subject following the head of the main VP of the sentence, as estimated from the parsed Brown corpus and shown in Table 1 (tree-search patterns used to obtain these figures are given in Appendix A). It seems, then, that structural expectations predict a processing penalty for extraposition in examples such as (2), just as collocational expectations. In both cases, however, probabilistic theories predict that the penalty for extraposition may be modulated by the effects of the specific context on the predictability of a relative clause—either extraposed or not—in the position in which it appears.

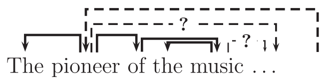
As seen in Table 1, expectation-based theories predict that non-projective dependencies—at least the extraposed-RC variety—are likely to be difficult to process in general due to their rarity. We will show, however, in Experiment 3 that under some circumstances the conditional probability of an extraposed RC in its context can be made quite high. These circumstances are of particular interest for theories of online sentence comprehension. On the one hand, one might expect that strong expectations for an extraposed RC in such contexts will eliminate any general processing difficulty observed with RC extraposition in other contexts. On the other hand, the formal complexity of non-projective dependency structures might make it difficult for comprehenders to deploy expectations effectively when processing non-projective dependencies. To make this argument clear, we spend this and the following two paragraphs laying out the types of situations in which a comprehender might have a strong expectation for a non-projective dependency structure. We begin with the relatively simple situation seen in Example (3), where *student* may yet be postmodified by an RC. The solid line between *a* and *student* indicates the already-constructed dependency between the subject's determiner and head noun; the thick dashed line pointing to *Yesterday* indicates the known dependency that will occur in the future between this adverbial and the main verb of the sentence (Gibson, 1998, 2000); and the thin dashed line with a question mark originating from *student* indicates the (probabilistic) expectation for this possible postmodification.

(3)


The strength of this expectation for a postmodifier will depend on the properties of the subject NP. For example, Wasow, Jaeger, and Orr (2006) and Jaeger (2006) demonstrated that definite-superlative NPs are more likely to have postmodifying RCs than indefinite NPs without adjectives. In Example (4), the definite-superlative content would thus create a stronger expectation for an upcoming RC, denoted by the question-marked dotted line (thicker than the corresponding line in (3)) originating from *student*.

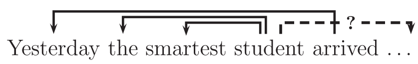
(4)


On an expectation-based theory, an immediately following RC would thus be easier to process in (4) than in (3). Contrasting expectation effects of this type have been shown especially clearly in RC attachment preferences. For example, Desmet, Brysbaert, and de Baecke (2002) showed that in Dutch, human NPs are more frequently postmodified by RCs than non-human NPs are. A complex NP onset, such as *the pioneer of the music*... in Example (5), would thus have differing strengths of expectation for RC postmodification of the two preceding nouns.

(5)


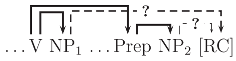
On expectation-based syntactic-comprehension theories, the relatively stronger expectation for attachment to the first noun (*pioneer*) predicts a processing benefit accrued for comprehension of RCs attaching to this noun, an effect documented by Mitchell and Brysbaert (1998).

Now let us consider a case similar to Example (4), but in which the main verb has also been seen, as in Example (6).

(6)


The prenominal content of the subject NP has created an expectation for a postmodifier, but any such postmodifier would now have to be extraposed, and its dependency on the subject NP would thus cross the verb → adverb dependency. It is possible that maintaining this strong expectation across the verb → adverb dependency in online comprehension is difficult or impossible due to the crossing of dependencies and ensuing non-projectivity. If this were true, then in the cases where *arrived* was indeed followed by an RC, the comprehender would fail to benefit from what should be a strong extraposed-RC expectation. Put another way, when an RC appears that is unexpected, it would be hard to process whether it was extraposed or not, because the comprehender (sensibly) failed to anticipate it; but when an RC appears that should be expected, it would be easy to process only if it were unextraposed, and thus the comprehender would fail to reap the appropriate benefits of expectation when the RC is extraposed.

It is thus of considerable interest to determine whether comprehenders can deploy strong syntactic expectations to facilitate processing even when non-projective dependencies are involved, since an inability to do so could provide one possible explanation for the comparative rarity of extraposition. A particularly strong test of this ability could be made by setting up a situation as in (7), in which some property of a post-verbal NP_1 sets up a strong expectation for a postmodifier that remains unmet during the processing of an ensuing verb-modifying prepositional phrase:

(7) 

In this situation, expectation-based theories predict that if this expectation is sufficiently strong, an RC appearing immediately after NP_2 may be as easy or easier to process if it modifies NP_1 than if it modifies NP_2, but only if expectations for a non-projective dependency can be effectively deployed across an existing dependency. We return to this prediction in Experiment 3.

1.3. Summary

Several leading theories of online syntactic comprehension predict comprehension costs for extraposed-RC structures in English, but the details of these predictions are different for each theory, as are the implications for larger questions regarding the distribution of non-projectivity in natural language. Investigating extraposed RCs in English thus poses an opportunity to refine our understanding of online syntactic comprehension, and at the same time may contribute part of an answer to questions of a more typological nature. The remainder of this paper presents three experiments bearing on these issues. Experiment 1 establishes the presence of a processing cost for the most common type of extraposed RCs in comparison with their non-extraposed alternates, but does not distinguish between the theories outlined above. Experiment 2 provides evidence distinguishing between predictions of derivational complexity and structural expectations on the one hand, versus those of locality/interference and collocational expectations on the other hand. Experiment 3 provides evidence distinguishing between predictions of structural expectations versus those of derivational complexity. We revisit the implications of these experimental results for our original questions in Section 5.

2. Experiment 1

In this study we compared online comprehension of subject-modifying English relative clauses in situ (that is, immediately following the modified noun) and relative clauses that are right-extraposed over an intransitive verb. In this contrast, as illustrated in Example (2), there is no local structural ambiguity as to the correct interpretation of the relative clause in either case, hence the contrast may reveal processing costs associated with extraposition independent of structural ambiguity resolution. We also tested the probabilistic-expectation account by varying the semantic class of the main-clause verb. In particular, it has been claimed that extraposition across an intransitive verb is facilitated when the verb comes from the class of presentative or presentational verbs (Givón, 1993; Aissen, 1975), such as *arrived*, *appeared* and *showed up*, in which the subject of the verb is being introduced into a scenario. If this is true, then seeing a presentative main verb might increase the comprehender's expectation for an RC extraposed from the subject NP to appear after the main verb. Hence we tested both presentative verbs and non-presentative verbs such as *performed*, *died* and *lied*. This results in a 2 × 2 factorial design as in Example (8):

(8) a. After the show, a performer who had really impressed the audience came on and everyone went wild with applause [presentative, RC in situ].
b. After the show, a performer came on who had really impressed the audience and everyone went wild with applause [presentative, RC extraposed].
c. After the show, a performer who had really impressed the audience bowed and everyone went wild with applause [non-presentative, RC in situ].
d. After the show, a performer bowed who had really impressed the audience and everyone went wild with applause [non-presentative, RC extraposed].

If RC extraposition leads to a greater processing cost, then we should see greater difficulty in Examples (8b) and (8d) than in (8a) and (8c). This pattern may show up as a main effect; or, if extraposition is easier to process across presentative verbs, then there should be an interaction between extraposition and verb type such that the reading time difference between extraposed and non-extraposed versions is smaller for the presentative versions of the items than for the non-presentative versions.

2.1. Participants

Forty participants from MIT and the surrounding community were paid for their participation. All were native speakers of English and were naive as to the purposes of the study.

2.2. Materials

Twenty-four items (listed in full in Appendix B) were constructed following the pattern of (8). Each item was initiated by a prepositional phrase that established a context for the sentence (e.g., *After the show* in (8)). The subject NP of the sentence occurred next, consisting of an indefinite determiner (*a/an*) and an occupation noun; *performer* in (8). In the non-extraposed versions, the relative clause occurred next, consisting of five or six words—often but not always a passive verb plus a prepositional phrase. The main verb phrase in the sentence occurred next (*came on* in (8c)), followed by a conjunction such as *and* or *but*, and finally five or six words making up a second clause.[6] The extraposed versions of each item were formed by shifting the relative clause past the main verb of the sentence. The non-presentative conditions were identical to the presentative conditions, except that the presentative verb was replaced with a non-presentative verb (*came on* vs. *bowed* in (8)). In addition to the target sentences, 96 filler sentences with various syntactic structures were included, including sentence materials from two other experiments. Each participant saw only one of the four versions of each item, according to a Latin-square design. The stimuli were pseudo-randomized separately for each participant, so that a target sentence never immediately followed another target sentence.

2.2.1. Verb presentativity: corpus study

If expectations based on linguistic experience are a determinant of extraposed RC processing difficulty, and if presentative verbs facilitate RC extraposition, then we might expect to see differences in the relative frequencies of extraposed RCs in corpora as a function of verb type for our materials. It turned out to be rather difficult to quantify the differences between presentative and non-presentative cases, however, because RC extraposition is rare enough that reliable frequency estimates were impossible to obtain using hand-parsed corpora. We therefore resorted to the largest publicly available corpus, the Google *n*-grams corpus, a compilation of the most frequent *n*-grams for $n \leqslant 5$ based on one trillion words of Web data (Brants & Franz, 2006). We used the word *who* as a proxy for detecting RC onsets, and for each of our items obtained relative-frequency estimates of the conditional probabilities $P(\text{who}|\text{VP})$ for the presentative and non-presentative variants of the VP. Averaging across items, this probability was $2 \times 10^{-4}(\pm 5 \times 10^{-5})$ for our presentative condition, and $4 \times 10^{-5}(\pm 2 \times 10^{-5})$ for our non-presentative condition. A two-sample non-parametric permutation test (Good, 2004) indicated that this probability was higher for the presentative condition than for the non-presentative at $p < 0.01$.

2.3. Procedure

Sentences were presented to participants in a non-cumulative moving-window self-paced procedure on a Mac or a PC computer running the Linger software (Rohde, 2005). Each trial began with a series of dashes displayed on the computer screen in place of the words in the sentence. The first press of the space bar revealed the first region in the sentence, and each subsequent press of the space bar revealed the next word in the sentence and masked the previous word. The sentence-initial adjunct (*After the show*), and most multi-word verb phrases such as *came on*, were presented in a single group, in order to avoid misinterpretation; otherwise, each word was presented individually as its own single-word group. Due to a programming error, four items (11, 13, 16, and 24) with two-word verb phrases in the non-presentative conditions were presented as two separate single-word groups; the analyses presented in Section 2.4 include measurements from only the first of these two groups.[7] The times between button presses were recorded to the nearest millisecond. Each sentence was followed by a yes-or-no comprehension question probing the participant's understanding of the content of the sentence. The study took an average of 40 min per participant to complete.

Table 2
Question–answering accuracy in Experiment 1.

	Extraposed	Unextraposed
Non-presentative	0.89	0.88
Presentative	0.93	0.93

Table 3
F-statistics for analysis of question–answering accuracy in Experiment 1.

	*F*1	*F*2
Presentative	5.80*	1.59
Extraposition	<1	<1
Pres × extrap	<1	<1

* $p < 0.05$

2.4. Results

2.4.1. Statistical analysis

Raw reading times were analyzed in each region as follows, unless otherwise specified. Measurements above 1500 ms were discarded; means and standard deviations were then computed in each condition, and any measurement more than three standard deviations above or below the mean was discarded. These procedures resulted in total loss of 0.94% of data to be analyzed. The remaining measurements were then analyzed in 2×2 by-participants and by-items ANOVAs.[8] In cases where a single region of

[6] It should be noted that in many of the items, the conjunction initiating the following clause was *and*, which could initially ambiguously attach to the preceding clause (the eventual interpretation), or to the preceding object NP as a conjoined NP (e.g., *who had really impressed the audience* in (8d)). This ambiguity was not present in the non-extraposed conditions. Consequently, any reading time differences in this region between extraposed and non-extraposed conditions would be difficult to interpret for these items.

[7] When these items are excluded from analysis altogether, the qualitative patterns of question answering accuracy and reading times are the same, and the crucial main effect of extraposition is generally more highly significant than when all items are included in analysis. Because excluding these four items leads to imbalance in the experimental lists, however, we present analyses using all items.

[8] Some readers may wonder why we do not use mixed-effects models (Baayen, Davidson, & Bates, 2008) for these analyses. In our view, the disadvantages of these analyses currently outweigh the advantages for traditional, balanced datasets such as ours: the issue of what random-effects structure to use to test for fixed effects has not been fully resolved; our by-subjects and by-items means are never missing cells and thus our data are fully balanced; the averaging process in computing these means leads to response variables more normally distributed than raw RTs are; the by-subjects and by-items proportions in our question–answering accuracy data are far from 0 and 1; and the ability of mixed-effects analysis to accommodate correlated control predictors is of minimal value for a factorial design with fully balanced data such as ours. We do, however, use mixed-effects models in Experiment 3 to analyze categorical data that are both severely imbalanced and where the proportions vary dramatically across condition (reaching zero in one case).

analysis constituted more than one group of words presented (e.g., words five and above of the RC), reading times for the trial were averaged across all groups in the region. Error bars in graphs are standard errors of by-subject means. In-text description of significant reading-time results is restricted to regions inside the relative clause, but for completeness, *F*-statistics and significance levels are given in tables for all regions.

2.4.2. Comprehension accuracy

Overall question–answering accuracy on experimental items was 91%. Tables 2 and 3 show question–answering accuracy by condition, together with the results of 2 × 2 ANOVAs. Accuracy was high across conditions. In ANOVAs, the only significant effect was a main effect of verb type, significant by participants but not by items.

2.4.3. Reading times

We divided the sentence into nine regions of analysis as depicted in (9):

(9) After the show | a performer | ({came on/bowed}) | who | had | really | impressed | the audience | ({came on/bowed}) | . . .

Fig. 2 shows average reading times for each region of analysis, and Table 4 gives the results of 2 × 2 ANOVAs by participants and by items for each region. At the main verb, we see a numerical pattern toward reading times being shorter in the non-presentative conditions; this main effect of verb type is marginal by participants and insignificant by items. This difference is plausibly due to the use of different lexical items across the two verb-type conditions. Across the first three words of the RC (Regions 4–6), we find a pattern toward reading times being longer in the extraposed condition; this main effect of extraposition is significant by participants in Region 4, significant by participants and marginal by items in Region 5, and marginal by items in Region 6. There is also a hint of an interaction between verb type and extraposition such that extraposed RCs are read more quickly in presentative than in non-presentative conditions; this pattern is marginal by participants in Region 8.

Since there is considerable variability across items in the content of the relative clauses (see Appendix B), we conducted a residual reading-time analysis across the first four regions of RC (Regions 4–7). Following Ferreira and Clifton (1986), we first computed a single linear regression of RT against region length (as measured in characters) for each participant, including as data all reading times measured for all the participant's non-practice trials, and then summed both the residual and raw RTs separately across these four regions. We discarded trials for which the summed raw RTs were above 6000 ms, and analyzed

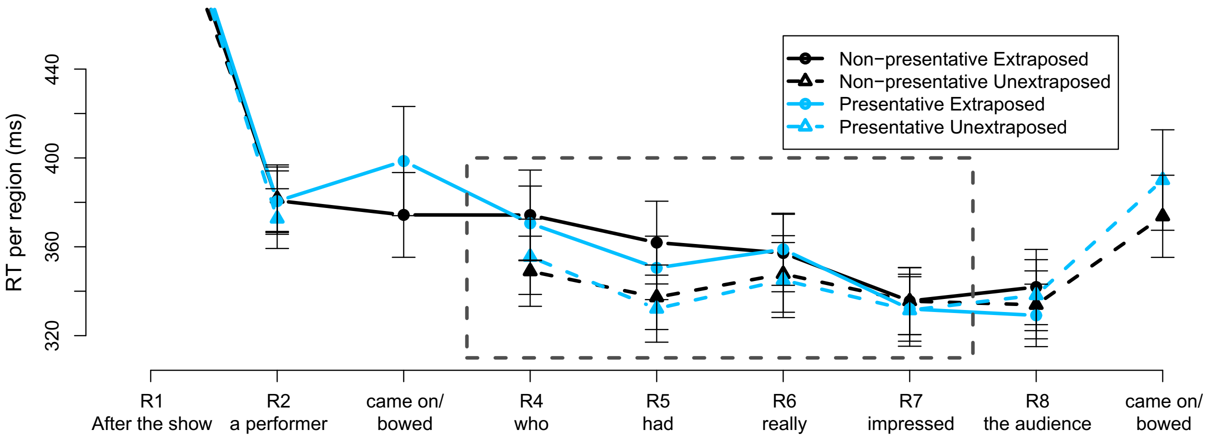


Fig. 2. Reading-time results as a function of region and condition for Experiment 1. Onset of the relative clause (first four words) is boxed.

Table 4
F-statistics for Experiment 1. In regions not listed, there were no effects were below $p = 0.1$.

	R1		R2		R4		R5		R6		R7		R8		MainVerb	
	F1	*F2*	*F1*	*F2*	*F1*	*F2*	*F1*	*F2*	*F1*	*F2*	*F1*	*F2*	*F1*	*F2*	*F1*	*F2*
Presentative	1.37	<1	<1	<1	<1	<1	3.02*	<1	<1	<1	<1	<1	<1	<1	3. 15*	1.09
Extraposition	<1	<1	<1	<1	4.27**	2.42	6.10**	3. 79*	2.32	3.30*	<1	<1	<1	<1	<1	<1
Pres × extrap	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	3.04*	1.64	<1	<1

* $p < 0.1$.
** $p < 0.05$.

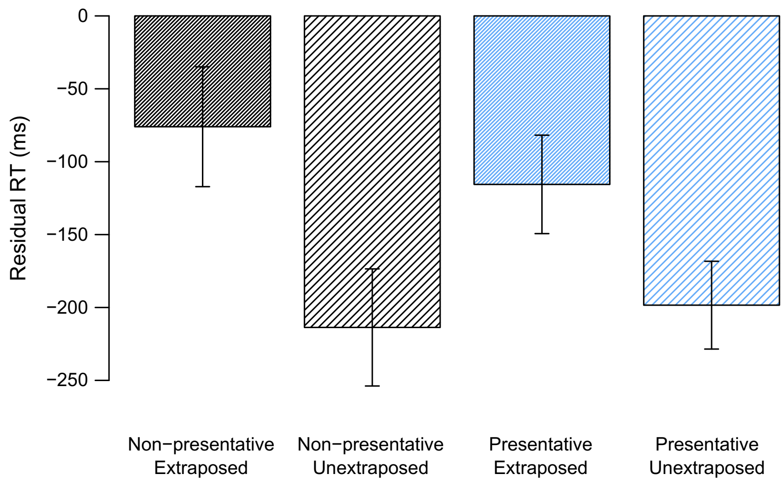


Fig. 3. Residual reading times over the first four words of the RC in Experiment 1.

Table 5
F-statistics for residual reading-time analysis across first four words of the RC in Experiment 1.

	*F*1	*F*2
Presentative	<1	<1
Extraposition	7.29*	6.23*
Pres × extrap	<1	<1

* $p < 0.05$.

residual RTs on the remaining trials as described in Section 2.4.1 (outlier removal procedures resulted in 1.1% total data loss in this analysis). Fig. 3 shows the average summed residual RT as a function of condition, and Table 5 reports results of a 2 × 2 ANOVA on these data. We see a highly significant main effect of RC extraposition in both participants and items analyses. Although there is a slight numerical trend toward an interaction such that extraposition is easier in the presentative condition than in the non-presentative condition, this interaction was far from significant.

2.5. Discussion

The results of Experiment 1 demonstrate that there is online comprehension difficulty associated with comprehending relative clauses extraposed from subject NPs across intransitive VPs. This effect of extraposition was most apparent at the beginning of the extraposed RC; by the fourth word of the RC, processing times were numerically near-identical across conditions. This experiment does not distinguish between derivational-complexity, decay/interference, or collocational/structural expectation-based theories as possible bases for comprehension difficulty in RC extraposition, although there is some weak circumstantial evidence against decay- and interference-based theories deriving from the lack of main-verb reading time sensitivity to RC extraposition. If memory decay or retrieval interference were important factors in determining reading times in this sentence, we might expect to see greater reading times at the main-clause verb when it is separated from the subject by an in situ RC (see also Jaeger, Fedorenko, Hofmeister, & Gibson (2008), who document several experiments in which manipulating the size of a subject-modifying RC has no effect on main-clause verb reading times). The results regarding whether verb type affects extraposition difficulty were inconclusive. The interaction predicted by Givon and Aissen's hypothesis—that a relative clause extraposed across a presentative verb should be easier to comprehend than one extraposed across a non-presentative verb—was not significant. However, the numerical pattern across the first four words of the RC (Fig. 3) was consistent with the hypothesis, and the lack of significance could derive from the weakness of the expectation manipulation and a corresponding lack of statistical power. A reviewer also correctly points out that multi-word VPs (mostly verb–particle combinations like *come on*) were more frequent in the presentative than in the non-presentative conditions, so any additional difficulty incurred from extraposition over the additional word (which both decay/interference and complexity-based theories might predict) could vitiate expectation-derived processing benefits.

Having demonstrated that RC extraposition can induce processing difficulty, we now turn to addressing the possible reasons for this difficulty. Since the presence of extraposition in Experiment 1 was conflated with linear distance of the RC from its attachment site, in the next two experiments we shift to a design that disentangles these two factors. Furthermore, in the remaining experiments in this paper we keep constant the word sequence immediately preceding the RC such that an experience-driven hypothesis driven purely by word collocations would be hard pressed to explain extraposition difficulty, simply because the collocations in question are too rare.

3. Experiment 2

This experiment was designed to distinguish between derivational complexity or structural expectations on the one hand versus decay, interference, or collocational expectations on the other. We achieved this by holding constant the words preceding the RC, and by manipulating whether the RC is extraposed independently of its linear distance from the noun it modifies. Sentences in this experiment involved an RC-preceding context of the form

(10) Verb Det_1 $Noun_1$ Preposition Det_2 $Noun_2$

where all these elements except for the preposition are held constant. We crossed the attachment site of the prepositional phrase (PP)—to NP_1 (NP-attachment) or to the verb (VP-attachment)—with the linear adjacency of the RC attachment—to NP_2 (adjacent) versus NP_1 (non-adjacent). The RCs were disambiguated with a set of redundant animacy- and plausibility-based cues starting at the relative pronoun (*who* can modify only animate NPs) and continuing throughout the RC. The four resulting conditions are illustrated in (11):

(11) a. The reporter interviewed the star of the movie which was filmed in the jungles of Vietnam [NP-attached PP, RC adjacent].
b. The reporter interviewed the star of the movie who was married to the famous model [NP-attached PP, RC non-adjacent].
c. The reporter interviewed the star about the movie which was filmed in the jungles of Vietnam [VP-attached PP, RC adjacent].
d. The reporter interviewed the star about the movie who was married to the famous model [VP-attached PP, RC non-adjacent].

Only the VP-attached PP, RC non-adjacent condition in (11d) involves an extraposed RC. If processing RC extraposition in comprehension involves a cost above and beyond the additive costs of PP attachment and the adjacency of the RC to the element it modifies, then we should see an interaction between these two factors in reading times within the RC itself, with super-additively high reading times in the extraposed-RC condition. Such an interactive pattern would be predicted by theories of derivational complexity in which extraposed modification (non-projective dependency) incurs a fundamentally greater processing cost than unextraposed modification, or by expectation-based theories in which structural frequencies play a role. Table 6 lists conditional probabilities of adjacent and non-adjacent RCs in the four conditions depicted in Example (11), based on relative-frequency estimation using the parsed Brown corpus (tree-matching patterns are given in Appendix A): extraposed RCs are far less expected than the other three types. Such an interactive pattern would not be predicted by theories in which the decisive factors are decay and/or interference based purely on linear distance. We return to collocational frequencies in Section 3.5.

Table 6
Syntactic conditional probabilities of RCs for the four conditions of Experiment 2. Note that the parsed Brown corpus contains 459,148 words.

Condition	RC conditional probability (%)	*n*
NP-attached PP, RC adjacent	4.07	2603
NP-attached PP, RC non-adjacent	4.07	2603
VP-attached PP, RC adjacent	3.25	13276
VP-attached PP, RC non-adjacent	0.14	13276

3.1. Participants

Forty-four participants from MIT and the surrounding community were paid for their participation. All were native speakers of English and were naive as to the purposes of the study

3.2. Materials

Twenty-four items (listed in full in Appendix C) were constructed following the pattern of (11). Each item consisted of a sentence-initial subject (determiner plus noun) followed by a word sequence with parts of speech as shown in (10), then the word *which* or *who* (depending on condition), and finally *was* plus four to seven more words to complete the relative clause and the sentence. The first post-verbal noun was always a singular human noun (usually denoting an occupation), and the second post-verbal noun was always a singular noun denoting an inanimate entity (e.g., *movie* in (11)), an organization-like (e.g., *college* or *company*), or an animate non-human entity (e.g., *dog*). Crucially, the second post-verbal noun never denoted a singular animate human entity, so that the relative pronoun *who* in the non-adjacent RC conditions should always bias RC attachment toward NP_1.

In addition to the target sentences, 120 filler sentences with various syntactic structures were included, including sentence materials from two other experiments. Each participant saw only one of the four versions of each sentence, according to a Latin-square design. The stimuli were pseudo-randomized separately for each participant, so that a

Table 7
Question–answering accuracy in Experiment 2.

	RC adjacent	RC non-adjacent
VP-attached	0.88	0.85
NP-attached	0.89	0.87

Table 8
F-statistics for analysis of question–answering accuracy in Experiment 2. No effects were significant.

	*F*1	*F*2
PP attachment	<1	<1
RC adjacency	1.28	<1
PP × adjacency	<1	<1

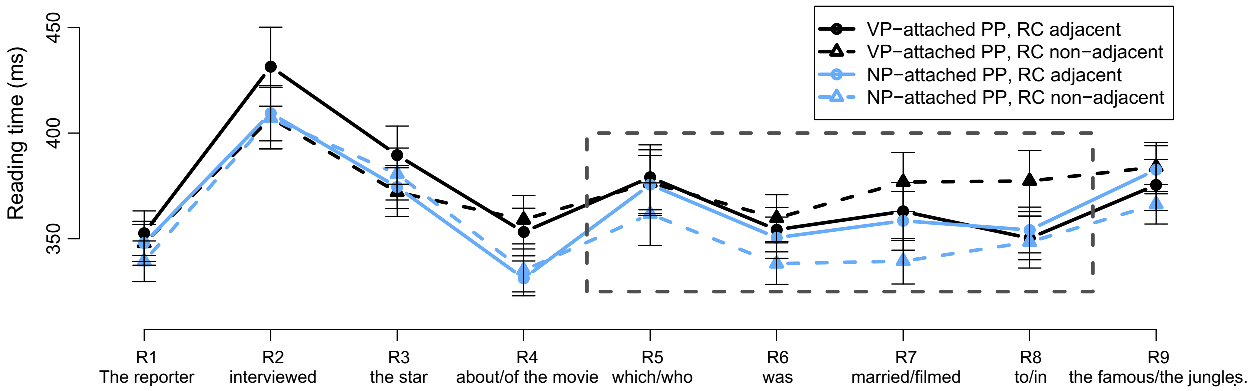


Fig. 4. Reading-time results as a function of region and condition for Experiment 2. Onset of the relative clause (first four words) is boxed.

Table 9
F-statistics for Experiment 2. In regions not listed, there were no effects were below $p = 0.1$. See Footnote 8 regarding the marginal effect in Region 3.

	R1		R2		R3		R4		R5		R6		R7		R8		R9	
	F1	*F2*	*F1*	*F2*	*F1*	*F2*	*F1*	*F2*	*F1*	*F2*	*F1*	*F2*	*F1*	*F2*	*F1*	*F2*	*F1*	*F2*
PP attachment	1.87	<1	1.25	1.45	<1	<1	16.08‡	8. 38†	<1	<1	3.33$^{\cdot}$	3.59$^{\cdot}$	8.12†	3.73$^{\cdot}$	3.06$^{\cdot}$	<1	1.14	<1
RC adjacency	1.57	1.09	1.76	1.37	<1	<1	<1	<1	<1	1.09	<1	<1	<1	<1	2.77	1. 09	<1	<1
PP × adjacency	<1	<1	1.64	1.03	3.40$^{\cdot}$	1.88	<1	<1	<1	<1	1.83	<1	4.16*	2.56	3.94$^{\cdot}$	1.51	4.12*	2.57

$^{\cdot}$ $p < 0.1$.
* $p < 0.05$.
† $p < 0.01$.
‡ $p < 0.001$

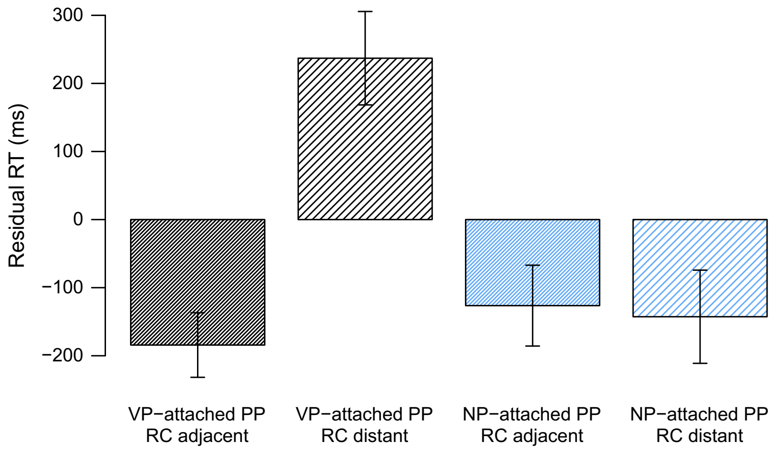


Fig. 5. Residual reading times over the first four words of the RC in Experiment 2.

Table 10
F-statistics for residual reading-time analysis across first four words of the RC in Experiment 2.

	F1	*F2*
PP attachment	13.23‡	10.60†
RC adjacency	10.12†	10.33†
PP × adjacency	14.01‡	9.78†

† $p < 0.01$.
‡ $p < 0.001$.

target sentence never immediately followed another target sentence.

3.3. Procedure

Sentences were presented to participants using the same moving-window self-paced reading technique as in Experiment 1. Every word was displayed individually.

The study took an average of 50 min per participant to complete.

3.4. Results

3.4.1. Statistical analysis

Statistical analysis procedures were identical to those in Experiment 1. Outlier removal procedures led to 2.5% of reading-time data being discarded.

3.4.2. Comprehension accuracy

Tables 7 and 8 show question–answering accuracy by condition, together with the results of 2×2 ANOVAs. There were no significant differences by condition; participants' overall comprehension accuracy was high across the board, though there is a hint of a trend toward greater difficulty for non-adjacent and especially extraposed RCs.

3.4.3. Reading times

Fig. 4 shows average per-word reading times for each region, and Table 9 gives the results of 2×2 ANOVAs by participants and by items for each region.[9] Within the relative clause we see a numerical trend toward an interaction with RTs in the extraposed-RC condition (VP/non-adjacent) highest. This interaction reaches significance by participants in Regions 7 and 9, and is marginal by participants in Region 8. Since these regions involve considerably different word sequences as a function of RC adjacency, we conducted a residual reading-time analysis across the first four words of the RC, using the same procedure as in Experiment 1 (2.0% data loss from outlier procedures). Fig. 5 shows the average residual RT per region as a function of condition, and Table 10 reports results of a 2×2 ANOVA on these data. We see significant main effects of both PP attachment and RC adjacency, driven by a significant interaction, such that the extraposed-RC condition is most difficult. Pairwise comparisons showed significant effects of PP attachment within the RC non-adjacent conditions ($F_1(1,43) = 20.62$, $p < .001$; $F_2(1,19) = 14.04$, $p = 0.001$) and of RC adjacency within the VP-attached PP conditions ($F_1(1,43) = 25.31$, $p < .001$; $F_2(1,19) = 19.31$, $p < .001$).

3.5. Discussion

Reading times within the RC show an interaction between PP attachment and RC adjacency, with reading times superadditively difficult in the extraposed-RC condition. This pattern emerges most clearly in the third word of the RC, but is evident already at the relative pronoun, the first available cue as to correct attachment. This is the pattern of results predicted by derivational-complexity and structural-expectation theories. It is problematic for decay- or retrieval interference-based accounts of the difficulty observed in Experiment 1 for RC extraposition, because these theories predict no effect of PP attachment on processing difficulty within the RC.

Close inspection of the word strings used in this experiment also affords us some ability to discriminate between theories of pure collocational expectations versus those of structural expectations. The minimal-length collocation that would be required to capture the four-way distinction among conditions would be a 4-gram—in (11), for example, *of/about the move that/who*. It is difficult to estimate the experience that an average college-age speaker might have with such a collocation, but as a crude estimate we consulted the 1-trillion-word Google *n*-grams corpus (Brants & Franz, 2006) for the frequencies of the relevant collocations in our materials. In order for a pure collocational-expectation theory to explain the interaction between PP attachment and RC adjacency, such speakers would on average need to have been exposed to each of the non-extraposed variants of such 4-grams (*of the movie that*, *of the movie who*, and *about the movie that* in our example) at least once. For this example, the total counts found in the Google *n*-grams corpus are 22256, 618, and 3141 respectively. If we were to estimate that the average college-age native English speaker has had lifetime exposure to no more than 350 million words of English with distribution similar to that of Web-based documents, the expected number of exposures to these collocations would be 7.8, 0.2, and 1.1.[10] For the other 19 items we find that the arithmetic mean occurrence of the non-extraposed condi-

[9] In this and Experiment 3, there are a few numerically small "precognitive" effects which reach some degree of statistical significance but which precede divergence of the input available at that point in the relevant experimental conditions—here, a marginal by-subjects interaction between PP type and RC adjacency at Region 3—before any differences between conditions are available to the comprehender. Because no effects are predicted in these cases, and because there are many such regions in each experiment, some sort of correction for multiple comparisons would be appropriate, but we are reluctant to impose a specific correction procedure since the time-series nature of SPR times makes it difficult to quantify precisely how many comparisons we should consider to have been made. However, even a relatively mild correction—e.g., Bonferroni correction for 10 tests—would render all these effects statistically insignificant. For completeness, we have left the original *p*-values in the tables reporting *F*-statistics, but we do not discuss these presumably spurious effects in the text. Additionally, for these two experiments we replicated the crucial effects seen in the RCs in other experiments not reported in detail here, and the "precognitive" effects do not replicate in these experiments; see Footnotes 11 and 12.

[10] This estimate can be obtained in a number of ways. Roy, Frank, and Roy (2009) collected 4260 h of audio recordings during months 9 through 24 of an American child's life, which they estimate contains under 10 million words of speech by or audible to the child; extrapolating this figure leads to about 300 million words over 20 years. Hart and Risley (1995) estimate that a 4-year old in a professional American family has heard roughly 50 million words; extrapolating this leads to about 250 million words over 20 years. Finally, Mehl, Vazire, Ramírez-Esparza, Slatcher, and Pennebaker (2007) estimated on the basis of 31 days of audio recordings of 396 university students that the average student speaks approximately 16,000 words a day. If the average speaker hears three times as many words a day, extrapolating over 20 years also leads to about 300 million words. The expected number of exposures to the collocation in question is almost certainly an overestimate, however, given that business-related documents are over-represented on the Web, in comparison with the life experience of most native-English speaker MIT college students.

Table 11
Relative frequency in parsed Brown corpus of RC postmodifier for various types of premodifying structure. Totals include both extraposed and unextraposed RCs.

Premodifier	Proportion RC postmodification (%)	*n*
a/an	8.1	10071
the	7.5	22193
only	25.4	307
the only	67.6	74
only those	100	1

tion with minimum count is 952, corresponding to an expected number of exposures of 0.3. This analysis tentatively suggests that the strings in question may simply be too rare for our results to find a likely explanation in purely collocational-expectation theories.[11] As demonstrated in Table 6, however, the structural configurations in question are sufficiently common for adult native speakers to have had ample direct experience of their relative frequencies, and predict the interactive pattern cleanly.

4. Experiment 3

We have suggested two possible sources for the comprehension difficulty associated with extraposed structures observed in the experiments reported thus far: derivational complexity or probabilistic expectations. On the former account, extraposed RCs should be uniformly more difficult to process than in situ RCs; on the latter account, it must be the case that in the stimuli we have used thus far, the RCs are less expected in the extraposed conditions than in the unextraposed conditions, and this difference in expectation is reflected in reading times and question–answering accuracy. If the latter account is correct, the difficulty seen with extraposed RCs should not be inevitable but rather contingent on the probabilistic expectations computed by the comprehender based on what precedes the particular extraposed RC in the sentence. If we can find some way of manipulating the comprehender's expectations for extraposed versus unextraposed RCs, we may be able to distinguish between the derivational-complexity and probabilistic-expectations accounts. In the following two experiments, we put this idea to the test.

We take advantage of recent work by Wasow et al. (2006; see also Jaeger, 2006, Levy & Jaeger, 2007) indicating that the probability of various types of NP postmodifiers is strongly dependent on the prenominal structure of the NP. Wasow et al., show for example, that in a dataset derived from the parsed Switchboard corpus of conversational English (Godfrey, Holliman, & McDaniel, 1992; Marcus et al., 1994), five times as many definite NPs (with the determiner *the*) have a postmodifying (non-subject-extracted) RC as do indefinite NPs (with the determiner *a/an*). Other types of premodifiers can create even stronger expectations—for example, the word *only* is associated with a considerably higher rate of RC postmodification. Intuitively, the reason for this involves semantics, pragmatics, and world knowledge: *only* imposes an exclusivity requirement on some aspect of the proposition encoded by the clause in which it appears (von Fintel, 1994, inter alia), and it seems to be a contingent fact about language use that part of establishing this exclusivity tends to involve refining the domain of nominal reference with a postmodifier. Table 11 shows the proportion of NPs possessing projective-dependency RC postmodifiers for various types of premodifying structure in the parsed Brown corpus (non-projective RC postmodifiers were too rare to obtain reliable statistics for in these cases). Although the *a*(*n*)/*the* contrast is minimal in the parsed Brown corpus (which is written, rather than spoken, English), the use of *only* considerably increases the expectation for a postmodifying RC.

On the probabilistic-expectations account, prenominal structure that establishes a strong expectation for a postmodifying RC should facilitate comprehension of such an RC when it is encountered. Some evidence exists in the literature demonstrating such effects: in particular, Frazier et al. (1995, Chapter 4) examined reading of sentences on the pattern given in (12).

(12) a. Max met the only one of Sam's employees who {has/have} teeth...
b. Max met the only supervisor with employees who {has/have} teeth...

Within the framework described in the present paper, in (12a) the prenominal content *the only* sets up an expectation for a postnominal modifier of *one* which remains unmet by the time the RC onset *who* is encountered. Thus the comprehender should expect attachment to the first noun (*one*) at this point, and should incur processing difficulty when attachment turns out to be to the second noun (*employees*, e.g., when the sentence continues with *have*). In (12b), in contrast, the expectation for a postnominal modifier has already been met by the phrase *with employees*, thus upon seeing the RC onset the comprehender's

[11] In another experiment with a similar design but whose details are omitted for reasons of space (originally presented in Gibson & Breen (2004)), we found the same reading-time results but the materials permit an even stronger case against purely collocational expectations. This experiment used items on the pattern of (i):

(i) a. The chairman consulted the executive of the companies that were making lots of money [NP-attached PP, RC adjacent].
b. The chairman consulted the executive of the companies that was making lots of money [NP-attached PP, RC non-adjacent].
c. The chairman consulted the executive about the companies that were making lots of money [VP-attached PP, RC adjacent].
d. The chairman consulted the executive about the companies that was making lots of money [VP-attached PP, RC non-adjacent].

Reading times indicated the same superadditive interaction with the extraposed condition (id) most difficult. Here, however, the shortest strings differentiating the conditions—*of/about the companies that was/were*—are 5-grams, the greatest expected number of exposures to any item in any non-extraposed condition is 0.7, and the arithmetic mean over all items in all non-extraposed conditions is 0.07. Additionally, the "precognitive" effect seen in Region 3 of Experiment 2 did not replicate.

expectations lean more toward attachment to the second noun (*employees*). This pattern of results was predicted for similar reasons by Frazier and Clifton, who argued that the semantic requirements of *only one* would modulate RC attachment preferences (though they did not cast the prediction within a probabilistic framework as we are doing here), and was confirmed in a self-paced reading study.

According to expectation-based theories, this expectation-based facilitation of NP-postmodifying structures should extend to extraposed RCs, as well—but only if (i) the corpus data and intuitions grounded in semantics, pragmatics, and world knowledge regarding postmodifier frequency generalize to non-projective RC structures, and (ii) comprehenders are able to maintain and take advantage of corresponding expectations outside of the projective-dependency domain. In the present experiment, we test a specific version of the probabilistic-expectations hypothesis in which (i) and (ii) are true, using a premodifying collocation, *only those*, which intuitively gives rise to a very strong expectation for a postmodifier. The collocation *only those* was too rare in parsed corpus data to obtain reliable frequencies of co-occurrence with postmodifying RCs, but in Section 4.3 we describe a completion study that corroborates this more specific intuition. In the comprehension study, we modified the design from Experiment 2 to cross extraposition with expectation, as in (13):

(13) a. The chairman consulted the executives about the company which was acquired recently by an aggressive rival firm [WEAK expectation for NP_1 postmodification, −extraposition].
b. The chairman consulted the executives about the company who were highly skilled and experienced in the industry [WEAK expectation for NP_1 postmodification, +extraposition].
c. The chairman consulted only those executives about the company which was acquired recently by an aggressive rival firm [STRONG expectation for NP_1 postmodification, −extraposition].
d. The chairman consulted only those executives about the company who were highly skilled and experienced in the industry [STRONG expectation for NP_1 postmodification, +extraposition].

Examples (13a)–(b) are identical in design to the VP-attached conditions of Experiment 2, hence the RC should be harder to process in the extraposed variant (13b) than in the unextraposed variant (13a). Examples (13c)–(d) differ in that the prenominal material *only those*, which modifies the direct-object (DO) noun *executives*, should create a strong expectation for a relative clause that postmodifies *executives*. We call this an expectation for NP_1 modification, since *executives* is the first noun in the post-verbal domain. This expectation is not satisfied by the immediately following constituent *about the company*, because this PP is a dependent of the verb rather than of NP_1. After encountering *about the company*, it becomes clear that any postmodifier of NP_1 that may appear later in the sentence cannot form a continuous constituent (i.e. a projective dependency) with it, but must rather be extraposed. If the comprehender nevertheless maintains a strong expectation for an NP_1 postmodifier beyond the continuous-constituent domain, however, we should see that reading of the extraposed RC in (13d) is facilitated relative to the weak-expectation variant, (13b). We note a further prediction that arises if online syntactic comprehension truly is *probabilistic*—that is, there are limited overall resources to be allocated among possible upcoming constituent types (formally, the probabilities assigned to upcoming constituent types must sum to 1; Jurafsky, 1996; Hale, 2001; Levy, 2008), so that increasing the expectation for one type of constituent through a manipulation entails that expectations for some other type or types of constituent must correspondingly decrease. If this is the case, then the increased expectation for an extraposed RC in the STRONG expectation conditions of (13) should have the effect of decreasing the expectation for other types of constituents, including an unextraposed RC modifying *company*. Therefore we predict additionally that the unextraposed RC should be harder to read in (13c) than in (13a). We tested these predictions in a self-paced reading study using sets of sentences as in (13).

4.1. Method

4.1.1. Participants

Thirty-two native English speakers at MIT participated in this study for cash compensation.

4.1.2. Materials

We constructed 24 items (listed in full in Appendix D) on the pattern of Example (13). Each item consisted of a sentence-initial subject (determiner plus noun), followed by a transitive verb, then a plural direct object ({*the*/*only those*} + *noun*), then a prepositional phrase consisting of a preposition plus a singular definite noun, then the relative clause. The contents of the RC were different in the extraposed and unextraposed conditions, and were constructed to serve as a strong cue as to the correct attachment. These test sentences were interspersed among 80 fillers, including 24 from one other experiment. Among the fillers we included six of the form seen in (14):

(14) The chairman consulted only those executives about the company which was acquired recently who had worked at that company previously.

Table 12
Question–answering accuracy in Experiment 3.

	Extraposed	Unextraposed
Strong-expectation	0.82	0.84
Weak-expectation	0.91	0.88

to ensure that comprehenders were exposed to cases of *only those* noun premodification in which the putative expectation for a postmodifier was satisfied, but only after a projectively-attached RC on the intervening NP (see Section 4.4 for more discussion). We also included one such item among the practice sentences.

4.1.3. Procedure

The procedure was the same as in previous experiments. The study took an average of 30 min per participant to complete.

Table 13
F-statistics for analysis of question–answering accuracy in Experiment 3.

	*F*1	*F*2
Expectation	5.64*	4.69*
Extraposition	<1	<1
Expect × extrap	1.31	<1

* $p < 0.05$.

4.2. Results

4.2.1. Statistical analysis

Due to a programming error, the question for item 21 in the unextraposed conditions was incorrectly written and the normative answer was incorrectly coded. We present analyses of question–answering data with this item omitted, though we include this item in reading-time analyses. Three participants did not answer comprehension questions to all items (experimental and fillers jointly) significantly above chance; data from these participants were discarded. Statistical analysis procedures for the remaining data were identical to those in previous experiments. Outlier removal procedures led to 1.9% of reading-time data being discarded.

4.2.2. Comprehension questions

Question answering accuracies and 2 × 2 ANOVA results are reported in Tables 12 and 13. We found a significant main effect of expectation, with accuracies lower in the strong-expectation conditions.

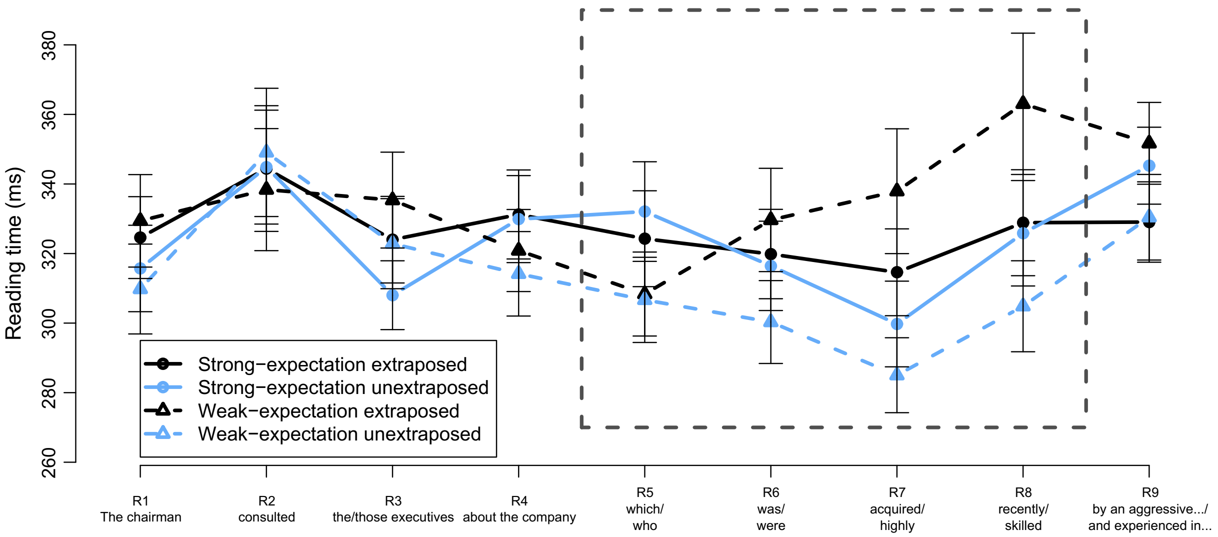


Fig. 6. Region-by-region reading times for Experiment 3.

Table 14
F-statistics for Experiment 3. In regions not listed, there were no effects were below $p = 0.1$.

	R1		R2		R3		R4		R5		R6		R7		R8		R9	
	*F*1	*F*2	*F*1	*F*2	*F*1	*F*2	*F*1	*F*2	*F*1	*F*2	*F*1	*F*2	*F*1	*F*2	*F*1	*F*2	*F*1	*F*2
Expectation	<1	<1	<1	<1	6.34*	1.85	7.87†	2.27	11.98†	6.40*	<1	<1	<1	<1	<1	<1	1.44	<1
Extraposition	9.19†	7.72*	<1	<1	8.31†	8.07†	<1	<1	<1	<1	5.69*	3.76$^{\cdot}$	17.25‡	28.74‡	12.66†	18.61‡	<1	<1
Expect × extrap	1.17	<1	<1	<1	<1	<1	<1	<1	<1	<1	5.07*	1.76	6.21*	4.25$^{\cdot}$	9.07†	4.38*	8.53†	3.57$^{\cdot}$

$^{\cdot}$ $p < 0.1$.
‡ $p < 0.001$.
† $p < 0.01$.
* $p < 0.05$.

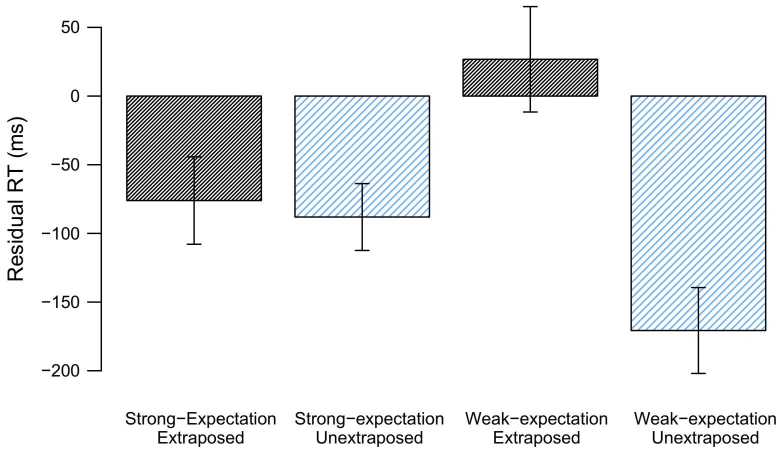


Fig. 7. Residual reading times over the first four words of the RC in Experiment 3.

Table 15
F-statistics for residual reading-time analysis across first four words of the RC in Experiment 3.

	*F*1	*F*2
Expectation	<1	<1
Extraposition	13.41[a]	16.69[b]
Expect × extrap	6.82[*]	14.90[b]

[a] $p < 0.01$
[b] $p < 0.001$.
[*] $p < 0.05$

4.2.3. Reading Times

Fig. 6 shows region-by-region reading times, and Table 14 gives the results of 2 × 2 ANOVAs by participants and by items for each region. (The word *only* is omitted from these analyses as it does not correspond to any word in the weak-expectation condition.) At Region 5 we find a significant main effect of expectation, with reading times slower in the strong-expectation condition. In Regions 6–9, we find an interaction between expectation and extraposition with extraposed RCs read more slowly than unextraposed RCs in weak-expectation conditions but not in strong-expectation conditions. This interaction is significant by both participants and items in Region 8, and significant by participants in Regions 6, 7, and 9. We also find main effects of extraposition in Regions 6–8, with unextraposed RCs read more quickly overall than extraposed RCs.

As in the previous two experiments, we conducted a residual reading-time analysis across the first four words of the RC, using the same methodology as described in Section 2.4.3 (2.6% data loss from outlier removal procedures). Fig. 7 shows the average residual RT per region as a function of condition, and Table 15 reports results of a 2 × 2 ANOVA on these data. Analysis reveals a significant interaction between expectation and extraposition. We also see a significant main effect of extraposition, with extraposed RC reading times slower overall. Pairwise comparisons reveal a significant effect of extraposition in the weak-expectation condition ($F_1(1,28) = 13.30$, $p = 0.001$; $F_2(1,23) = 26.41$, $p < .001$) but not in the strong-expectation condition (*Fs* < 1); the effect of expectation in the extraposed condition is marginal by subjects, significant by items ($F_1(1,28) = 3.82$, $p = 0.061$; $F_2(1,23) = 6.90$, $p = 0.015$), and significant in the unextraposed condition ($F_1(1,28) = 4.29$, $p = 0.048$; $F_2(1,23) = 8.49$, $p = 0.008$).[12]

4.3. Completion study

We followed up this self-paced reading study with a completion study designed to estimate comprehenders' expectations for NP_1- versus NP_2-modifying RCs in the strong- versus weak-expectation condition, for two reasons. First, completion study results may serve as corroborating evidence for the intuition and corpus data suggesting that the use of *only those* as a premodifier truly increases the expectation for a modifying RC. Second, the absolute magnitude of the extraposition effect is numerically larger in the weak-expectation condition (189 ms in favor of unextraposed RCs) than in the strong-expectation condition (145 ms in favor of extraposed RCs), which could possibly be interpreted as an overall processing penalty for extraposition, but could alternatively arise in a purely probabilistic framework if the relative expectations for NP_1- versus NP_2-attaching RCs are more balanced in the strong-expectation than in the weak-expectation condition. For both these reasons, it is of considerable interest to quantify the precise strengths of these expectations in the *the* versus *only those* conditions.

[12] In another self-paced reading study omitted for reasons of space, with design similar to Experiment 3 but attachment disambiguation through auxiliary number marking as described in Footnote 11, we replicated the interaction in RC reading times. There was no trace of the "precognitive" effects seen in regions 1 or 3 of Experiment 3.

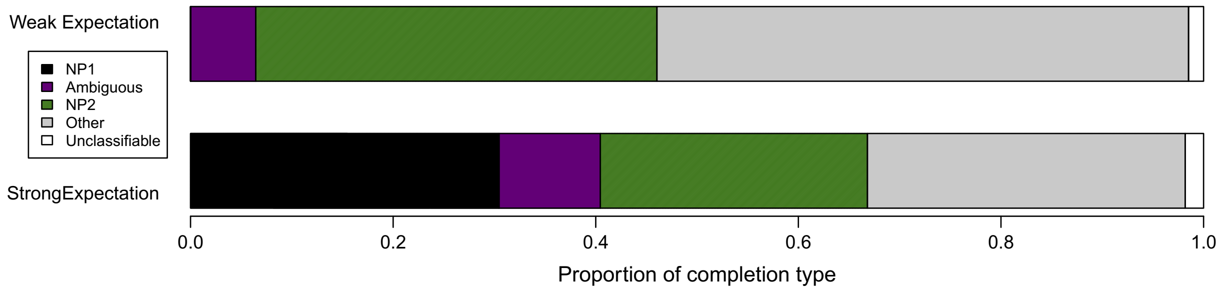


Fig. 8. Completion-study results for Experiment 3.

The completion study used the pre-relative clause sentence fragments from the self-paced reading study items as prompts, as in Example (15):

(15) a. The chairman consulted the executives about the company...
b. The chairman consulted only those executives about the company...

Twenty-two native English speakers participated in this study for cash compensation. For each participant, half the items were presented in the weak-expectation *the* condition (15a) and the other half were presented in the strong-expectation *only those* condition (15b); which items were presented in which condition was rotated across participants. Two native English speaker research assistants naive to the goals of the study coded each completion as (i) an extraposed RC attaching to the first of the two post-verbal nouns (NP_1), (ii) an in situ RC attaching to the second of the two post-verbal nouns (NP_2), (iii) an RC ambiguous between attachment to NP_1 and NP_2, (iv) any other type of clearly interpretable continuation that is not an RC, and (v) unclassifiable. Coding was done conservatively, with RCs for which there was any doubt as to the proper attachment between NP_1 and NP_2 coded as ambiguous. Cases of disagreement among annotators were discarded. Examples of cases (i–iv) are given in (16):

(16) a. The reporter interviewed only those actors about the movie *who had a speaking role* (strong-expectation, NP_1).
b. The architect consulted only those carpenters about the project *that was about to begin* (strong-expectation, NP_2).
c. The judge queried the lawyers about the evidence *that was used to indict the suspect* (weak-expectation, NP_2).
d. The principal criticized only those instructors for the program *that had just started* (strong-expectation, ambiguous).
e. The socialite praised the hostesses for the party *because it was great* (weak-expectation, other).
f. The chairman consulted only those executives about the company *issues* (strong-expectation, other).

Fig. 8 shows summary results of this coding. In the weak-expectation condition there were no NP_1-attaching RC continuations, whereas in the strong-expectation condition there were more NP_1-attaching continuations than NP_2-attaching continuations; this latter difference was highly significant by a likelihood-ratio test on mixed-effects logit models with subject-specific random intercepts and item-specific random slopes ($\chi^2(1) = 21.17$, $p \ll 0.001$).[13] However, the preference for NP_2 attachment in the weak-expectation condition was significantly greater than the preference for NP_1 in the strong-expectation condition by a likelihood-ratio test ($\chi^2(1) = 30.79$, $p \ll 0.001$). Thus the strengths of expectations closely match those of the reading-time study.

4.4. Discussion

The key results of the self-paced reading study in this experiment were the interaction between expectation for a relative clause modifying the direct object (NP_1; *executives* in (13)) and the attachment site of a relative clause following the verb-modifying PP (*about the company* in (13)) on question answering accuracies, and even more crucially on reading times in the onset of the RC. When the expectation for an NP_1-modifying relative clause is weak, an unextraposed (NP_2-modifying) RC is read considerably faster than an extraposed (NP_1-modifying) RC. When the expectation for an NP_1-modifying RC is strong, on the other hand, the reading-time difference is neutralized. This interactive pattern can be understood as a consequence of probabilistic syntactic expectations: in the presence of a strong expectation that NP_1 should have a postmodifier, the comprehender may expect more strongly that the next constituent is an extraposed NP_1 modifier than an unextraposed NP_2 modifier. The main effect of extraposition is consistent with an effect of derivational

[13] Models with subject-specific random slopes failed to converge. We use the likelihood-ratio test here instead of a Wald z-test because the absence of NP_1 responses in the weak-expectation condition leads to an extremely large coefficient and an accompanying inflated standard error for the fixed effect of *the/only those* and makes the Wald statistic unreliable. This is well known to be a problem with the Wald statistic for logit models with large coefficient estimates (Agresti, 2002; Menard, 1995).

complexity, though it might also reflect the differing lexical content of extraposed and unextraposed RCs. Additionally, the completion study indicates that the relative expectation for an unextraposed RC in the weak-expectation condition is in fact stronger than the relative expectation for an extraposed RC in the strong-expectation condition, suggesting that this effect could be due to an asymmetry in the strengths of the pertinent probabilistic expectations.

We are not aware of any plausible analysis of the crucial interaction found here that does not include a probabilistic-expectation component. The best non-probabilistic analysis that we were able to come up with would involve a combination of two factors: (A) a fundamental processing cost to processing extraposed structures, and (B) a categorical infelicity in the strong-expectation unextraposed condition induced when no NP_1-postmodifier is found to satisfy the uniqueness requirements imposed by the premodifier *only those*. On this analysis, factor (A) would determine the pattern in the weak-expectation contexts, and factor (B) would determine the pattern in the strong-expectation contexts. However, what crucially militates against this analysis—in addition to parsimony concerns–is that during reading of an unextraposed RC it is still too early for the comprehender to know that NP_1 will not ultimately have an extraposed postmodifier appearing later in the sentence. For example, in (14), repeated below for convenience, an extraposed RC appears immediately after the unextraposed RC.

(14) The chairman consulted only those executives about the company which was acquired recently who had worked at that company previously.

As mentioned in Section 4.1.2, we included a number of sentences with this form as fillers and as a practice sentence in the present study to ensure that comprehenders had specific experience of such a possibility. Categorical infelicity could thus only be evaluated when the end of the sentence is reached; on such a theory, we should thus not see reading-time effects in an example like (13c) early in the RC. The probabilistic-expectations account avoids these difficulties: having a strong expectation that an extraposed RC immediately follows NP_2 entails that the expectation for an unextraposed RC at that position must be weak, leading to predicted reading-time effects consistent with our empirical results.

5. General discussion

The results of our three experiments can be summarized as follows:

- Relative clauses extraposed from simple [determiner + noun] NPs across a verb are harder to process than their corresponding in situ variants.
- Relative clauses extraposed from a direct object NP across a PP are harder to process than in situ relative clauses modifying either the direct object (but following the PP) or the PP-internal NP.
- Nevertheless, a preceding context (specifically, NP-internal premodifiers) that sets up a strong expectation for a relative clause modifying a given noun can strongly facilitate comprehension of an extraposed RC modifying that noun;

These results are supportive of the structural-expectations account described in Section 1.2.3. The results of Experiment 2 provide evidence in favor of derivational complexity and/or probabilistic expectations over retrieval/interference and/or collocational expectations. In particular, the linear arrangement of NPs and RCs disentangled distance (and number of intervening NPs) from extraposition, with extraposition being the crucial factor leading to processing difficulty; and the word collocations in question are too infrequent for experimental participants to be likely to have had direct experience with them. Experiment 3 provides evidence in favor of expectations over derivational complexity (and over retrieval/interference): in an arrangement where an RC could in principle modify either the immediately preceding noun phrase (NP_2), or alternatively an earlier noun phrase (NP_1) through extraposition, giving NP_1 a strong unfulfilled expectation for a postmodifier can neutralize the difficulty associated with extraposition.

The experiments reported here can thus be added to a number of recent results demonstrating that expectations for not only specific upcoming words (Ehrlich & Rayner, 1981; Kutas & Hillyard, 1980, 1984) but also upcoming constituent types (Jaeger et al., 2008; Lau, Stroud, Plesch, & Phillips, 2006; Staub & Clifton, 2006; Staub, Clifton, & Frazier, 2006) can facilitate online sentence comprehension. Our results go beyond these findings in one crucial respect, however, being the first experimental demonstration that comprehenders use syntactic expectations for non-projective dependency structures in online comprehension.[14] It has been shown (e.g., Boland et al., 1995; Traxler & Pickering, 1996) that in the processing of filler-gap dependencies such as *That's the {garage/pistol} with which the heartless killer* SHOT..., verbs that are more predictable given the filler are processed more easily than verbs that are less predictable. In this previous work, however, the relative facilitation could conceivably be attributed to plausibility differences computed as the semantic content of filler and governing verb is combined, rather than to prediction. In our Experiment 3, however, there are signs that the crucial interactive pattern starts to emerge already at the relative pronoun *who/which*, at which point it is not clear how plausibility differences alone would give rise to the interactive pattern observed between prenominal content and RC attachment. Rather, the more natural explanation is that comprehenders had formulated expectations about the likely attachment site of any relative clause before they had encountered the relative pronoun, that these beliefs are

[14] The heavy noun-phrase shift constructions used in Staub et al. (2006), along with related verb–particle constructions are treated as involving discontinuous constituents in some syntactic analyses, but they do not involve non-projectivity, and correspondingly some syntactic analyses (e.g., Pollard & Sag, 1994) treat these as strictly continuous-constituent constructions.

consistent with the relevant conditional probabilities that we have estimated from corpora and from our completion study, and that the reading time pattern observed reflects the differential consistency of these expectations with the attachment cues of animacy and agreement available in the first two words of the relative clause.

The expectation-based analysis also explains the results of Francis (2010), who found that sentences containing subject-modifying RCs in English (like those of our Experiment 1) were read more quickly with the RC in situ than with it extraposed when the RC was short, but more quickly with the RC extraposed than with it in situ when the RC was long. Since the distribution of RC extraposition is precisely that longer RCs are more often extraposed than shorter RCs (Francis, 2010; Hawkins, 1994; Uszkoreit et al., 1998), the comprehender's encountering a relative pronoun in an extraposed setting should set up an expectation that the RC is likely to be long. On this analysis, comprehenders in Francis's experiment reaped benefits when this expectation was met (fast reading for long extraposed RCs) and paid a cost when it was not (relatively slow reading for short extraposed RCs). Expectation-based theories thus unify our results with those of Francis (2010): the former demonstrate the effect of context on the expectation for *whether* an RC will be encountered at any given point in a sentence, while the latter demonstrate the effect of context on the expectation, given that an RC is encountered, for *what the RC and the rest of the sentence will be like*.

Taken together with related findings, our results thus have implications for the nature of the representations used in the online computation of sentence structure in comprehension. As depicted in Fig. 1, there are several formal means of encoding the non-projective dependencies induced by right-extraposition into the kinds of phrase structure representations on which most models of incremental processing are based. Crucially, our experiments show that a strictly context-free phrase-structure representation *without* any percolation of missing-RC information out of the NP – e.g., Fig. 1b without slashes – would be insufficient to account for the full range of syntactic comprehension effects in online sentence processing, if phrase-structure locality were taken to encode independence assumptions about events in the tree—that is, what happens outside of a given node is independent of what happens inside a given node. The reason for this is that in Experiment 3, we found that the prenominal content of an NP can affect comprehenders' expectations about how likely a relative clause is to appear in a position that, in phrase-structure terms, is strictly outside the NP. We are left with two alternatives: either the syntactic representations computed online must allow information inside a node to influence expectations about what will happen outside a node (Fig. 1a, b, and d), or they must allow the explicit representation of discontinuous constituents (Fig. 1c). A formalism for describing the knowledge deployed in online syntactic comprehension should therefore be at least as expressive as either a context-free grammar with a slash-passing component (as in Generalized Phrase Structure Grammar; Gazdar et al., 1985), or a mildly context-sensitive formalism such as Tree-Adjoining Grammar (Joshi et al., 1975), Combinatory Categorial Grammar (Steedman, 2000), or Minimalist Grammar (Stabler, 1997).[15] Any of these formalisms—when coupled with a probabilistic component along the lines of Resnik (1992) and others—would be adequate to express this knowledge; as noted by Joshi et al. (1991), weak generative capacity is equivalent among many of these formalisms.

Finally, let us point out that although our results hold out the prospect for understanding the typical difficulty associated with English RC extraposition as derivative of construction frequency and expectation-based syntactic processing, we have not provided an explanation for the rarity of non-projectivity in the first place, either in English or cross-linguistically. Our results suggest that this rarity may not itself derive from inherent difficulty for the comprehender, as Experiment 3 showed that the difficulty of RC extraposition can be completely neutralized in some contexts. Although thorough study of alternative sources of explanation is well beyond the scope of this paper, we note here that speaker choices in RC extraposition in particular are likely to implicate both information structure (Huck & Na, 1990; Rochemont & Culicover, 1990) and grammatical weight (Arnold, Wasow, Losongco, & Ginstrom, 2000; Gibson, 1998; Hawkins, 1994, 2004; Szmrecsanyi, 2004; Wasow, 1997, 2002; Yngve, 1960), both of which induce linear-ordering preferences for independent functional motivations. It may be the case that the information-structural and grammatical-weight conditions under which speakers would choose extraposition and other non-projective structures simply occur infrequently. Such a possibility may be a good starting point for future investigations of the source of the distribution of non-projectivity, both within and across languages.

Acknowledgments

Portions of this work have benefited from feedback from presentations at the 2004 and 2008 CUNY Sentence Processing Conferences, and presentation in 2009 colloquia at the Departments of Linguistics at UC San Diego and UC Berkeley. We are grateful to Elena Churilov and Natalie Katz for assistance in coding the sentence-completion data of Experiment 3. Mike Frank provided valuable suggestions and discussion regarding the estimation of a college-age student's lifetime linguistic exposure (Footnote 10); Matthew Dryer provided valuable discussion regarding the cross-linguistic distribution of extraposition. We would also like to thank David Beaver, Micha Breakstone, Brady Clark, Timothy Desmet, Gabriel Doyle, Janet Fodor, Mike Frank, Colin Phillips, Hannah Rohde, Carson Schutze, Tom Wasow, Florian Wolf, and three anonymous reviewers for valuable comments and feedback; and Jarrod Hadfield, Luca Borger, and David Duffy for suggestions regarding data analysis. Parts of this work were supported by NSF Grant IIS-0953870 and NIH Grant HD065829 to R.L. All errors are our own responsibility.

[15] Note that the coindexing of distant nodes in the base-generated analysis illustrated in Fig. 1d also introduces expressivity beyond that of context-free grammars, though the precise complexity of such a formalism remains debatable (Ristad, 1993; Manaster Ramer, 1995).

Appendix A. Tree-search patterns for corpus frequencies reported

The following table lists patterns for Table 1, which were applied using the Tregex tree-search tool (Levy & Andrew, 2006). The first three lines are for Example (2a), the last three are for Example (2b). These patterns were applied to a version of the parsed Brown corpus in which $X \rightarrow XY$ adjunction structures were flattened along the lines described in Johnson (1998), to simplify tree search. These searches used a version of the headship rule (invoked by the operators `<#` and `<<#`) that treated VPs as the heads of nested VPs, so that for a complex VP such as *may have arrived*, the pattern fragment `VP <<# (__ !< __)` would pick out the lowest *arrived* as head (across which extraposed RCs would land), rather than the auxiliary verb *may*. In interpreting these patterns, it is important to know that the Penn Treebank annotation practice has been to bracket extraposed relative clauses as daughters of the VP node in the main clause, although this contravenes mainstream syntactic analyses that capture extraposition with slash-passing or movement (see Fig. 1 on the VP bracketing, extraposition from subject would be *downward* slash-passing or movement).

Probability	Conditioning pattern	Count	Outcome pattern	Count
$P(w_i \mid w_{i-1})$	`woman`	186	`woman . who`	13
$P(w_i \mid c_{i-1})$	`NN`	57236	`NN . who`	252
$P(RC \mid \text{context})$	`/^NP-SBJ/ <# __`	54502	`/^NP-SBJ/ <# (__ $+ (@SBAR </^WH/))`	308
$P(w_i \mid w_{i-1})$	`arrived`	33	`arrived . who`	0
$P(w_i \mid c_{i-1})$	`VBD`	27935	`VBD . who`	9
$P(RC \mid \text{context})$	`/^NP-SBJ/ !< (@SBAR !< -NONE-) $+ (VP << # (__ !< __))`	49292	`/^NP-SBJ/!< (@SBAR !< -NONE-) < (SBAR < (-NONE- </ \*ICH\*-([0–9]+)/#1%i)) $+ (VP << # (__ !< __ . (/^SBAR-([0–9]+)/#1%i < /^WH/)))`	4

The following table lists patterns used to obtain counts used in Section 3. (Counts restricted to cases where the VP-attached PP is headed by *about* yield similar patterns, but the counts are low enough as to be unreliable.)

PP/RC Type	Conditioning pattern	count	Outcome pattern	count
NP/adj.	`@NP > @VP <# (__ $+ (@PP <# (__ < of)))`	2603	`@NP > @VP <# (__ $+(@PP <# (__ < of $+ (@NP <# (__ $+ (@SBAR </^WH/))))))`	106
NP/non-adj.	`@NP > @VP <# (__ $+ (@PP <<# of))`	2603	`@NP > @VP <# (__ $+ ((@PP <<# of) $+ (@SBAR </^WH/)))`	106
VP/adj.	`@NP > @VP $+ (@PP <<# __ <- (@NP <# __))`	13276	`@NP > @VP $+ (@PP <<# __ <- (@NP <# (__ $+ (@SBAR < /^WH/))))`	432
VP/non-adj.	`@NP > @VP $+ (@PP <<# __ <- (@NP <# __))`	13276	`@NP < (@SBAR < (-NONE- </^\*ICH\* -([0-9]+)/#1%i )) > @VP $+ (@PP <<# __ <- (@NP <# __) $+ (/^SBAR-([0-9]+)$/#1%i < /^WH/))`	18

The following table lists patterns for Table 11.

Probability	Conditioning pattern	count	Outcome pattern	count
a/an	`@NP < (DT < a|an)`	10071	`@NP < (DT < a|an $++ (@SBAR </^WH/))`	821
the	`@NP < (DT < the)`	22193	`@NP < (DT < the $++ (@SBAR </^WH/))`	1666
only	`@NP < (__ , only)`	307	`@NP < (__ , only $++ (@SBAR </^WH/))`	78
the only	`@NP < (DT < (the . only))`	74	`@NP < (DT < (the . only) $++ (@SBAR </^WH/))`	50
only those	`@NP < (DT < those , only)`	1	`@NP < (DT < those , only $++ (@SBAR </^WH/))`	1

Appendix B. Materials for experiment 1

Underscores indicate word sequences presented as a single region. Items 11, 13, 16, and 24 had the main-clause VP presented as two regions in non-presentative conditions; in all other cases, the main-clause VP was a single region.

1.
   (a) Presentative, RC in situ: In_the_last_scene, a character who was wounded in the battle

appeared and the heroine wept when she saw him.

(b) Presentative, RC extraposed: In_the_last_scene, a character appeared who was wounded in the battle and the heroine wept when she saw him.

(c) Non-presentative, RC in situ: In_the_last_scene, a character who was wounded in the battle died and the heroine wept when she saw him.

(d) Non-presentative, RC extraposed: In_the_last_scene, a character died who was wounded in the battle and the heroine wept when she saw him.

2. During_the_opera, a villain {appeared/lied} who attempted to trick the heroine but the hero came_in and told everyone the truth.
3. At_night, a ghost {materialized/howled} who supposedly was the murdered child and so no one wanted to sleep upstairs alone.
4. After_dinner, a musician {arrived/performed} who was hired for the wedding and the guests danced until midnight.
5. During_the_conference, a researcher {arrived/spoke} who had won the Nobel Prize and the rest of the participants were very excited.
6. After_the_climb, an amateur {came_in/fainted} who had complained all day long but he was ignored by almost everyone at the camp.
7. Yesterday, a customer {came_in/complained} who usually buys lattes every day but the manager wasn't here to see her.
8. After_the_singer, a comedian {came_on/performed} who was famous for his impersonations and the audience fell into hysterical laughter.
9. After_the_show, a performer {came_on/bowed} who had really impressed the audience and everyone went wild with applause.
10. During_the_presentation, an executive {dropped_in/interjected} who was known for inappropriate remarks and the room turned silent with anticipation.
11. Before_last_call, a drunk {dropped_in/passed out} who often bothered the young women and so the manager quickly called him a cab.
12. At_closing, an old lady {entered/remained} who was shopping for her grandchildren but the employees felt bad about kicking her out.
13. During_the_meeting, a parent {entered/spoke up} who was pushing for less homework but the school board didn't want to listen to her.
14. After_a_while, a platoon {passed_by/attacked} who was hiding in the mountains and the villagers fled their homes.
15. At_the_market, a woman {passed_by/apologized} who hit people with her bag and a man asked her to be more careful.
16. At_nine-thirty, a student {ran_in/woke up} who was late for the test but the professor wouldn't let anyone start late.
17. At_the_hospital, a man {ran_in/cried} whose wife had been severely injured but the doctors were eventually able to save her.
18. After_midnight, an entertainer {showed_up/danced} who was hired for the party but the neighbors began to complain about the noise.
19. After_the_class, a student {showed_up/apologized} whose attendance was far from perfect but the professor wouldn't let him make up the test.
20. Yesterday, a patient {stopped_in/complained} who had missed his noon appointment so the doctor agreed to see him early tomorrow.
21. This_morning, a manager {stopped_in/stayed} who is rarely in the office so the employees were very well behaved.
22. On_Saturday, a thief {turned_up/confessed} who was suspected in several crimes after detectives had given up hope of solving the cases.
23. Late_last_week, a boy {turned_up/escaped} who was kidnapped by a cult and the national media descended on his town.
24. Last_week, a relative {came_over/fell ill} who was celebrating her 90th birthday and the rest of the family came to see her.

Appendix C. Materials for experiment 2

For item 1, we present all four conditions; for all other items we present only the conditions in which the PP is VP-attaching, from which the other conditions can be inferred. The VP-attaching conditions always used the preposition *about*; the NP-attaching conditions always used the preposition *of*.

1.

(a) VP/non-local: The reporter interviewed the star about the movie who was married to the famous model.

(b) NP/non-local: The reporter interviewed the star of the movie who was married to the famous model.

(c) VP/local: The reporter interviewed the star about the movie which was filmed in the jungles of Vietnam.

(d) NP/local: The reporter interviewed the star of the movie which was filmed in the jungles of Vietnam.

2. The student petitioned the instructor about the college who was writing a thesis on Philosophy.
The student petitioned the instructor about the college which was founded in the 18th century.
3. The socialite praised the hostess about the party who was preparing a fresh batch of punch.
The socialite praised the hostess about the party which was organized to celebrate the Oscars.
4. The parent called the teacher about the class who was giving bad grades to foreign students.
The parent called the teacher about the class which was held every Wednesday after lunch.
5. The neighbor approached the owner about the dog who was building a doghouse over the property line.
The neighbor approached the owner about the dog which was barking late at night.

6. The policeman questioned the driver about the bus who was directing tourists to the restricted ruins.
   The policeman questioned the driver about the bus which was broken down in front of the museum.
7. The chairman consulted the executive about the company who was playing golf at the country club.
   The chairman consulted the executive about the company which was merging with an internet start-up.
8. The republican challenged the president about the nation who was elected by the left-wing opposition.
   The republican challenged the president about the nation which was located within disputed territory.
9. The reporter approached the victim about the attack who was injured by the suicide bomber.
   The reporter approached the victim about the attack which was planned by the opposition to the government.
10. The principal questioned the member about the clique who was mouthing off to teachers.
    The principal questioned the member about the clique which was gathering by the bleachers after school.
11. The homeowner consulted the architect about the house who was worried about being behind schedule.
    The homeowner consulted the architect about the house which was constructed beside a lake.
12. The sportscaster interviewed the captain about the team who was leading his team to the championship.
    The sportscaster interviewed the captain about the team which was hosting the state tournament.
13. The colonel cautioned the commander about the platoon who was ordering the troops to continue fighting.
    The colonel cautioned the commander about the platoon which was thrown into disarray after heavy casualties.
14. The critic complimented the director about the play who was asked to write the screenplay.
    The critic complimented the director about the play which was opening to rave reviews nationwide.
15. The salesman called the buyer about the rifle who was looking for antiques from the war.
    The salesman called the buyer about the rifle which was manufactured in France before the war.
16. The diner praised the chef about the feast who was trained in the classical tradition.
    The diner praised the chef about the feast which was prepared from authentic ingredients.
17. The activist petitioned the sponsor about the bill who was speaking out against immigration.
    The activist petitioned the sponsor about the bill which was proposed to curb illegal immigration.
18. The officer cautioned the driver about the Explorer who was talking on the phone while driving.
    The officer cautioned the driver about the Explorer which was leaking air from its front tires.
19. The scientist challenged the inventor about the drug who was claiming to have found a cure for cancer.
    The scientist challenged the inventor about the drug which was causing cancer in laboratory animals.
20. The host complimented the author about the book who was being interviewed on all the talk shows.
    The host complimented the author about the book which was autographed for the entire audience.

Appendix D. Materials for Experiment 3

For item 1, we present all four conditions; for the remaining items we present only the low-expectation conditions, with the unextraposed and extraposed RCs presented in that order for each item. For all items, the weak-expectation conditions used *the* as the prenominal material in the main-clause object NP, whereas the strong-expectation conditions used *only those*.

1.
   (a) Weak expectation, unextraposed: The chairman consulted the executives about the company which was acquired recently by an aggressive rival firm.
   (b) Weak expectation, extraposed: The chairman consulted the executives about the company who were highly skilled and experienced in the industry.
   (c) Strong expectation, unextraposed: The chairman consulted only those executives about the company which was acquired recently by an aggressive rival firm.
   (d) Strong expectation, extraposed: The chairman consulted only those executives about the company who were highly skilled and experienced in the industry.
2. The reporter interviewed the actors about the movie (which was pretty scary and difficult to watch/who were television stars as well as film stars).
3. The student petitioned the professors regarding the course (which was badly overenrolled and needed a bigger lecture hall/who were lecturing quickly and were quite hard to understand).
4. The agent approached the publicists about the photoshoot (which was taking place immediately before the fashion show/who were complaining loudly about the inconsistent lighting).
5. The socialite praised the hostesses for the party (which was held yesterday in a luxurious ballroom/who were greeting guests enthusiastically at the door).
6. The publisher complimented the editors on the magazine (which was read widely by children aged twelve and below/who were in charge of the content but not the advertising).
7. The counselor consoled the students about the competition (which was clearly rigged by the teachers who organized it/who were unfairly excluded from the finals due to poor judging).

8. The principal criticized the instructors for the program (which was run poorly by an inexperienced substitute teacher/who were clearly unqualified to teach specialized topics).
9. The visitor approached the owners about the cat (which was scratching furniture and were meowing loudly and constantly/who were professing skepticism regarding the existence of pet allergies).
10. The producer complimented the directors on the documentary (which was about skateboarding culture in southern California/who were giving speeches at the awards ceremony later on).
11. The officer questioned the guides about the expedition (which was competently led by an eccentric French professor/who were highly experienced at identifying poisonous snakes).
12. The nanny consulted the babysitters about the virus (which was highly virulent and causing concern among the parents/who were getting sick themselves and were caring for the child).
13. The candidate criticized the senators regarding the attack (which was obviously hateful and invoked racial stereotypes/who were opposing reform and were up for re-election).
14. The reporter called the agents about the scandal (which was broadcast continuously on the evening news/who were allegedly spies for a competitor company).
15. The superhero interrogated the henchmen about the plot (which was threatening lives and planned by a supervillain/who were loyal followers of the brilliant supervillain).
16. The colonel praised the captains for the maneuver (which was clearly decisive in the outcome of the battle/who were deeply respected by the rank and file soldiers).
17. The advertiser approached the hosts about the show (which was broadcast Saturdays at midnight for half an hour/who were interviewing pundits and asking controversial questions).
18. The fan petitioned the coaches regarding the strategy (which was completely unsuccessful in nearly every game this season thus far/who were yelling violently at the referees whenever the opposition scored).
19. The stockholder queried the employees about the policy (which was highly controversial and unfair to senior staff/who were immediately fired after complaining in public).
20. The father interrogated the suitors about the venue (which was too informal and was also uncomfortably dark/who were very wealthy and serious about about marriage).
21. The judge queried the lawyers about the evidence (which was tampered with by certain members of the police force/who were persuasively arguing that the defendant was incompetent).
22. The reporter interviewed the pollsters about the election (which was still unpredictable and might require a runoff/who were frequently quoted in all the best-known newspapers).
23. The detective questioned the witnesses about the crime (which was carefully planned by the infamous gang/who were willingly testifying in the court case).
24. The architect consulted the carpenters about the project (which was highly complex but promised to pay a handsome wage/who were clearly enthusiastic and willing to work overtime).

References

Agresti, A. (2002). *Categorical data analysis* (2nd ed.). Wiley.

Aissen, J. (1975). Presentational-*There* insertion: A cyclic root transformation. In *Proceedings of the Chicago Linguistic Society* (vol. 11, pp. 1—14).

Anderson, J. R. (1990). *The adaptive character of human thought*. Lawrence Erlbaum.

Arnold, J. E., Wasow, T., Losongco, A., & Ginstrom, R. (2000). Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering. *Language, 76*(1), 28–55.

Austin, P., & Bresnan, J. (1996). Non-configurationality in Australian aboriginal languages. *Natural Language & Linguistic Theory, 14*, 215–268.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412.

Bach, E., Brown, C., & Marslen-Wilson, W. (1986). Crossed and nested dependencies in German and Dutch: A psycholinguistic study. *Language & Cognitive Processes, 1*(4), 249–262.

Bever, T. G. (1988). The psychological reality of grammar: A student's eye view of cognitive science. In W. Hirst (Ed.), *Giving birth to cognitive science: A Festschrift for George A. Miller*. Cambridge University Press.

Boland, J. E., Tanenhaus, M. K., Garnsey, S. M., & Carlson, G. N. (1995). Verb argument structure in parsing and interpretation: Evidence from *wh*-questions. *Journal of Memory and Language, 34*, 774–806.

Boston, M. F., Hale, J. T., Kliegl, R., Patil, U., & Vasishth, S. (2008). Parsing costs as predictors of reading difficulty: An evaluation using the Potsdam sentence corpus. *Journal of Eye Movement Research, 2*(1), 1–12.

Brants, T. & Franz, A. (2006). Web 1T 5-gram version 1.

Bunt, H. (1996). Formal tools for describing and processing discontinuous constituency. In H. Bunt & A. van Horck (Eds.), *Discontinuous constituency, number 6 in natural language processing* (pp. 63–84). Mouton de Gruyer.

Chomsky, N. (1981). *Lectures on government and binding*. Mouton de Gruyer.

Clifton, C., & Frazier, L. (1989). Comprehending sentences with long distance dependencies. In G. Carlson & M. Tanenhaus (Eds.), *Linguistic structure in language processing* (pp. 273–317). Dordrecht: Kluwer.

Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2001). *Introduction to algorithms* (2nd ed.). MIT Press.

Crocker, M., & Brants, T. (2000). Wide-coverage probabilistic sentence processing. *Journal of Psycholinguistic Research, 29*(6), 647–669.

Culicover, P. W., & Rochemont, M. S. (1990). Extraposition and the complement principle. *Linguistic Inquiry, 21*(1), 23–47.

Culy, C. (1985). The complexity of the vocabulary of Bambara. *Linguistics and Philosophy, 8*, 345–351.

Demberg, V., & Keller, F. (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition, 109*(2), 193–210.

Desmet, T., Brysbaert, M., & de Baecke, C. (2002). The correspondence between sentence production and corpus frequencies in modifier attachment. *The Quarterly Journal of Experimental Psychology, 55A*, 879–896.

Earley, J. (1970). An efficient context-free parsing algorithm. *Communications of the ACM, 13*(2), 94–102.

Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior, 20*, 641–655.

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

Ferrer i Cancho, R. (2006). Why do syntactic links not cross? *Europhysics Letters, 76*(6), 1228–1234.

Fodor, J. A., Bever, T. G., & Garrett, M. F. (1974). *The psychology of language: An introduction to psycholinguistics and generative grammar*. New York: McGraw-Hill.

Francis, E. (2010). Grammatical weight and relative clause extraposition in English. *Cognitive Linguistics, 21*(1), 35–74.

Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading*. London: Erlbaum.

Frazier, L., & Clifton, C. Jr., (1995). *Construal*. MIT Press.

Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition, 6*(4), 291–325.

Gazdar, G., Klein, E., Pullum, G., & Sag, I. (1985). *Generalized phrase structure grammar*. Harvard.

Gibson, E. (1991). A computational theory of human linguistic processing: Memory limitations and processing breakdown. PhD thesis, Carnegie Mellon.

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 A. Marantz, Y. Miyashita, & W. O'Neil (Eds.), *Image language, brain* (pp. 95–126). MIT Press.

Gibson, E. (2006). The interaction of top-down and bottom-up statistics in the resolution of syntactic category ambiguity. *Journal of Memory and Language, 54*, 363–388.

Gibson, T. & Breen, M. (2004). Processing crossed dependencies in English. In *Presented at the 17th CUNY conference on Human Sentence Processing*.

Givón, T. (1993). English grammar: A function-based introduction. In *English grammar: A function-based introduction* (Vol. II). John Benjamins.

Godfrey, J. J., Holliman, E. C., & McDaniel, J. (1992). Switchboard: Telephone speech corpus for research and development. In *Proceedings of the IEEE international conference on acoustics, speech, & signal processing* (pp. 517–520).

Good, P. I. (2004). *Permutation, parametric, and bootstrap tests of hypotheses* (3rd ed.). Springer.

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

Gordon, P. C., Hendrick, R., & Johnson, M. (2004). Effects of noun phrase type on sentence complexity. *Journal of Memory and Language, 51*(1), 97–114.

Grodner, D., & Gibson, E. (2005). Some consequences of the serial nature of linguistic input. *Cognitive Science, 29*(2), 261–290.

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of the second meeting of the North American chapter of the Association for Computational Linguistics* (pp. 159–166).

Hale, J. (2003). The information conveyed by words in sentences. *Journal of Psycholinguistic Research, 32*(2), 101–123.

Hale, J. (2006). Uncertainty about the rest of the sentence. *Cognitive Science, 30*(4), 609–642.

Hale, K. (1983). Warlpiri and the grammar of non-configurational languages. *Natural Language and Linguistic Theory, 1*(1), 5–47.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Brookes Publishing Company.

Hawkins, J. A. (1994). A performance theory of order and constituency. Cambridge.

Hawkins, J. A. (2004). *Efficiency and complexity in grammars*. Oxford University Press.

Huck, G. J., & Na, Y. (1990). Extraposition and focus. *Language, 66*(1), 51–77.

Hudson, R. A. (1984). *Word grammar*. Oxford–New York: Basil Blackwell.

Jaeger, T. F. (2006). Redundancy and syntactic reduction in spontaneous speech. PhD thesis, Stanford University, Stanford, CA.

Jaeger, T. F., Fedorenko, E., Hofmeister, P., & Gibson, E. (2008). Expectation-based syntactic processing: Anti-locality outside of head-final languages. Oral presentation at CUNY 2008.

Johnson, M. (1998). PCFG models of linguistic tree representations. *Computational Linguistics, 24*(4), 613–632.

Joshi, A., Shanker, K. V., & Weir, D. (1991). The convergence of mildly context-sensitive grammar formalisms. In P. Sells, S. Shieber, & T. Wasow (Eds.), *Foundational issues in natural language processing*. MIT Press.

Joshi, A. K. (1985). How much context-sensitivity is necessary for characterizing structural descriptions – Tree Adjoining Grammars. In: D. Dowty, L. Karttunen, & A. Zwicky (Eds.), Natural language processing – Theoretical, computational, and psychological perspectives. Cambridge.

Joshi, A. K., Levy, L. S., & Takahashi, M. (1975). Tree adjunct grammars. *Journal of Computer and System Sciences, 10*(1).

Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science, 20*(2), 137–194.

Kaan, E., & Vasić, N. (2004). Cross-serial dependencies in Dutch: Testing the influence of NP type on processing load. *Memory & Cognition, 32*(2), 175–184.

Kobele, G. M. (2006). Generating copies: An investigation into structural identity in language and grammar. PhD thesis, UCLA.

Konieczny, L. (2000). Locality and parsing complexity. *Journal of Psycholinguistic Research, 29*(6), 627–645.

Konieczny, L., & Bormann, T. 2001. The influence of extraposition on the acceptability of German SC/RC- and RC/SC- double embedded sentences. In *Presented at the 7th Annual Conference on Architectures and Mechanisms for Language Processing (AMLaP)*. Saarbruecken, Germany, 20–22 September.

Kruijff, G. -J. M. & Vasishth, S. (2003). Quantifying word order freedom in natural language: Implications for sentence processing. In *Proceedings of the architectures and mechanisms for language processing conference*.

Kuhlmann, M. & Nivre, J. (2006). Mildly non-projective dependency structures. In *Proceedings of COLING/ACL*.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*(4427), 203–205.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307*, 161–163.

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

Lau, E., Stroud, C., Plesch, S., & Phillips, C. (2006). The role of structural prediction in rapid syntactic analysis. *Brain & Language, 98*, 74–88.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition, 106*, 1126–1177.

Levy, R. & Andrew, G. (2006). Tregex and Tsurgeon: Tools for querying and manipulating tree data structures. In *Proceedings of the 2006 conference on Language Resources and Evaluation*.

Levy, R. & Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. In *Proceedings of the 20th conference on Neural Information Processing Systems (NIPS)*.

Levy, R. & Manning, C. (2004). Deep dependencies from context-free statistical parsers: Correcting the surface dependency approximation. In *Proceedings of ACL*.

Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science, 29*, 1–45.

Lewis, R. L., Vasishth, S., & Van Dyke, J. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Science, 10*(10).

MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review, 109*(1), 35–54.

Manaster Ramer, A. (1995). Book review: The language complexity game. *Computational Linguistics, 21*(1), 124–131.

Marcus, M. P., Santorini, B., & Marcinkiewicz, M. A. (1994). Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics, 19*(2), 313–330.

Marr, D. (1982). *Vision*. San Francisco: Freeman.

McDonald, R., Pereira, F., Ribarov, K., & Hajič, J. (2005). Non-projective dependency parsing using spanning tree algorithms. In *Proceedings of ACL*.

McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language, 38*(3), 283–312.

Mehl, M. R., Vazire, S., Ramírez-Esparza, N., Slatcher, R. B., & Pennebaker, J. W. (2007). Are women really more talkative than men? *Science, 317*(5834), 82.

Mel'cuk, I. (1988). *Dependency syntax: Theory and practice*. SUNY Press.

Menard, S. (1995). *Applied logistic regression analysis*. Sage.

Miller, G. A. (1962). Some psychological studies of grammar. *American Psychologist, 17*, 748–762.

Miller, P. (2000). Strong generative capacity: The semantics of linguistic formalism. Cambridge.

Mitchell, D., & Brysbaert, M. (1998). Challenges to recent theories of language differences in parsing: Evidence from Dutch. In D. Hillert (Ed.), *Sentence processing: A crosslinguistic perspective*. *Syntax and semantics* (Vol. 31, pp. 313–336). Academic Press.

Mitchell, D. C. (1994). Sentence parsing. In M. Gernsbacher (Ed.), *Handbook of psycholinguistics*. New York: Academic Press.

Mitchell, D. C., Cuetos, F., Corley, M., & 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.

Narayanan, S. & Jurafsky, D. (1998). Bayesian models of human sentence processing. In *Proceedings of the twelfth annual meeting of the Cognitive Science Society*.

Narayanan, S. & Jurafsky, D. (2002). A Bayesian model predicts human parse preference and reading time in sentence processing. In *Advances in neural information processing systems* (vol. 14, pp. 59–65).
Nederhof, M.-J. (1999). The computational complexity of the correct-prefix property for TAGs. *Computational Linguistics, 25*(3), 345–360.
Nivre, J. & Nilsson, J. (1999). Pseudo-projective dependency parsing. In *Proceedings of the annual meeting of the Association for Computational Linguistics* (Vol. 43, pp. 99–106).
Pollard, C. (1984). Generalized phrase structure grammars, Head grammars, and natural languages. PhD thesis, Stanford.
Pollard, C., & Sag, I. (1994). *Head-driven phrase structure grammar*. Chicago, Stanford: University of Chicago Press, CSLI Publications.
Resnik, P. (1992). Probabilistic Tree-Adjoining Grammar as a framework for statistical natural language processing. In *Proceedings of ACL*.
Ristad, E. S. (1993). *The language complexity game*. Cambridge, MA: MIT Press.
Roark, B., Bachrach, A., Cardenas, C., & Pallier, C. (2009). Deriving lexical and syntactic expectation-based measures for psycholinguistic modeling via incremental top-down parsing. In *Proceedings of EMNLP*.
Rochemont, M. S., & Culicover, P. W. (1990). *English focus constructions and the theory of grammar. Cambridge studies in linguistics* (Vol. 63). Cambridge University Press.
Rohde, D. (2005). *Linger: A flexible platform for language processing experiments*, version 2.88. Available from: http://tedlab.mit.edu/~dr/Linger.
Roy, B. C., Frank, M. C., & Roy, D. (2009). Exploring word learning in a high-density longitudinal corpus. In *Proceedings of the 31st annual meeting of the Cognitive Science Society*.
Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science, 237*(4820), 1317–1323.
Shieber, S. M. (1985). Evidence against the context-freeness of natural language. *Linguistics and Philosophy, 8*, 333–343.
Simpson, J. (1983). Aspects of Walrpiri morphology and syntax. PhD thesis, MIT.
Slobin, D. I. (1966). Grammatical transformations and sentence comprehension in childhood and adulthood. *Journal of Verbal Learning and Verbal Behavior, 5*, 219–227.
Smith, N. J. & Levy, R. (2008). Optimal processing times in reading: A formal model and empirical investigation. In *Proceedings of the 30th annual meeting of the Cognitive Science Society*.
Spivey, M. J., & Tanenhaus, M. K. (1998). Syntactic ambiguity resolution in discourse: Modeling the effects of referential content and lexical frequency. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 24*(6), 1521–1543.
Spivey-Knowlton, M. (1996). Integration of linguistic and visual information: Human data and model simulations. PhD thesis, University of Rochester.
Stabler, E. P. (1997). Derivational minimalism. In C. Retoré (Ed.), *Logical aspects of computational linguistics* (pp. 68–95). Springer.
Staub, A., & Clifton, C. (2006). Syntactic prediction in language comprehension: Evidence from either ... or. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 32*(2), 425–436.
Staub, A., Clifton, C., Jr., & Frazier, L. (2006). Heavy NP shift is the parser's last resort: Evidence from eye movement. *Journal of Memory and Language, 54*(3), 389–406.
Steedman, M. (2000). *The syntactic process*. Cambridge, MA: MIT Press.
Szmrecsanyi, B. (2004). On operationalizing syntactic complexity. In G. Purnelle, C. Fairon, & A. Dister (Eds.), *Proceedings of the 7th international conference on textual data statistical analysis* (pp. 1032–1039).
Tanenhaus, M. K., & Trueswell, J. C. (1995). Sentence comprehension. In J. L. Miller & P. D. Eimas (Eds.), *Speech, language, and communication* (pp. 217–262). Academic Press.
Traxler, M. J., & Pickering, M. J. (1996). Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language, 35*, 454–475.
Uszkoreit, H., Brants, T., Duchier, D., Krenn, B., Konieczny, L., Oepen, S., et al. (1998). Studien zur performzorientieren Linguistik: Aspekte der Relativsatzextraposition im Deutschen. *Kognitionswissenschaft, 7*, 129–133.
Valiant, L. G. (1975). General context-free recognition in less than cubic time. *Journal of Computer and System Sciences, 10*, 308–315.
Vijay-Shanker, K., & Joshi, A. K. (1985). Some computational properties of Tree Adjoining Grammars. In *Proceedings of ACL*.
von Fintel, K. (1994). Restrictions on quantifier domains. PhD thesis, University of Massachusetts.
Wasow, T. (1997). Remarks on grammatical weight. *Language Variation and Change, 9*, 81–105.
Wasow, T. (2002). *Postverbal behavior*. CSLI.
Wasow, T., Jaeger, T. F., & Orr, D. (2006). Lexical variation in relativizer frequency. In H. Wiese & H. Simon (Eds.), Proceedings of the workshop on expecting the unexpected: Exceptions in grammar at the 27th annual meeting of the German Linguistic Association, University of Cologne, Germany. DGfS.
Yngve, V. (1960). A model and an hypothesis for language structure. In *Proceedings of the American Philosophical Society* (pp. 444–466).
Younger, D. H. (1967). Recognition and parsing of context-free languages in time n^3. *Information and Control, 10*(2), 189–208.