The Processing of Extraposed Structures in English

Roger Levy[a], Evelina Fedorenko[b], Mara Breen[c], Ted 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 at Amherst, Tobin Hall, 135 Hicks Way, Amherst, MA 01003, USA*

Abstract

In most languages, most of the syntactic dependency relations found in any given sentence are *projective*: there are no crossings between word-word dependencies in the sentence. Some syntactic dependency relations, however, are *non-projective*, involving word-word dependencies that cross each other. Crossing 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 crossing dependencies, and whether factors known to influence processing of conventional dependencies also affect crossing-dependency processing. We report four self-paced reading studies, together with corpus and completion studies, investigating the comprehension difficulty associated with the crossing 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 reverse the difficulty associated with extraposition. Our experiments rule out accounts based purely on derivational complexity and/or dependency locality in terms of linear positioning. This is the first demonstration that comprehenders maintain probabilistic syntactic expectations that persist beyond projective-dependency structures, and suggests that it may be possible to explain observed patterns of comprehension difficulty associated with extraposition entirely through probabilistic expectations.

Keywords: Sentence comprehension, Syntactic Complexity, Parsing, Word Order, Memory and Language, Self-paced Reading, Frequency, Prediction

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.[1] For example, in the sentence

[1]Portions of this work have benefited from feedback from presentations at the 2004 and 2008 CUNY Sentence Processing Conferences, and presentation in a 2009 colloquium at the Department of Linguistics at

(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 dependency relationships is projective if no two dependencies 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). The sentence in (2b), in contrast, is non-projective: the dependency between *Yesterday* and *arrived* crosses the dependency between *woman* and *who*.

(2) a. Yesterday a woman who I knew arrived.

b. Yesterday a woman arrived who I knew.

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 a crossing dependency even if *Yesterday* were omitted.

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). There are several formal means of representing non-projective dependency in phrase-structure trees; Figure 1 illustrates three alternatives,

UC San Diego. We are grateful to Giulia Pancani for assistance in coding the sentence-completion data of Experiment 4. Mike Frank provided useful suggestions and discussion regarding the estimation of a college-age student's lifetime linguistic exposure (Footnote 8). We would also like to thank David Beaver, Brady Clark, Timothy Desmet, Gabriel Doyle, Janet Fodor, Mike Frank, Colin Phillips, Hannah Rohde, Carson Schutze, Tom Wasow and Florian Wolf for valuable comments and feedback; and Jarrod Hadfield, Luca Borger, and David Duffy for suggestions regarding data analysis. All errors are our own responsibility.

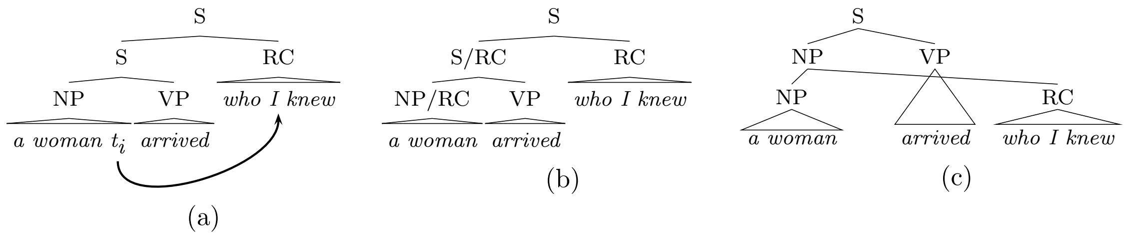


Figure 1: Different phrase-structure representations for the non-projective dependency structure in (2b). In 1a, the non-projective dependency is characterized by movement from an NP-internal trace. In 1b, the non-projective dependency is Represented by the missing-RC information transmitted between the NP and the top S categories. In 1c, non-projectivity is directly represented as a discontinuous constituent.

including movement with traces as in Government and Binding (Chomsky, 1981), slash-passing as in Generalized Phrase Structure Grammar (Gazdar, Klein, Pullum & Sag, 1985), and discontinuous-constituency derivation trees as in Tree Adjoining Grammars (Joshi, Levy & Takahashi, 1975), Head Grammars (Pollard, 1984), and Discontinuous Phrase Structure Grammar (Bunt, 1996).

This paper reports an investigation into the online comprehension of crossing dependencies, motivated on the one hand by their formal complexity and on the other by their rarity—both within and among languages. The remainder of the introduction briefly discusses the cross-linguistic distribution and computational complexity of non-projectivity, and frames the major questions to be addressed in the remainder of the paper.

1.1. Non-projectivity cross-linguistically

Although it is clear from examples such as (2b) that non-projective dependency structures do arise in natural language, non-projectivity is the exception rather than the norm. In English, for example, the only sources of non-projectivity are right extraposition (of relative clauses such as in (2b), prepositional phrases, and comparative-construction complements) and unbounded extraction (*wh-* question formation, topicalization, and relativization). Cross-linguistically, although these and other grammatical constructions (for example, complex verb phrases in Dutch, as examined by Bach, Brown & Marslen-Wilson, 1986 and others) are available to give rise to crossing dependency, their distribution is nevertheless much more limited than that of projective-dependency constructions. The best-known case study may be that of relativization: as the work of Keenan & Comrie (1977) and Hawkins (1999) demonstrates, the types of relativizations that would give rise to crossing dependencies are less common than the simplest form, subject relativizations, in which dependencies typically do not cross.

A brief discussion is also in order here regarding the numerous languages such as Latin, Russian, and Finnish that have what is called "free word order". Although in these languages the relative positioning of words standing in a dependency relation is flexible, for the most part this flexibility is restricted to orderings that maintain projective dependencies. In Russian, for example, the simple subject-verb-object (SVO) transitive sentence in (3) has an adjective premodifying each of the subject and object noun; morphological case marking on the adjectives and nouns unambiguously determines the word-word dependency relations and the grammatical functions of each word in the sentence. Any other permutation of the major constituents of the sentence, as shown in (4), can also be acceptable given appropriate context (Krylova & Khavronina, 1988, inter alia).

(3) **Staryj professor** uvidel *molodogo studenta*.
old.NOM professor.NOM noticed young.ACC student.ACC
"The old professor noticed the young student."

(4)
a. **Staryj professor** *molodogo studenta* uvidel.
b. Uvidel **staryj professor** *molodogo studenta*.
c. Uvidel *molodogo studenta* **staryj professor**.
d. *Molodogo studenta* **staryj professor** uvidel.
e. *Molodogo studenta* uvidel **staryj professor**.

But this word-order flexibility is limited to the major constituents of the sentence, and does not create crossing dependencies. For example, strong constraints exist against arbitrarily separating adjectives from the nouns they modify, as illustrated by the infelicity of Example (5) below:

(5) **Molodogo* **staryj** uvidel *studenta* **professor**.
young.ACC professor.NOM noticed old.NOM student.ACC

This type of separation can create crossing dependencies; in (5), for example, the dependency between *staryj* and *professor* crosses the dependency between *uvidel* and *studenta*. Word-order flexibility that creates felicitous crossing dependency is less widespread than major-constituent word order freedom, and it is only in a small number of languages such as Warlpiri (Hale, 1983) that something approaching complete word-order freedom seems to exist. In the languages for which it has been possible to quantify the frequency of crossing dependency, it has been clear that projective dependencies are far more common (Kruijff & Vasishth, 2003; Levy & Manning, 2004). Even for Warlpiri, the absence to date of quantitative corpus analysis means that it is not clear how common crossing dependencies truly are.

1.2. Computational complexity of non-projectivity

In addition to their rarity, non-projective dependencies are more complex than projective dependencies to compute. 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 (Younger, 1967; Earley, 1970).[2] In order to parse sentences using grammars allowing

[2]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.

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 (Vijay-Shanker & Joshi, 1985; Nederhof, 1999).[3,4]

1.3. Sources of the rarity of non-projectivity

The foregoing discussion gives rise to the question of why non-projective dependency structures are rare in the first place. Let us consider a number of logical possibilities:

1. **Acquisition bias:** Humans may possess an innate acquisition bias (part of "universal grammar") toward learning grammars in which non-projectivity is dispreferred or disallowed;[5]
2. **Production constraints:** The pressures of real-time language production may tend to disfavor discontinuous realization of constituents (Gennari & MacDonald, 2008, 2009);
3. **Information-structure preference:** It is well known that word order can carry information-structural content and it may be that the information-structural circumstances that make non-projective linearizations desirable are rare;
4. **Comprehension cost:** If non-projectivity carries with it some inherent processing cost in language comprehension, this processing cost could tend to make non-projectivity rare, via audience design (Clark & Murphy, 1982) or transmission bias (Kirby, 1999).

No subset of the possibilities outlined above are mutually exclusive, and there is merit in investigating each possibility in its own right, using appropriate methodology. In this paper,

[3]Some types of nonprojectivity 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.

[4]McDonald, Pereira, Ribarov & 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 types of crossing dependency that may be introduced, 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 crossing dependency, those characterized by the *mildly context-sensitive* class of formal languages (Shieber, 1985; Culy, 1985; Joshi, Shanker & Weir, 1991)—those characterized by Tree-Adjoining Grammar (Joshi et al., 1975), Head Grammar (Pollard, 1984), and Combinatory Categorial Grammar (Steedman, 2000)—though see Kobele (2006) for more recent developments.

[5]This possibility would entail that grammars can include not only specifications of what structures are categorically allowed and disallowed (Chomsky, 1965), but also specification of relative preferences for different structural forms, since it is clear that non-projectivity is dispreferred in some cases more strongly than others. Since work in recent years has made it clear that such grammars encoding gradient preference can be formally described using probabilistic models (e.g., Bod, 1998; Boersma & Hayes, 2001) and innate biases can be formally stated over these models as a part of Bayesian statistical inference (e.g., Goldsmith, 2001; Perfors, Tenenbaum & Regier, 2006; Goldwater, Griffiths & Johnson, 2009), we find no difficulty with this logical possibility.

we focus on the last possibility, that of *comprehension cost*. In Section 2 we describe our research strategy for investigating the online comprehension of non-projective dependencies, and why—and under what conditions—one might expect comprehension costs to arise in the processing of non-projective dependency, based on leading theories of online syntactic comprehension. Sections 1 through 6 present experiments designed to tease apart the predictions of different theories; in the process, we gain new insights into online syntactic comprehension that are valuable in their own right. Section 7 concludes.

2. Online comprehension of non-projective structure

Both their rarity and computational complexity raise the question of whether crossing dependencies pose any special burden in human sentence comprehension. If they do pose such a burden, a question of causal direction follows: are they difficult because they are rare, or are they rare because they are difficult? Despite the prominence of crossing-dependency constructions in theoretical syntax, relatively little work has addressed these questions. Perhaps the best-known study is Bach et al. (1986)'s comparative study of crossing versus nested dependencies in German versus Dutch, 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 cannot be generalized and do not definitively address issues of processing cost. 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. As a result, many important questions regarding the processing of crossing-dependency structures remain to be addressed. In particular, if it were to turn out that structural properties of non-projective dependency posed special difficulties for the human sentence processor that could not be reduced to known effects of construction frequency, this finding could serve as a crucial explanatory link (in the spirit of Hawkins, 1994, 2004) between formal treatments of non-projectivity and generalizations regarding the restricted distribution of non-projective structures within and across languages.

We thus focus on the following questions regarding the online comprehension of crossing-dependency structures:

1. Are crossing-dependency structures any harder to comprehend than projective-dependency structures?
2. If so, under what conditions?
3. How can these differences in comprehension be understood with respect to existing theories of online comprehension?
4. What are the further implications of the answers to Questions 1–3 for the study of crossing dependencies in natural language?

The research strategy we adopt to address these questions is as follows. In general, we are interested in comparing the processing cost of projective-dependency structures with the cost of non-projective dependency structures of the same type, controlling for factors other than projectivity that may independently influence processing cost. We follow the tradition in experimental psycholinguistics of using *reading times* as a proxy measure of processing costs (e.g., Thibadeau, Just & Carpenter, 1982; Mitchell, 1984; Grodner & Gibson, 2005, among many others). On the hypothesis that the construction of syntactic relations is incremental (supported by a wide range of existing evidence and also by the experiments presented here), we can expect most processing costs to be measurable within the first few words of the establishment of a (projective or non-projective) dependency relationship. We can thus compare reading times upon and immediately after completion of projective versus non-projective dependency relations to draw inferences about their relative processing costs. For reasons of methodological convenience we focus here on the processing of non-projective dependency relations in English, though we emphasize that achieving a wider cross-linguistic picture would be of great interest.

Given this research strategy, focusing on right extraposition turns out to have several advantages. Unlike the situation with many filler-gap dependencies in English, relative clause 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, it 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. To set the stage for our answers to the above questions, the next section outline the predictions made by existing theories for the processing of extraposed relative clauses.

2.1. 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 (Frazier & Fodor, 1978; Frazier, 1987; Clifton & Frazier, 1989), 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 (Jurafsky, 1996; Narayanan & Jurafsky, 1998; Crocker & Brants, 2000; Narayanan & Jurafsky, 2002). Because right-extraposition does not necessarily involve any local structural ambiguity, as is the case in (2b), if locally unambiguous cases could be shown to induce processing difficulty, then theories that exclusively cover ambiguity resolution would not be sufficient to capture constraints involved in crossing-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., Gibson, 1991)—that do 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.

2.1.1. Derivational complexity

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 Figure 1a: namely, that right-extraposed RCs (e.g., (2b)) will be more difficult to process than their corresponding unextraposed variants (e.g., (2a)).

2.1.2. 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). On 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.

2.1.3. Probabilistic expectations

There are several theories that predict differential difficulty in unambiguous contexts on the basis of probabilistic expectations, including surprisal (Hale, 2001; Demberg & Keller, 2008; Levy, 2008; 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

	$P(w_i \mid w_{i-1})$	$P(w_i \mid c_{i-1})$	$P(RC \mid \text{syntactic context})$
(2a)	0.07	0.0044	0.00565
(2b)	0	0.0003	0.00008

Table 1: Collocational and syntactic conditional probabilities of a relative clauses in Example (2). c_{i-1} denotes the syntactic part of speech of the previous word. Probabilities are relative-frequency estimates.

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:

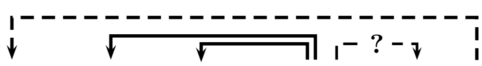
- **Collocational expectations.** Comprehenders may attend to a limited number of words immediately preceding a given word in order to formulate expectations. As an extreme case, in (2), the probability of the relative pronoun *who* when it appears might be conditioned on only the previous word or the previous word class. On the assumption that expectations are set *rationally* (Marr, 1982; Shepard, 1987; Anderson, 1990) 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 (2a), the expectation for *who* should be identified with the probability that *woman* will be immediately post-modified 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 $P(\text{RC} \mid \text{context})$, 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.

Neither collocational-expectation nor structural-expectation-based approaches as described thus far would provide an explanatory link between comprehension cost and the rarity of non-projective dependencies. However, one might suppose that the formal complexity of non-projective dependency structures might make it difficult for comprehenders to deploy expectations effectively when processing crossing dependencies. To make this argument clear, let us consider the situation in incremental comprehension in Example (6) below, 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.

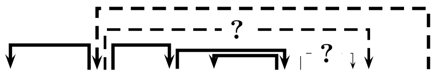
(6) Yesterday a student . . .

Now consider the case where some already-seen property of the subject NP creates a stronger—yet non-categorical—expectation for a postmodifying RC. For example, Wasow, Jaeger & Orr (2006) and Jaeger (2006) demonstrated that definite-superlative NPs are more likely to have postmodifying RCs than indefinite NPs without adjectives. In Example (7) below, the definite-superlative content would thus create a stronger expectation for an upcoming RC, denoted by the thicker question-marked dotted line originating from *student*.



(7) Yesterday the smartest student . . .

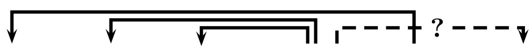
On an expectation-based theory, an immediately following RC would thus be easier to process in (7) than in (6). One way in which these contrasting expectation effects have been shown to come out especially clearly is in RC attachment preferences. For example, Desmet, Brysbaert & de Baecke (2002) showed that in Dutch, human NPs are more frequently postmodified by RCs than non-human RCs are. A complex NP onset, such as *the pioneer of the music. . .* in Example (8) below, would thus have differing strengths of expectation for RC postmodification at the two levels.



(8) The pioneer of the music ...

On expectation-based syntactic-comprehension theories, the relatively stronger expectation for high attachment than for low attachment predicts a processing benefit accrued for comprehension of a high-attaching RC in this case, an effect that was documented by Mitchell & Brysbaert (1998).

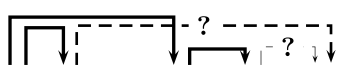
Now let us consider a case similar to Example (7), but in which the strongly expected potential postmodifying RC would be extraposed and thus cross another dependency, as in Example (9) below.



(9) Yesterday the smartest student arrived ...

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 nonprojectivity. If this were true, then in the cases where *arrived* was indeed followed by an RC (and these cases would have to be common on the previously mentioned assumption that expectation-setting is rational), 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. Such an asymmetry in processing difficulty, combined with factors of audience design and/or transmission bias (Section 1.3), could lead to extraposed RCs being rare.

It is thus of considerable interest to determine whether comprehenders can deploy strong syntactic expectations to facilitate processing even when crossing 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 (10) below, in which some property of a postverbal NP_1 sets up a strong expectation for a postmodifier that remains unmet during the processing of a preposition which modifies the verb and governs a following NP_2:



(10) ... V NP_1 ... Prep NP_2 [RC]

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 can be effectively deployed across crossing dependencies. We return to this prediction in Section 6.

2.2. 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 four 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. Experiments 2 and 3 provide 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 4 provides evidence distinguishing between predictions of structural expectations versus those of derivational complexity. We revisit the implications of these experimental results for both our psycholinguistic and typological questions in the General Discussion.

3. 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 (11) below:

(11) 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 (11b) and (11d) than in (11a) and (11c). 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.

3.1. 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.

3.2. Materials

Twenty-four items (listed in full in Appendix B) were constructed following the pattern of (11) above. Each item was initiated by a prepositional phrase that established a context for the sentence (e.g., *After the show* in (11)). The subject NP of the sentence occurred next, consisting of an indefinite determiner (*a/an*) and an occupation noun, such as *performer* in (11). 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 (e.g., *came on* in (11c)), 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 (11)). 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.

3.2.1. 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 $n \leq 5$ 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

[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 (11d)). 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.

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 $p < 0.01$.

3.3. 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, 2001). 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, there were four items (11, 13, 16, and 24) with two-word verb phrases in the non-presentative conditions which were presented as two separate single-word groups; the analyses presented in Section 3.4 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 minutes per participant to complete.

3.4. Results

3.4.1. Statistical analysis

Reading times were analyzed in each region as follows, unless otherwise specified. Measurements above 1500ms 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 measured. The remaining measurements were then analyzed in 2×2 by-participants and by-items ANOVAs. In cases where a single region of 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.

3.4.2. 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.

[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.

	Extraposed	Unextraposed
Non-presentative	0.89	0.88
Presentative	0.93	0.93

Table 2: Question-answering accuracy in Experiment 1

	F1	F2
Presentative	5.80*	1.59
Extraposition	<1	<1
Pres×Extrap	<1	<1

Table 3: F-statistics for analysis of question-answering accuracy in Experiment 1 ($^{*}p < 0.05$)

	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

Table 4: F-statistics for Experiment 1 ($^{\cdot}p < 0.1, ^{*}p < 0.05$). In regions not listed, there were no effects were below $p = 0.1$.

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.

3.4.3. Reading times

We divided the sentence into nine regions of analysis as depicted in (12) below:

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

Figure 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. The crucial result for our study is evident across the first three words of the RC, where 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 a pattern toward 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 through 7). We first computed a linear regression of RT against region length (as measured in characters) for each participant, following Ferreira & Clifton (1986), 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 6000ms, and analyzed residual

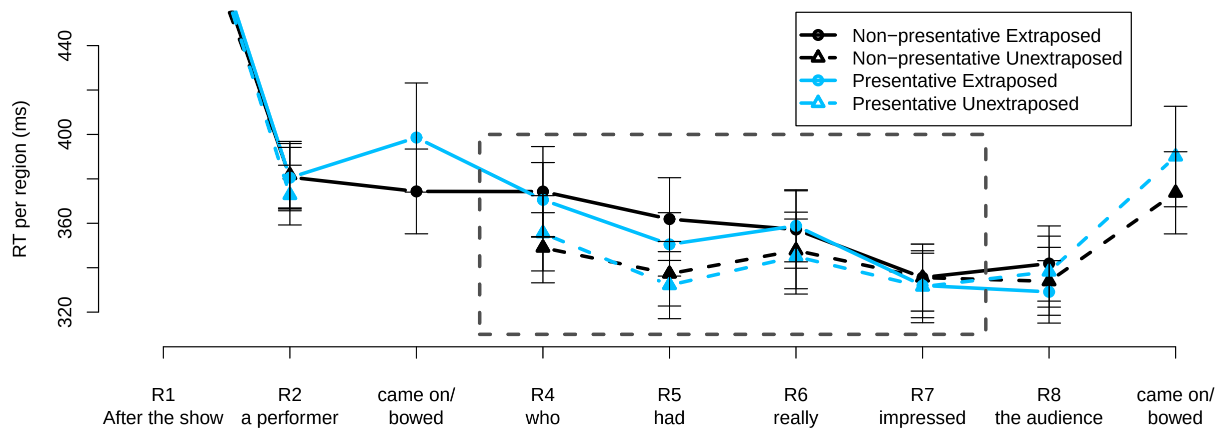


Figure 2: Reading-time results as a function of region and condition for Experiment 1. Onset of the relative clause (first four words) is boxed.

RTs on the remaining trials as described in Section 3.4.1. Figure 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-verb condition than in the non-presentative condition, this interaction was far from significant.

3.5. Discussion

The results of Experiment 1 demonstrate that there is online processing 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 relative clause; by the fourth word of the relative clause, 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 (Figure 3) was consistent with the hypothesis, and the lack of significance could derive from the weakness of the

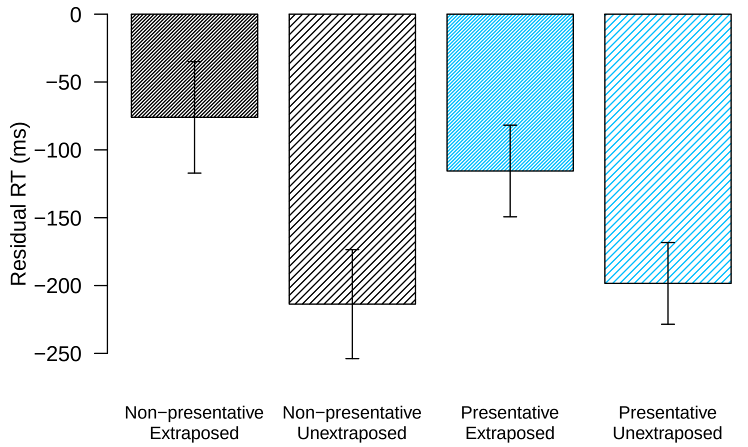


Figure 3: Residual reading times over the first four words of the RC in Experiment 1

	F1	F2
Presentative	<1	<1
Extraposition	7.29*	6.23*
Pres×Extrap	<1	<1

Table 5: F-statistics for residual reading-time analysis across first four words of the RC in Experiment 1 ($^*p < 0.05$).

expectation manipulation and a corresponding lack of statistical power.

Having demonstrated that RC extraposition can be associated with processing difficulty, we now turn to addressing the possible sources of 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.

4. Experiment 2

This experiment is designed to distinguish between derivational complexity or structural expectations on the one hand versus decay, interference, or collocational expectations on the other. We achieve 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 involve an RC-preceding context of the form

(13) 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 Noun_1 (NP-attached) or to the verb (VP-attached)—with the adjacency of the RC attachment—to Noun_2 (adjacent) versus to Noun_1 (non-adjacent). The four resulting conditions are illustrated in (14) below:

(14) a. The chairman consulted the executive of the companies that were making lots of money. [NP-attached PP, RC adjacent]

Condition	RC conditional probability	n
NP-attached PP, RC adjacent	3.96%	2552
NP-attached PP, RC non-adjacent	3.78%	30127
VP-attached PP, RC adjacent	3.32%	6576
VP-attached PP, RC non-adjacent	0.35%	1139

Table 6: Syntactic conditional probabilities of RCs for the four conditions of Experiment 2

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]

Only the VP-attached PP, RC non-adjacent condition in (14d) 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 (crossing 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 (14), 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 4.5.

4.1. Participants

Forty-two 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.

4.2. Materials

Twenty-four items (listed in full in Appendix C) were constructed following the pattern of (14) above. Each item consisted of a sentence-initial subject (determiner plus noun) followed by a word sequence with parts of speech as shown in (13) above, then the word *that* followed by the word *was* or *were* (depending on condition) and finally four or five more words to complete the relative clause and the sentence. In every sentence, one of the two post-verbal main-clause nouns was plural and the other was singular; plural/singular ordering was balanced across items. In addition to the target sentences, 96 filler sentences

	RC adjacent	RC non-adjacent
VP-attached	0.85	0.54
NP-attached	0.81	0.63

Table 7: Question-answering accuracy in Experiment 2

	F1	F2
PP attachment	1.22	<1
RC adjacency	$97.95^{\ddagger}$	$18.82^{\ddagger}$
PP×Adjacency	$8.78^{\dagger}$	$4.27^{\cdot}$

Table 8: F-statistics for analysis of question-answering accuracy in Experiment 2 ($^{\cdot}p < 0.1$, $^{\dagger}p < 0.01$, $^{\ddagger}p < 0.001$)

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.

4.3. 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 45 minutes per participant to complete.

4.4. Results

4.4.1. Statistical analysis

Statistical analysis procedures were identical to those in Experiment 1. Outlier removal procedures led to 2.1% of reading-time data being discarded.

4.4.2. Comprehension Accuracy

Overall question-answering accuracy on experimental items was 71%, demonstrating a greater than average difficulty of these sentences than of those in Experiment 1. Tables 7 and 8 show question-answering accuracy by condition, together with the results of 2 × 2 ANOVAs. There was a significant interaction between PP attachment and RC adjacency, with question answering in the extraposed-RC condition (VP/non-adjacent) lowest at 54%.

4.4.3. Reading times

For analysis purposes, we divided the materials into nine regions, as shown in (15) below: (1) the subject NP, consisting of *the* plus a noun; (2) the verb, (3) the determiner (always *the*) and noun of the direct object; (4) a prepositional phrase, consisting of *of/about the* plus a noun; (5-8) each of the first four words of the relative clause (as in Experiment 1); and (9) the rest of the words in the sentence:

(15) The chairman | consulted the executives | of/about the company | that | was/were | making | lots | of money.

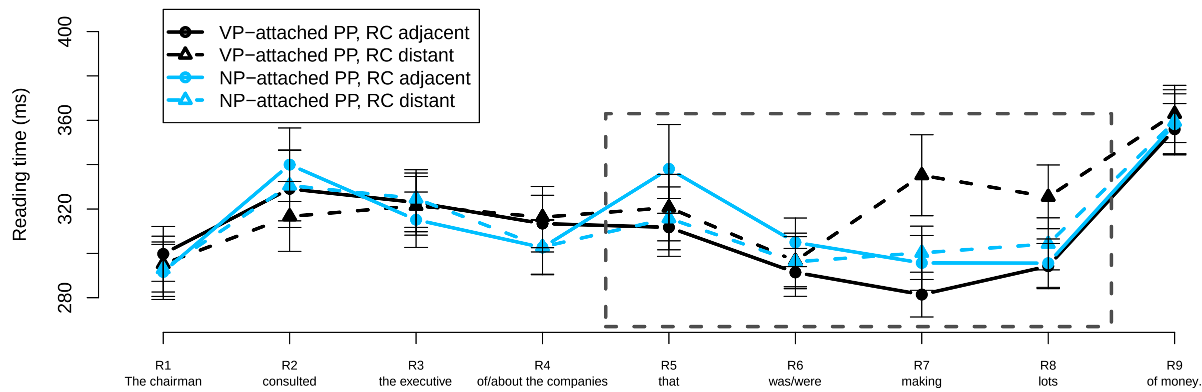


Figure 4: Reading-time results as a function of region and condition for Experiment 2. Onset of the relative clause (first four words) is boxed.

Figure 4 shows region-by-region reading times, and Table 9 gives the results of 2×2 ANOVA analyses by participants and by items for each region. Before the relative clause, at Region 2, we find a main effect of PP attachment marginal by participants, presumably due to chance since readers have not yet encountered the PP. At Region 4, we find a significant main effect of PP attachment, with faster reading times in the NP-attached condition than in the VP-attached position. This is likely due to differences in word length and frequency between the prepositions *of* and *about*. The crucial result of this study emerges inside the relative clause at Region 7, where we find a significant main effect of RC adjacency, with non-adjacent RCs read more slowly, driven primarily by a significant interaction such that the extraposed-RC condition (VP/non-adjacent) is read most slowly. At Region 8 we find a similar pattern, albeit with the interaction significant only by participants, plus a main effect of PP attachment driven by the same interaction.

As with Experiment 1, we conducted a residual reading-time analysis across the first four regions of the RC (Regions 5 through 8), using the same methodology as described in Section 3.4.3. Figure 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 a significant interaction between the attachment site of the PP and adjacency of the RC, with reading times highest in the extraposed condition (VP/non-adjacent). Pairwise comparisons showed significant effects of PP attachment within the RC non-adjacent conditions ($F1(1,41) = 12.58, p < .001; F2(1,23) = 12.17, p = 0.002$) and of RC adjacency within the VP-attached PP conditions ($F1(1,41) = 16.08, p < .001; F2(1,23) = 36.55, p < .001$).

4.5. Discussion

In reading times at the third and fourth words of the relative clause, as well as in residual times as a whole over the first four words of the RC, we see an interaction between PP attachment and RC adjacency (albeit significant only by participants at the third word), with reading times superadditively highest in the extraposed-RC condition. This is the pattern

	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	2.17	<1	$3.29^{\cdot}$	1.92	<1	<1	6.44^{*}	7.72^{*}	1.31	<1	1.45	1.05	2.11	1.20	2.65	$2.98^{\cdot}$	<1	<1
RC adjacency	<1	<1	2.36	1.29	<1	<1	<1	<1	<1	<1	<1	<1	$19.75^{\ddagger}$	$12.16^{\dagger}$	$11.79^{\dagger}$	7.53^{*}	<1	<1
PP×Adjacency	<1	<1	<1	<1	1.29	<1	<1	<1	$3.25^{\cdot}$	2.37	1.28	1.65	$8.72^{\dagger}$	$8.52^{\dagger}$	4.67^{*}	2.81	<1	<1

Table 9: F-statistics for Experiment 2 ($^{\cdot}p < 0.1, ^{*}p < 0.05, ^{\dagger}p < 0.01, ^{\ddagger}p < 0.001$). In regions not listed, there were no effects below $p = 0.1$.

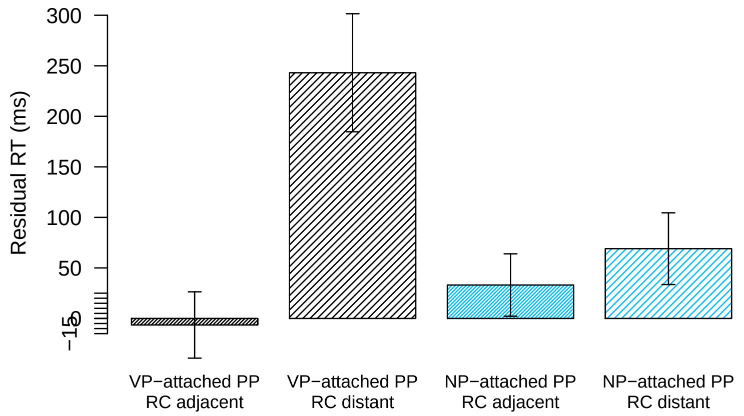


Figure 5: Residual reading times over the first four words of the RC in Experiment 2

	F1	F2
PP attachment	6.62^{*}	4.82^{*}
RC adjacency	$12.99^{\ddagger}$	$18.52^{\ddagger}$
PP×Adjacency	$12.31^{\dagger}$	$12.72^{\dagger}$

Table 10: F-statistics for residual reading-time analysis across first four words of the RC in Experiment 2 ($^{*}p < 0.05, ^{\dagger}p < 0.01, ^{\ddagger}p < 0.001$).

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.

This pattern is also problematic for theories of pure collocational expectations simply on the basis of the complexity of the collocation that would be required. The minimal-length collocation that would be required to capture the four-way distinction among conditions would be a 5-gram—in (14), for example, *of/about the companies that was/were*. 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 Google n-grams corpus (Brants & Franz, 2006) for the frequencies of the relevant collocations in our materials. As an example of the sparsity of such collocational data, the 5-gram *of the companies that were* has a count of 2,098 in this dataset; this was the item/condition combination with the highest count. 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 this collocation would be 0.7.[8]

[8]This estimate can be obtained in a number of ways. Roy, Frank & Roy (2009) collected 4260 hours

The remaining 5-grams in question occur considerably less frequently—the arithmetic mean over the non-extraposed conditions is 200.4—so that the average experimental participant is unlikely to have had any direct experience with most of the relevant collocations in their lifetime. As demonstrated in Table 6, however, the structural configurations in question are relatively common, and predict the interactive pattern cleanly.

5. Experiment 3

Although Experiment 2 supported the interpretation of Experiment 1 as indicating a processing cost associated with RC extraposition measurable in terms of both question-answering accuracy and on-line reading times, the use of number agreement as the sole disambiguating cue raises possible methodological concerns. Some experimental participants in Experiment 2 stated that there were noun-verb agreement errors in the items that they had read, although no experimental sentences (including fillers) had actual agreement errors on the intended syntactic analysis. The low question-answering accuracy in the non-local conditions of Experiment 2 (see Table 7), in particular in the extraposed condition, highlights these concerns. The high reading times observed in the extraposed condition might thus be a symptom of perceived ungrammaticality rather than of elevated processing difficulty in the construction of the intended grammatical analysis. To address this issue, we devised a new set of materials with syntactic structures similar to those used in Experiment 2, but whose disambiguating cues were based on animacy and plausibility, rather than number agreement. These new materials are exemplified in (16) below:

(16)
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]

In (16a) and (16c), the relative clause *which was filmed in the jungles of Vietnam* is most plausible as modifying the noun *movie*. In (16b) and (16d), in contrast, the relative clause

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 & 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 & 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.

who was married to the famous model is most plausible as modifying the noun *star*. Furthermore, the relative pronoun *who* in (16b) and (16d) is a strong cue toward non-adjacent attachment to *star* (cf. Experiment 2 in which the relative pronoun *that* leaves RC attachment ambiguous). As in Experiment 2, if the cause of the difficulty of processing the VP-attached, non-adjacent condition is due to the presence of extraposition (crossing dependency), then there should be super-additively high reading times during the relative clause in (16d) compared to the other three conditions. If animacy- and plausibility-based cues toward correct RC attachment are more reliable than the number-agreement cues used in Experiment 2, we might also expect to find that question answering accuracy is higher overall in this experiment, and that it differs less across conditions.

5.1. 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.

5.2. Materials

Twenty-four items (listed in full in Appendix D) were constructed following the pattern of (16) above. Each item consisted of a sentence-initial subject (determiner plus noun) followed by a word sequence with parts of speech as shown in (13) above, 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 (16)), 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 to the first post-verbal noun.

The first post-verbal noun was always a singular human noun, and the second post-verbal noun was always a singular noun and never denoted a single human. 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 target sentence never immediately followed another target sentence.

5.3. 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 minutes per participant to complete.

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

Table 11: Question-answering accuracy in Experiment 3

	F1	F2
PP attachment	<1	<1
RC adjacency	1.28	<1
PP×Adjacency	<1	<1

Table 12: F-statistics for analysis of question-answering accuracy in Experiment 3. No effects were significant.

	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	$16.08^{\ddagger}$	$8.38^{\dagger}$	<1	<1	$3.33^{\cdot}$	$3.59^{\cdot}$	$8.12^{\dagger}$	$3.73^{\cdot}$	$3.06^{\cdot}$	<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

Table 13: F-statistics for Experiment 3 ($^{\cdot}p < 0.1$, $^{*}p < 0.05$, $^{\dagger}p < 0.01$, $^{\ddagger}p < 0.001$). In regions not listed, there were no effects were below $p = 0.1$.

5.4. Results

5.4.1. 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.

5.4.2. Comprehension accuracy

Overall question-answering accuracy on experimental items was much higher than for Experiment 2 at 87%. Tables 11 and 12 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.

5.4.3. Reading times

Figure 6 shows region-by-region reading times, and Table 13 gives the results of 2×2 ANOVAs by participants and by items for each region. Before the relative clause, in Region 3, we find an interaction marginal by participants between PP attachment and RC adjacency on reading time, which we ascribe to chance since the reader has not yet seen the PP or RC. In Region 4 (the post-object PP) we find a significant main effect of PP attachment similar to that found in Experiment 2.

During 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 Experiments 1 and 2. Figure 7 shows the average residual RT per

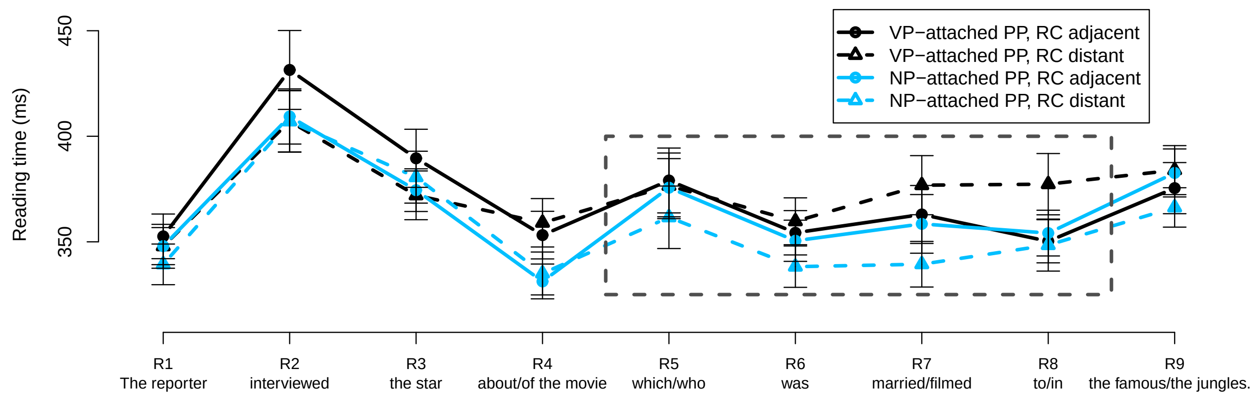


Figure 6: Reading-time results as a function of region and condition for Experiment 3. Onset of the relative clause (first four words) is boxed.

region as a function of condition, and Table 14 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 ($F1(1,43) = 20.62, p < .001; F2(1,19) = 14.04, p = 0.001$) and of RC adjacency within the VP-attached PP conditions ($F1(1,43) = 25.31, p < .001; F2(1,19) = 19.31, p < .001$).

5.5. Discussion

The results of Experiment 3 support derivational-complexity and structural-expectation based accounts of RC extraposition difficulty, and are problematic for decay/interference accounts based on linear distance. Additionally, the present results allay methodological concerns about Experiment 2: question answering accuracy in Experiment 3 is uniformly higher than in the previous experiment, and differs only minimally (and not significantly) across conditions.

6. Experiment 4

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

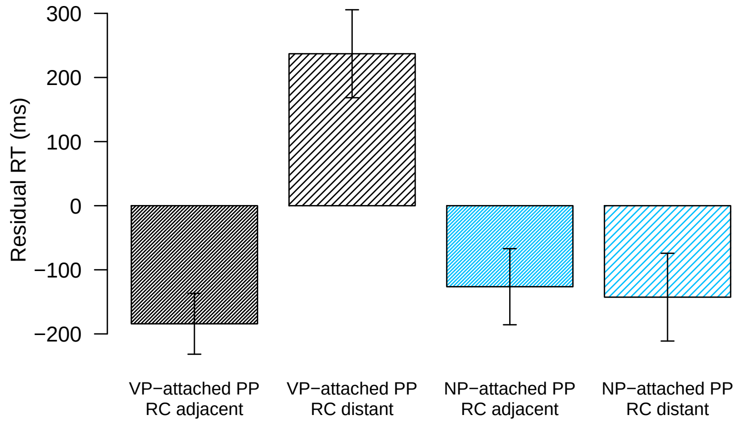


Figure 7: Residual reading times over the first four words of the RC in Experiment 3

	F1	F2
PP attachment	$13.23^{\ddagger}$	$10.60^{\dagger}$
RC adjacency	$10.12^{\dagger}$	$10.33^{\dagger}$
PP×Adjacency	$14.01^{\ddagger}$	$9.78^{\dagger}$

Table 14: F-statistics for residual reading-time analysis across first four words of the RC in Experiment 3 ($^{\dagger}p < 0.01$,$^{\ddagger}p < 0.001$).

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 this experiment, 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., for example, note 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 15 shows the proportion of NPs possessing projective-dependency RC postmodifiers for various types of premodifying structure in the parsed Brown corpus (crossing-dependency 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. This facilitation should extend to extraposed RCs, as well—but only if

Premodifier	Proportion RC postmodification	n
a/an	8.8%	10071
the	7.8%	22193
only	24.6%	313
the only	64.9%	74
only those	100%	1

Table 15: Relative frequency of RC postmodifier for various types of premodifying structure. Totals include both extraposed and unextraposed RCs.

(i) the intuition and corpus data regarding postmodifier frequency generalize to crossing-dependency RC structures, and (ii) comprehenders are able to maintain and take advantage of corresponding expectations outside of the projective-dependency domain. In this experiment, we tested 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 co-occurrence frequencies with postmodifying RCs, but in Section 6.2.4 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 (17) below:

(17) a. The chairman consulted the executives about the company which was making lots of money. [LOW expectation, −extraposition]
b. The chairman consulted the executives about the company who were making lots of money. [LOW expectation, +extraposition]
c. The chairman consulted only those executives about the company which was making lots of money. [HIGH expectation, −extraposition]
d. The chairman consulted only those executives about the company who were making lots of money. [HIGH expectation, +extraposition]

Examples (17a)—(17b) are identical to the VP-attached conditions of Experiment 2, hence the RC should be harder to read in the extraposed variant (17b) than in the unextraposed variant (17a). Examples (17c)—(17d) differ in that the prenominal material *only those*, which modifies the direct-object (DO) noun *executives*, should create a high expectation for a relative clause that postmodifies *executives*. 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 the DO. After encountering *about the company*, it becomes clear that any postmodifier of the DO 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 their high expectation for a DO postmodifier beyond the continuous-constituent domain, however, we should see that reading of the extraposed RC in (17d) is facilitated relative to the low-expectation variant, (17b). 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, 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 HIGH expectations conditions of (17) 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 (17c) than in (17a). Depending on the strength of the expectation induced by the premodifier *only those*, we might even see a reversal in the high-expectation condition, such that the extraposed RC ((17d)) is easier to read than the unextraposed RC ((17c)). We tested this prediction in a self-paced reading study using sets of sentences as in (17).

6.1. Method

6.1.1. Participants

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

6.1.2. Materials

We constructed twenty-four items (listed in full in Appendix E) on the pattern of Example (17). 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. Due to concerns about comprehension accuracy raised in Experiment 2, and in order to minimize any attachment garden-path effects, we used relative pronouns giving animacy cues: *who* for the extraposed conditions and *which* for the unextraposed conditions. The second word of the RC was the auxiliary *were* in the extraposed conditions and *was* in the unextraposed conditions. The remainder of the RC concluded the sentence and was the same across conditions. These test sentences were interspersed among 100 fillers, including 64 from two other experiments.

6.1.3. Procedure

The procedure was the same as in previous experiments. The study took an average of 45 minutes per participant to complete.

6.2. Results

6.2.1. Statistical analysis

Statistical analysis procedures were identical to those in Experiment 1. Due to a programming error, one of the conditions in item 24 was incorrectly entered.[9] We present analyses with this item omitted; results with the item included are qualitatively the same. Outlier removal procedures led to 2.2% of reading-time data (not including Item 24) being discarded.

[9] In the low-expectation unextraposed condition, the word *union* was substituted for *outfit*.

	Extraposed	Unextraposed
Non-presentative	0.77	0.80
Presentative	0.59	0.82

Table 16: Question-answering accuracy in Experiment 4

	F1	F2
Expectation	$7.83^{\dagger}$	$9.74^{\dagger}$
Extraposition	$44.48^{\ddagger}$	$8.92^{\dagger}$
Expect×Extrap	$21.89^{\ddagger}$	$14.23^{\dagger}$

Table 17: F-statistics for analysis of question-answering accuracy in Experiment 4 ($^{\dagger}p < 0.01$,$^{\ddagger}p < 0.001$)

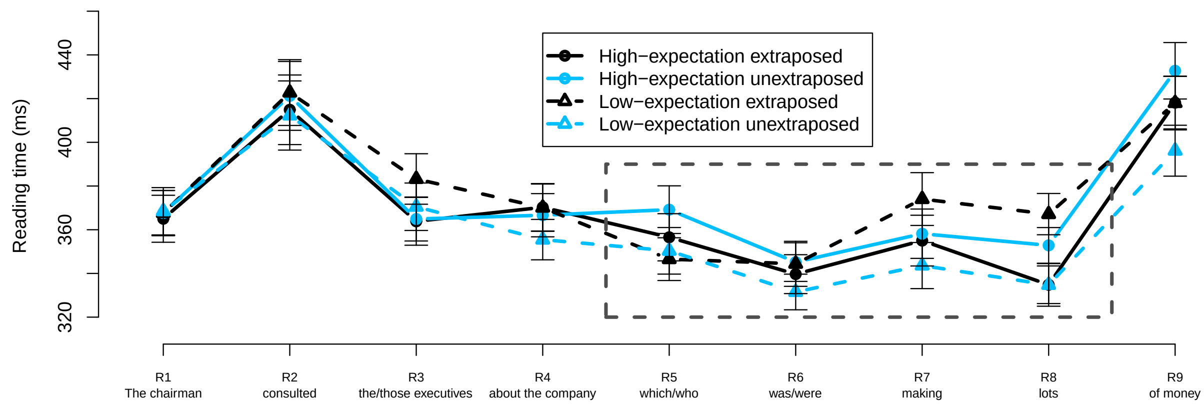


Figure 8: Region-by-region reading times for Experiment 4

6.2.2. Comprehension Questions

Question answering accuracies 2×2 ANOVA results are reported in Tables 16 and 17. We found significant main effects of both expectation and extraposition, driven by a significant interaction between expectation and extraposition such that accuracies were lowest in the low-expectation extraposed condition. The results for the low-expectation conditions are comparable to the results of Experiment 2.

6.2.3. Reading Times

Figure 8 shows region-by-region reading times, and Table 18 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 low-expectation condition.) Before the relative clause, we find in Region 3 (the post-verbal object) a significant main effect of expectation presumably due to the differing prenominal structure (*the* versus *only those*). In Region 4 we find a main effect of extraposition significant by participants, which we ascribe to chance, since none of the RC has yet been revealed. The crucial results of this experiment occur in the relative clause, where we find a significant main effect of expectation at Region 5 (the relative pronoun), with reading times slower in the high-expectation condition. In Regions 6 through 9, we find an interaction between expectation and extraposition in the

	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
Expectation	<1	<1	<1	<1	5.37^*	5.92^*	2.07	2.02	$7.52^\dagger$	5.57^*	<1	1.31	<1	<1	2.20	2.41	$8.73^\dagger$	$4.27^\cdot$
Extraposition	<1	<1	<1	<1	1.47	<1	4.03^*	<1	2.08	<1	<1	<1	4.43^*	<1	$2.80^\cdot$	<1	<1	<1
Expect×Extrap	<1	<1	1.72	1.14	2.47	1.87	1.49	<1	<1	<1	$3.76^\cdot$	1.64	$9.72^\dagger$	7.40^*	$25.63^\ddagger$	$19.41^\ddagger$	$8.86^\dagger$	$9.12^\dagger$

Table 18: F-statistics for Experiment 4 ($^\cdot p < 0.1, ^* p < 0.05, ^\dagger p < 0.01, ^\ddagger p < 0.001$). In regions not listed, there were no effects were below $p = 0.1$.

direction of extraposition being more difficult in low-expectation conditions, but not in high-expectation conditions. This interaction is marginal by participants in Region 6 (*was/were*), and significant by both participants and items in Regions 7, 8, and 9. There are also main effects of extraposition (Region 7) and expectation (Region 9) driven by this interaction.

As in the previous three experiments, we conducted a residual reading-time analysis across the first four words of the RC, using the same methodology as described in Section 3.4.3. Figure 9 shows the average residual RT per region as a function of condition, and Table 19 reports results of a 2×2 ANOVA on these data. This analysis reveals similar effects as observed in raw RTs in Regions 7 and 8: a main effect of extraposition significant by subjects, with extraposed RC reading times slower overall, and a significant interaction between expectation and extraposition such that extraposed RCs are read numerically more slowly than unextraposed RCs in the low-expectation condition, but numerically faster in the high-expectation condition. Pairwise comparisons within each expectation condition revealed this effect of extraposition to be significant in the low-expectation condition ($F1(1,71) = 7.66, p = 0.007; F2(1,22) = 11.43, p = 0.003$), and in the high-expectation condition ($F1(1,71) = 9.94, p = 0.002; F2(1,22) = 10.40, p = 0.004$), although the size of the effect is numerically larger in the low-expectation condition (189ms) than in the high-expectation condition (145ms). Pairwise comparisons within each extraposition condition revealed the expectation effect to be significant in the unextraposed condition ($F1(1,71) = 13.29, p < .001; F2(1,22) = 45.71, p < .001$); in the extraposed condition, the effect was insignificant by subjects and marginal by items ($F1(1,71) < 1; F2(1,22) = 3.40, p = 0.079$).

6.2.4. 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 high- versus low-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 low-expectation condition (189ms in favor of unextraposed RCs) than in the high-expectation condition (145ms 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 high-expectation than in the low-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.

The completion study used the pre-relative clause sentence fragments from the self-paced

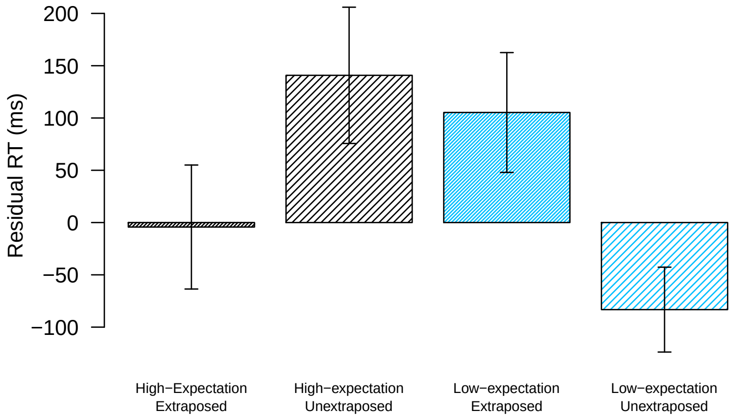


Figure 9: Residual reading times over the first four words of the RC in Experiment 4

	F1	F2
Expectation	1.31	2.48
Extraposition	<1	<1
Expect×Extrap	$12.00^{\ddagger}$	$21.90^{\ddagger}$

Table 19: F-statistics for residual reading-time analysis across first four words of the RC in Experiment 4 ($^{\ddagger}p < 0.001$).

reading study items as prompts, as in Example (18) below:

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

Sixteen native English speakers at MIT participated in this study for cash compensation. For each participant, half the items were presented in the low-expectation *the* condition (18a) and the other half were presented in the high-expectation *only those* condition (18b); which items were presented in which condition was rotated across participants. A native English-speaker research assistant, together with one of the authors (EF), 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 postverbal 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. Examples of cases (i–iv) are given in (19) below:

(19) a. The publisher complimented only those editors on the magazine *who completed their work on time.* (high-expectation, NP_1)
b. The nanny consulted only those babysitters about the virus *that was spreading around the school.* (high-expectation, NP_2)
c. The agent approached the publicist about the photoshoot *that is to take place in a few hours.* (low-expectation, NP_2)
d. The chairman consulted only those executives about the company *that he disagreed with about the loan.* (high-expectation, ambiguous)
e. The candidate criticized the senators for the attack *against her character.* (low-

expectation, other)

f. The agent approached only those publicists about the photoshoot *schedule conflict.* (high-expectation, other)

The proportion of each completion type is shown in Figure 10. The most notable result is that in the low-expectation condition, not a single clearly NP_1-modifying extraposed RC was produced, whereas in the high-expectation condition, most of the RCs produced were extraposed NP_1 modifiers (68% if all ambiguous cases are considered unextraposed NP_2 attachments, and 88% or 91% for the more realistic situations in which ambiguous cases are either ignored or treated as extraposed NP_1 attachments respectively). To analyze this trend statistically, we considered only those cases coded as unambiguous NP_1 or NP_2 attachments—145 trials in total. Because these data are imbalanced, we analyzed the relative preference for NP_1 versus NP_2 attachment with a mixed-effects logit model (Agresti, 2002; Jaeger, 2008), allowing participant- and item-specific random slopes for each condition. A likelihood-ratio test revealed a highly significant improvement in model log-likelihood for a model with a fixed effect of expectation versus a model with no fixed effect of expectation ($\chi^2(1) = 122.55, p < 10^{-6}$).[10] This result corroborates intuition and corpus data in demonstrating that the use of prenominal *only those* strongly changed expectations for what types of RC continuations might be encountered.

We also addressed the issue of *strength of expectation* raised by the signs of a main effect of extraposition on RC reading time in the self-paced reading study, by testing for an effect of expectation *on frequency of occurrence of the preferred outcome*—that is, whether the proportion of NP_2-attaching RCs among the 45 unambiguously-attached RCs produced in the low-expectation condition significantly exceeds the proportion of NP_1-attaching RCs among the 100 unambiguously-attached RCs produced in the high-expectation condition. In a mixed-effects logit analysis with participant- and item-specific random intercepts, a likelihood-ratio test revealed a significant contribution of expectation ($\chi^2(1) = 20.21, p < 10^{-5}$).[11] This result suggests that the relative preference for an NP_2 attachment in the low-

[10]With random intercepts alone, `lme4` reported false convergence. Although the likelihood-ratio test for models differing only in fixed-effects structure may be anti-conservative, this anti-conservativity is minimal for cases such as ours in which the number of degrees of freedom of the test is small compared to the residual degrees of freedom in the more complex model (Pinheiro & Bates, 2000, pp. 87–92). Although hypothesis testing based on the Wald statistic is often recommended for mixed logit models (e.g., Jaeger, 2008), in our case the absence of NP_1 responses in the low-expectation condition leads to an extremely large coefficient and an accompanying inflated standard error for this fixed effect 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 (Menard, 1995; Agresti, 2002). For completeness, we analyzed the data with Fisher's exact test (which involves no multi-level structure), which rejects the null hypothesis at $p < 10^{-16}$; and we also conducted a Bayesian mixed-effect logit analysis implemented in JAGS (Plummer, 2003) with participant- and item-specific random slopes, whose posterior distribution indicated with greater than 99% confidence a significant effect of expectation condition.

[11]We also tried adding participant- and item-specific random slopes; however, when the random-slope model did not have a fixed effect of condition (required for model comparison), `lme4` failed to return reasonable random-effect covariance matrices; we believe that this failure is due to the absence of NP_1 responses

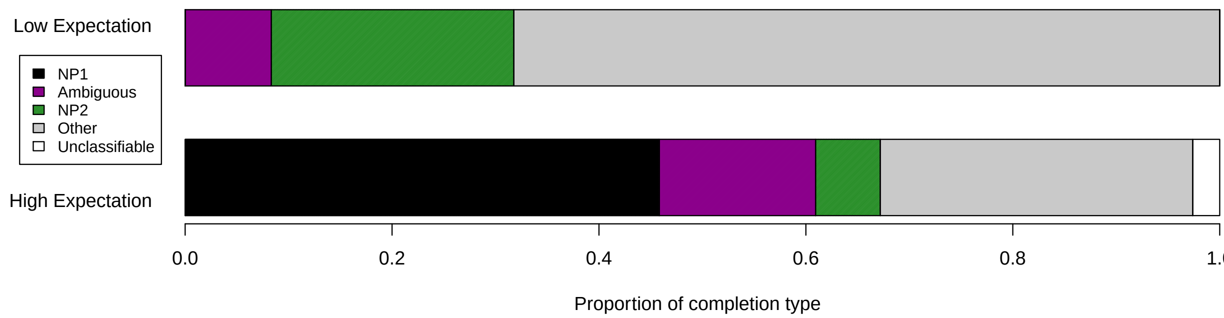


Figure 10: Completion-study results

expectation condition is in fact somewhat stronger than the relative preference for an NP_1 attachment in the high-expectation condition.

6.2.5. Discussion

The crucial result of the self-paced reading study in this experiment was the interaction between the expectation for a relative clause modifying the direct object (*executives* in (17)) and the attachment site of a relative clause following the verb-modifying PP (*about the company* in (17)) on reading time within the RC. When the expectation for a DO-modifying relative clause is low, an unextraposed (PP-modifying) RC is considerably easier to read than an extraposed (DO-modifying) RC, and participants are much less accurate at answering questions involving sentences with extraposed RCs in these conditions. When the expectation for a DO-modifying RC is high, on the other hand, the difficulty in answering questions almost completely disappears, and the reading-time difference reverses completely: unextraposed RCs are now harder to process than extraposed RCs. This interactive pattern can be understood as a consequence of probabilistic syntactic expectations: in the presence of a strong expectation that the DO should have a postmodifier, the comprehender may expect more strongly that the next constituent is an extraposed DO modifier than an unextraposed PP modifier. There was a weak numerical trend consistent with an effect of derivational complexity, evident in the fact that the absolute size of the numerical difference between processing times in the extraposed versus unextraposed conditions times was smaller in high-expectation conditions than in low-expectations. However, the main effect of extraposition was far from significant, and the completion study indicates that the relative expectation for an unextraposed RC in the low-expectation condition is in fact stronger than the relative expectation for an extraposed RC in the high-expectation condition, suggesting that any

in the low-expectation condition, combined with the relatively smaller contrast in raw proportions than in the previous analysis. For completeness, we analyzed the data with Fisher's exact test (which involves no multi-level structure), which rejects the null hypothesis at $p = 0.025$; and we also conducted a Bayesian mixed-effect logit analysis implemented in JAGS (Plummer, 2003) with participant- and item-specific random slopes, whose posterior distribution indicated with greater than 99% confidence a significant effect of expectation condition.

such 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 interaction 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 high-expectation unextraposed condition induced when no DO-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 low-expectation contexts, and factor (B) would determine the pattern in the high-expectation contexts. What crucially militates against this analysis, however, is that during reading of an unextraposed RC it is still too early for the comprehender to know that the DO will not ultimately have an extraposed postmodifier appearing later in the sentence. For example, in (20) below an extraposed RC appears immediately after the unextraposed RC.

(20) The chairman consulted only those executives about the company which was making lots of money who had worked at that company previously.

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 (17c) 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.

7. General Discussion

The results of our four 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 high 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 2.1.3. The results of Experiments 2 and 3 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 4 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 reverse the pattern of difficulty so that reading times in the extraposed variant are in fact lower than those in the unextraposed variant.

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 (Lau, Stroud, Plesch & Phillips, 2006; Staub & Clifton, 2006; Staub, Clifton & Frazier, 2006; Jaeger et al., 2008) 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.[12] 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 4, in contrast, relatively little semantic information was available at the disambiguation point that could give rise to plausibility differences, and 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 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.

This result has implications for the nature of the representations used in the online computation of sentence structure in comprehension. As depicted in Figure 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—

[12]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 crossing dependencies, and correspondingly some syntactic analyses (e.g., Pollard & Sag, 1994) treat these as strictly continuous-constituent constructions.

e.g., Figure 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 4, 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 (Figures 1a and 1b), or they must allow the explicit representation of discontinuous constituents (Figure 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). 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 return to the distributional and typological questions we raised at the outset of the paper: that non-projective dependencies are relatively infrequent in most of the languages in which they exist, including English. On our interpretation of the data presented here, the comprehension difficulty associated with the more common types of extraposed RC structures is a consequence of particular distributional facts of English—no additional factors are necessary to explain the key results of this paper. Although some aspects of Experiment 4 were consistent with a derivational-complexity penalty for extraposition, our completion study suggested that this apparent penalty could easily be due to an asymmetry of expectation strength across expectation condition. Thus, our primary results with respect to the question of why non-projectivity is rare in natural language are negative: However, this negative result brings us a step toward a sharper view of possible answers to the original question. We conclude by speculating on remaining possible explanations for why non-projectivity is rare.

One possible source of the rarity of non-projectivity could be the general effects (rather than effects specific to non-projectivity) of decay (Gibson, 1998, 2000) and interference (Gordon et al., 2001, 2004; Lewis & Vasishth, 2005; Lewis et al., 2006) in sentence comprehension. Although our results demonstrated that patterns of RC comprehension difficulty in English are not reducible to principles of decay and interference, nothing we have found here precludes decay and interference playing a role as a general determinant of comprehension difficulty along the lines found by Grodner & Gibson (2005), Gordon, Hendrick, Johnson & Lee (2006), Demberg & Keller (2008), and others. With specific respect to extraposition, since unextraposed postmodifiers are generally closer to the nouns they modify and thus involve shorter-length dependencies with fewer interveners that could interfere with retrieval, they would be likely to be favored in retrieval-based theories. The idea that dependency-length minimization has played a general role in shaping natural-language grammars and language-internal distributional preferences has been explored by a number of researchers including Hawkins (1994, 2004) and Wasow (2002), and has gained recent empirical support from studies of written English (Temperley, 2007; Gildea & Temperley, 2007) and Rumanian (Ferrer i Cancho, 2006) showing that empirically observed dependency distances are considerably closer to being minimal than random, distance-insensitive linearization would predict. Ferrer i Cancho in particular shows through simulations that dependency trees arranged so as to minimize linear dependency distance also turn out to have small numbers of crossing dependencies. On the other hand, Park & Levy (2009) replicate the results of Gildea & Temperley (2007) for English but also show that in German, empirically observed dependency distances are closer to random than they are to minimal, suggesting that the cross-linguistic picture may be rather complicated and requires further work to elucidate fully.

Moving beyond comprehension-based explanations, the discussion returns to the alternative possibilities of acquisition bias, speaker-centric preference, and information structure. We close simply by noting that speaker-centric preference is a natural next place to look—in particular at the well-established role of *constituent weight* in linear ordering preferences. Although it is unclear whether the weight of a given constituent is best measured in terms of syllables, words, nodes, or some other metric (Wasow, 1997, 2002; Jaeger, 2006), it has been observed since Yngve (1960) and Ross (1967) that "larger" constituents in the English postverbal domain tend to be placed later in the sentence (Hawkins, 1994; Wasow, 1997; Arnold, Wasow, Losongco & Ginstrom, 2000; Wasow, 2002). (There is a correlation between constituent weight and discourse newness, and there is evidence (Arnold & Lao, 2008) that speakers expect later postverbal constituents to be newer.) One explanation that has been proposed several times for this observation is that in the postverbal domain, placing a larger constituent before a smaller constituent would require the producer to remember the beginning of the smaller constituent for a longer time, whereas the smaller-before-larger ordering imposes this memory burden for a shorter time (Yngve, 1960; Hawkins, 1994; Wasow, 1997; Gibson, 1998, p. 52; Wasow, 2002). Choosing to extrapose a relative clause could create a similar burden of memory during production of the intervening material. It is already established that extraposition in written German is less common the longer the potentially intervening material (Uszkoreit, Brants, Duchier, Krenn, Konieczny, Oepen & Skut, 1998). For English, even in the cases where potentially intervening material is short (e.g., *I put some very heavy books on the table that were given to me by my cousin*), the Heavy Noun-Phrase Shift option is also available, which would shift the entire noun phrase past the intervener (*I put on the table some very heavy books that were given to me by my cousin*). It may be the case that, even when the circumstances under which extraposition is possible do arise, the linear ordering variant minimizing memory burden during production rarely involves extraposition. All of the possibilities outlined within these three classes of explanation deserve future investigation.

Agresti, A. (2002). *Categorical Data Analysis* (second ed.). Wiley.

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

Anderson, J. R. (1990). *The Adaptive Character of Human Thought*. Lawrence Erlbaum.

Arnold, J. & Lao, S.-Y. C. (2008). Put in last position something previously unmentioned: Word order effects on referential expectancy and reference comprehension. *Language and Cognitive Processes*, *23*(2), 282–295.

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.

Bach, E., Brown, C., & Marslen-Wilson, W. (1986). Crossed and nested dependencies in German and Dutch: A psycholinguistic study. *Language and 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.

Bod, R. (1998). *Beyond Grammar: An Experience-Based Theory of Language*. CSLI Press.

Boersma, P. & Hayes, B. (2001). Empirical tests of the gradual learning algorithm. *Linguistic Inquiry*, *32*(1), 45–86.

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.

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. (1965). *Aspects of the Theory of Syntax*. Cambridge.

Chomsky, N. (1981). *Lecture on Government and Binding*. Mouton de Gruyer.

Clark, H. H. & Murphy, G. L. (1982). Audience design in meaning and reference. In J. F. L. Ney & W. Kintsch (Eds.), *Language and Comprehension*, volume 9 of *Advances in Psychology* (pp. 287–297). North Holland.

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* (Second ed.). MIT Press.

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

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, Jr., C. (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.

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

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.

Gennari, S. P. & MacDonald, M. C. (2008). Semantic indeterminacy in object relative clauses. *Journal of Memory and Language*, *58*, 161–187.

Gennari, S. P. & MacDonald, M. C. (2009). Linking production and comprehension processes: The case of relative clauses. *Cognition*, *111*(1), 1–23.

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.

Gildea, D. & Temperley, D. (2007). Optimizing grammars for minimum dependency length. In *Proceedings of ACL*.

Givón, T. (1993). English grammar: A function-based introduction. In *English Grammar: A function-based introduction*, volume 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).

Goldsmith, J. (2001). Unsupervised learning of the morphology of a natural language. *Computational Linguistics*, *27*(2), 154–198.

Goldwater, S., Griffiths, T. L., & Johnson, M. (2009). A Bayesian framework for word segmentation: Exploring the effects of context. *Cognition*, *112*(1), 21–54.

Good, P. I. (2004). *Permutation, Parametric, and Bootstrap Tests of Hypotheses* (Third ed.). Springer.

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

Gordon, P. C., Hendrick, R., Johnson, M., & Lee, Y. (2006). Similarity-based interference during language comprehension: Evidence from eye tracking during reading. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *32*(6), 1304–1321.

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. (1999). Processing complexity and filler-gap dependencies. *Language*, *75*(2), 244–285.

Hawkins, J. A. (2004). *Efficiency and Complexity in Grammars*. Oxford University Press.

Jaeger, T. F. (2006). *Redundancy and Syntactic Reduction in Spontaneous Speech*. PhD thesis, Stanford University, Stanford, CA.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446.

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.

Keenan, E. J. & Comrie, B. (1977). Noun phrase accessibility and universal grammar. *Linguistic Inquiry*, *8*(1), 63–99.

Kirby, S. (1999). *Function, Selection, and Innateness: The Emergence of Language Universals*. Oxford University Press.

Kobele, G. M. (2006). *Generating Copies: An investigation into Structural Identity in Language and Grammar*. PhD thesis, UCLA.

Kruijff, G.-J. M. & Vasishth, S. (2003). Quantifying word order freedom in natural language: Implications for sentence processing. Proceedings of the Architectures and Mechanisms for Language Processing conference.

Krylova, O. & Khavronina, S. (1988). *Word Order in Russian*. Moscow, USSR: Russky Yazyk Publishers.

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 and 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.

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*, volume 31 of *Syntax and Semantics* (pp. 313–336). Academic Press.

Mitchell, D. C. (1984). An evaluation of subject-paced reading tasks and other methods for investigating immediate processes in reading. In D. Kieras & M. A. Just (Eds.), *New methods in reading comprehension*. Hillsdale, NJ: Earlbaum.

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*, volume 14, (pp. 59–65).

Nederhof, M.-J. (1999). The computational complexity of the correct-prefix property for TAGs. *Computational Linguistics*, *25*(3), 345–360.

Park, Y. A. & Levy, R. (2009). Minimal-length linearizations for mildly context-sensitive dependency trees. In *Proceedings of the North American Chapter of the Association for Computational Linguistics – Human Language Technologies (NAACL-HLT) conference*.

Perfors, A., Tenenbaum, J., & Regier, T. (2006). Poverty of the stimulus? a rational approach. In *Proceedings of the 28th Annual Conference of the Cognitive Science Society*.

Pinheiro, J. C. & Bates, D. M. (2000). *Mixed-Effects Models in S and S-PLUS*. Springer.

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*.

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: University of Chicago Press and Stanford: CSLI Publications.

Resnik, P. (1992). Probabilistic Tree-Adjoining Grammar as a framework for statistical natural language processing. In *Proceedings of ACL*.

Rohde, D. (2001). *Linger: a flexible platform for language processing experiments*.

Ross, J. R. (1967). *Constraints on Variables in Syntax*. PhD thesis, MIT.

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.

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. & 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, Jr, C., & Frazier, L. (2006). Heavy NP shift is the parser's last resort: Evidence from eye movement. *Journal of Memory and Language*.

Steedman, M. (2000). *The Syntactic Process*. Cambridge, MA: MIT Press.

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.

Temperley, D. (2007). Minimization of dependency length in written English. *Cognition*, *105*, 300–333.

Thibadeau, R., Just, M., & Carpenter, P. A. (1982). A model of the time course and content of reading. *Cognitive Science*, *6*, 157–203.

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., & Skut, W. (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 Wiese, H. & Simon, H. (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.

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 bracketing 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 Figure 1on 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 <<# (__ !< __))`	49284	`/^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 4. (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.)

<table>
<tr><th>PP/RC Type</th><th>Conditioning pattern</th><th>count</th><th>Outcome pattern</th><th>count</th></tr>
<tr><td>NP/adj.</td><td>@NP > @VP <# (__ $+ (@PP <<# of))</td><td>2552</td><td>@NP > @VP <# (__ $+ (@PP <<# of $+ (@SBAR </^WH/)))</td><td>101</td></tr>
<tr><td>NP/non-adj.</td><td>@NP > @VP <# (__)</td><td>30127</td><td>@NP > @VP <# (__ $+ (@SBAR </^WH/))</td><td>1139</td></tr>
<tr><td>VP/adj.</td><td>@NP > @VP $+ (@PP <- @NP <# __)</td><td>6576</td><td>@NP > @VP $+ (@PP <- (@NP <# (__ $+ (@SBAR < /^WH/))))</td><td>218</td></tr>
<tr><td>VP/non-adj.</td><td>@NP > @VP <# (__ $+ (@SBAR </^WH/))</td><td>1139</td><td>@NP < (@SBAR < (-NONE- < /^*ICH*-([0-9]+)/#1%i)) > @VP $+ (@PP <- (@NP <# __) $+ (/^SBAR-([0-9]+)$/#1%i < /^WH/))</td><td>4</td></tr>
</table>

The following table lists patterns for Table 15.

<table>
<tr><th>Probability</th><th>Conditioning pattern</th><th>count</th><th>Outcome pattern</th><th>count</th></tr>
<tr><td>a/an</td><td>@NP < (DT < a|an)</td><td>10071</td><td>@NP < (DT < a|an $++ (@SBAR </^WH/))</td><td>874</td></tr>
<tr><td>the</td><td>@NP < (DT < the)</td><td>22193</td><td>@NP < (DT < the $++ (@SBAR </^WH/))</td><td>1719</td></tr>
<tr><td>only</td><td>@NP < (__ , only)</td><td>313</td><td>@NP < (__ , only $++ (@SBAR </^WH/))</td><td>77</td></tr>
<tr><td>the only</td><td>@NP < (DT < (the . only)</td><td>74</td><td>@NP < (DT < (the . only) $++ (@SBAR </^WH/))</td><td>48</td></tr>
<tr><td>only those</td><td>@NP < (DT < those , only)</td><td>313</td><td>@NP < (DT < those , only $++ (@SBAR </^WH/))</td><td>77</td></tr>
</table>

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

There were four versions of each item, crossing PP attachment (VP, NP) and relative clause adjacency (non-adjacent, adjacent). All four conditions are provided for the first example. For the remaining items, only the VP/non-adjacent condition is presented. The remaining versions of each item can be formed as in the first example. The first 12 items included relative clauses with plural verbs, whereas the second twelve included relative clauses with singular verbs.

1. (a) VP, non-local: The chairman consulted the executive about the companies that was making lots of money.
   (b) NP, non-local: The chairman consulted the executive of the companies that was making lots of money.
   (c) VP: local: The chairman consulted the executive about the companies that were making lots of money.
   (d) NP, local: The chairman consulted the executive of the companies that were making lots of money.

2. The reporter interviewed the star about the movies that was getting very good reviews.
3. The student petitioned the professor about the courses that was demanding too much effort.
4. The agent approached the publicist about the models that was demanding a different photographer.
5. The socialite praised the hostess about the parties that was pleasing all the guests.
6. The publisher complimented the editor about the magazines that was winning many journalism awards.
7. The teacher called the parent about the children that was speaking two languages fluently.
8. The principal criticized the instructor about the classes that was refusing to admit women.
9. The relative approached the owner about the cats that was making guests feel uncomfortable.
10. The producer complimented the director about the films that was receiving many critics' praise.
11. The officer questioned the guide about the tourists that was exploring the restricted ruins.
12. The nanny consulted the babysitter about the children that was worrying about the parents.
13. The opponent criticized the representatives about the senator that were appearing on TV ads.
14. The reporter called the agents about the firm that were representing several famous athletes.
15. The superhero interrogated the henchmen about the villain that were refusing to take prisoners.
16. The colonel praised the captains about the platoon that were winning honors in combat.
17. The advertiser approached the hosts about the show that were appealing to young women.
18. The fan petitioned the owners about the team that were refusing to allow rookies.
19. The stockholder queried the employees about the company that were predicting low annual earnings.
20. The father interrogated the suitors about the bachelorette that were looking to get married.
21. The judge queried the representatives about the defendant that were entering a guilty plea.
22. The reporter interviewed the pollsters about the politician that were hoping for a victory.
23. The detective questioned the victims about the mobster that were fearing retaliation for testimony.
24. The architect consulted the assistants about the carpenter that were building the new house.

Appendix D. Materials for experiment 3

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 E. Materials for experiment 4

For every item, the low-expectation conditions used *the* as the prenominal material in the main-clause object NP, whereas the high-expectation conditions used *only those*.

1. (a) Low expectation, unextraposed: The chairman consulted the executives about the company which was making lots of money.
   (b) Low expectation, extraposed: The chairman consulted the executives about the company who were making lots of money.
   (c) High expectation, unextraposed: The chairman consulted only those executives about the company which was making lots of money.
   (d) High expectation, extraposed: The chairman consulted only those executives about the company who were making lots of money.
2. The reporter interviewed the actors about the movie who were getting very good reviews.
3. The student petitioned the professors regarding the course who were demanding too much effort.

4. The agent approached the publicists about the photoshoot who were needing more careful lighting.
5. The socialite praised the hostesses for the party who were pleasing all the guests.
6. The publisher complimented the editors on the magazine who were winning many journalism awards.
7. The counselor consoled the students about the group who were making class slow down.
8. The principal criticized the instructors for the program who were refusing to admit women.
9. The visitor approached the owners about the cat who were making guests feel uncomfortable.
10. The producer complimented the directors on the documentary who were receiving significant critical acclaim.
11. The officer questioned the guides about the expedition who were exploring the restricted ruins.
12. The nanny consulted the babysitters about the virus who were making the parents worry.
13. The candidate criticized the senators for the attack who were appearing on TV ads.
14. The reporter called the agents about the firm who were representing several famous athletes.
15. The superhero interrogated the henchmen about the organization who were refusing to take prisoners.
16. The colonel praised the captains for the platoon who were winning honors in combat.
17. The advertiser approached the hosts about the show who were appealing to young women.
18. The fan petitioned the coaches regarding the team who were refusing to consider rookies.
19. The stockholder queried the employees about the agency who were predicting low annual earnings.
20. The father interrogated the suitors about the venue who were interesting to his daughter.
21. The judge queried the lawyers about the evidence who were losing credibility very quickly.
22. The reporter interviewed the pollsters about the party who were expecting an easy victory.
23. The detective questioned the victims about the gang who were fearing retaliation for testimony.
24. The architect consulted the carpenters about the outfit [incorrectly *union* in the low expectation, unextraposed condition] who were protesting the working conditions.