COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 43 (2019) e12714

ISSN: 1551-6709 online
DOI: 10.1111/cogs.12714

Rare Constructions Are More Often Sentence-Initial

David Temperley

Eastman School of Music, University of Rochester

Received 27 March 2018; received in revised form 12 November 2018; accepted 10 January 2019

Abstract

Main clause phenomena (MCPs) are syntactic constructions that occur predominantly or exclusively in main clauses. I propose a processing explanation for MCPs. Sentence processing is easiest at the beginning of the sentence (requiring less search); this follows naturally from widely held assumptions about sentence processing. Because of this, a wider variety of constructions can be allowed at the beginning of the sentence without overwhelming the sentence-processing mechanism. Unlike pragmatic and grammatical accounts of MCPs, the processing account predicts avoidance of MCPs in non-initial main clauses (non-initial coordinate clauses and premodified clauses). A corpus study supports these predictions, but it is somewhat inconclusive. A further corpus study examines another type of syntactic construction, premodifying adjunct phrases ("openers"); the prediction here is that less common types of opener will be especially avoided in non-initial contexts. The prediction is confirmed, supporting the processing view of rare constructions.

Keywords: Syntax; Main clause phenomena; Sentence processing; Corpus analysis; Audience design

1. Introduction

It has long been known that certain syntactic constructions occur predominantly, if not exclusively, in main clauses. An example is NP topicalization, in which a non-subject NP (normally a direct object) is moved to the beginning of the clause. This feels quite natural in main clauses (1a) but seems awkward, if not incorrect, in embedded clauses (1b) and adverbial clauses (1c):

(1) a. Mushroom pizza she can't stand.
b. ? I forgot that mushroom pizza she can't stand.
c. ? She won't eat with us if mushroom pizza she can't stand.

Correspondence should be sent to David Temperley, Eastman School of Music, University of Rochester, 26 Gibbs St., Rochester, NY 14604. E-mail: dtemperley@esm.rochester.edu

Another example is participle preposing, in which a fronted passive or progressive participle phrase is followed by an inverted auxiliary and subject. This, too, seems much more acceptable in main clauses (2a) than in dependent clauses ((2b) and (2c)):

(2) a. Carrying the flag was one of the town's leading citizens.
b. ? I forgot that carrying the flag was one of the town's leading citizens.
c. ? I'll be happy if carrying the flag is one of the town's leading citizens.

A number of other constructions seem to follow this pattern; other examples of "main clause phenomena" (MCPs), as they are known, will be presented below.

Several explanations have been put forth for main clause phenomena. Emonds (1969, 1976), who was the first to study them in depth, suggested a purely syntactic explanation: the transformations needed to produce such constructions can only occur in "root" clauses (that is, main clauses). Other syntactic explanations of main clause phenomena have been put forth by den Besten (1983), Heycock (2005), and Haegeman (2010). Others have looked to pragmatic factors to explain main clause phenomena. Hooper and Thompson (1973) propose that main clause constructions are permissible in clauses that constitute "assertions"; this includes main clauses, but it also includes some embedded or adverbial clauses. Thus (3) is grammatical, while (1b) above is not:

(3) She said that mushroom pizza she can't stand.

Green (1976), however, finds the criterion of assertion to be ill-defined in some cases and at odds with the facts in others. Recent surveys of the topic (Aelbrecht, Haegeman, & Nye, 2012; Heycock, 2005) emphasize the unresolved state of these issues; in the words of Aelbrecht et al. (2012, p. 8), the MCP concept "appears to encompass an array of distinct phenomena for which a unified analysis is not readily available."

In what follows, I propose a processing explanation for main clause phenomena. The gist of the explanation is that sentence processing is easier at the beginning of the sentence, and this permits a wide variety of syntactic constructions that might overwhelm the sentence-processing mechanism in non-initial contexts. If some constructions are to be avoided in sentence-internal contexts, it makes sense that the constructions subject to this treatment would be relatively infrequent (even in initial contexts) and therefore presumably less essential for communication; main clause phenomena satisfy this criterion, as we will see. My focus here is on English, though at the end of the study I briefly consider the relevance of the theory to other languages.

To some extent, the predictions of the grammatical and pragmatic accounts of MCPs described earlier align with those of the processing account: Both types of account predict a lower frequency of MCP usage in dependent clauses than in main clauses. However, one prediction of the processing account does not appear to be made by any grammatical or pragmatic theory: that main clause phenomena will occur predominantly in *sentence-initial* main clauses, as opposed to sentence-internal main clauses, such as those with premodifying adjunct phrases (4a) and non-initial coordinate clauses (4b).

(4) a. ? At the parade, *carrying the flag was one of the town's leading citizens.*
b. ? The band played and *carrying the flag was one of the town's leading citizens.*

If this prediction is confirmed, it will strongly suggest that processing factors play a role in main clause phenomena, though it will not rule out a role for pragmatic factors as well. While a complete avoidance of MCPs in non-initial contexts would support the processing theory most strongly, a relative avoidance (i.e., a lower frequency in non-initial contexts than in initial contexts) is also compatible with the theory, as I will explain. The processing account also makes a further prediction that is not made by other accounts: that within the general category of MCPs, constructions that are especially rare will be especially avoided in sentence-internal contexts. Again, the logic is that if certain constructions are to be excluded from sentence-internal usage, it would be desirable to exclude those constructions that are very rare anyway.

I will present a corpus analysis, using text from the *Wall Street Journal*, to explore the use of main clause phenomena in written English and to test my processing explanation for them. The corpus analysis offers tentative support for the current hypothesis, but it is somewhat inconclusive. I then consider another syntactic phenomenon that allows further testing of the general idea that rare constructions are avoided in sentence-internal contexts: premodifying adjunct phrases, sometimes called "openers." While openers in general are very common, some syntactic types of opener are much more common than others; for example, prepositional-phrase openers like (5a) are much more common than predicate noun-phrase openers like (5b).

(5) a. In the park, John chased the dog.
b. A fast runner, John chased the dog.

The current theory predicts that less frequent types should be particularly avoided in sentence-internal contexts, and this prediction is borne out.

Main clause constructions and openers generally convey thoughts that could also be expressed in other, more normative, ways. In the case of main clause constructions, this is made clear by Hooper and Thompson (1973), who show "untransformed" (but still grammatical) versions of each MCP type. For example, sentences (1a) and (2a) above could be rephrased as follows:

(6) a. She can't stand mushroom pizza.
b. One of the town's leading citizens was carrying the flag.

There are subtle pragmatic differences between these sentence pairs—the fronted constituents *mushroom pizza* and *carrying the flag* function as the discourse topic in (1a) and (2a) but not in (6)—but the essential meaning is the same. The idea of MCPs as non-normative or *marked* variants of other syntactic constructions is supported by the corpus data presented below. With openers, too, the information contained in the premodifying phrase can usually be placed elsewhere in the sentence with much the same meaning (compare with (5a) and (5b) above):

(7) a. John chased the dog in the park.
b. John, a fast runner, chased the dog.

This view of MCP and opener usage suggests that these might be regarded as aspects of syntactic choice, similar in some respects to other widely studied phenomena such as dative alternation:

(8) a. I gave John the book.
b. I gave the book to John.

In the case of dative alternation, as with the MCP and opener constructions discussed above, there are two variants of an underlying form, essentially identical in meaning but with different pragmatic implications. Dative alternation has been shown to be influenced by processing factors (Arnold, Wasow, Losongco, & Ginstrom, 2000; Bresnan, Cueni, Nikitina, & Baayen, 2007), as have other aspects of syntactic choice (e.g., optional complementizer usage: Jaeger, 2010), so it is natural to look for such factors in the case of MCPs and openers as well.

The paper is organized as follows. I first present a view of sentence processing that leads naturally to the idea that rare constructions will be avoided in sentence-internal contexts—or at least, raises this as a plausible possibility. I then present two corpus analyses, one of MCP usage and the other of openers. In both cases, I focus especially on predictions that are made by the current view and not by other accounts: (a) that rare constructions will be avoided more in sentence-internal main clauses than in sentence-initial ones, and (b) that within each category (MCPs or openers), the sentence-initial bias will be especially strong for rarer constructions. In the final section of the paper, I consider some further issues and implications of the current study: some possible problems with the current theory (and alternative explanations), its relevance to other languages, and possible connections with domains outside sentence processing.

2. Motivating the sentence-initial bias for rare constructions

In recent years, a number of models of sentence processing have been proposed (Brants & Crocker, 2000; Dubey, Keller, & Sturt, 2013; Frank, 2009; Gibson, 2000; Hale, 2006; Jurafsky, 1996; Levy, 2008; Lewis & Vasishth, 2005; Stevenson & Smolensky, 2006). The theory of sentence-initial bias for rare constructions that I present here does not assume any specific model, but it makes certain assumptions about language and sentence processing that are shared by many of these models (though perhaps not all).

At a purely syntactic level, ordinary natural language is full of ambiguities, meaning that there are often multiple syntactically correct analyses for a given sentence. A long sentence may have hundreds or thousands of interpretations, as ambiguities within short segments of a sentence may combine in many different ways (Church & Patil, 1982). Normally, we are able to identify the correct (intended) interpretation out of all the possible ones; we do this by bringing to bear many kinds of knowledge and information, such as the probabilities of different syntactic constructions (Jurafsky, 1996; Levy, 2008), the

plausibility of different verb-argument combinations (Trueswell & Tanenhaus, 1994), and contextual and pragmatic information (Crain & Steedman, 1985; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995).

It is not possible to parse effectively in a purely "local" fashion—by analyzing small segments of a sentence in isolation. Consider these two sentences:

(9) a. Flying planes can be dangerous, can't they.
b. Flying planes can be dangerous, can't it.

One can only identify the correct interpretation of *flying planes*—plural noun with participle modifier in (9a), gerund phrase in (9b)—by considering the rest of the sentence. In some cases, our preferred interpretation of the first part of a sentence may be revised when additional words are added; that is, some reanalysis may occur (Lewis, 1998, 2000). Because the interpretation of one part of a sentence depends on other parts of it, sentence processing must be "holistic": We must construct and evaluate analyses of the entire sentence (or at least, a sizable portion of a sentence—see below). Some parsers in computational linguistics adopt efficient search procedures that find the optimal analysis without generating them all (Collins, 2003; Stolcke, 1995). This approach is valid if only syntactic knowledge is being considered; however, it becomes unworkable when contextual and pragmatic information is brought to bear (Dubey et al., 2013). The only wide-coverage sentence-processing model that incorporates pragmatic information—that of Dubey et al. (2013)—adopts an "n-best" approach, generating and evaluating complete analyses of the sentence.

There are limits to the holistic, exhaustive nature of parsing. It seems fairly clear that, in processing the thirtieth word of a 40-word sentence, we are not still maintaining multiple interpretations of the first few words. Experimental evidence for this comes from so-called "digging-in" effects. Tabor and Hutchins (2004) asked subjects to judge the grammaticality of sentences such as these:

(10) a. As the author wrote the essay grew quickly.
b. As the author wrote the essay describing Babylon grew quickly.

When *the essay* is first encountered, it could either be the direct object of *wrote* or the subject of the main clause. In general, the direct-object interpretation is favored in such situations, but in these sentences, the word *grew* excludes this interpretation, requiring reanalysis of *the essay* (and also *wrote*, which switches from transitive to intransitive). Subjects were much more likely to judge (10b) as ungrammatical than (10a). This suggests that in (10b), the segment requiring reanalysis was so far back that it could no longer be reanalyzed. Lewis and Vasishth (2005) find similar effects.

Putting these assumptions together leads to a view of sentence processing along the following lines. There is a window, consisting of a few words (perhaps five or six, judging by previous examples), within which some kind of holistic, exhaustive parsing takes place: Multiple analyses are generated and considered. (The "window" idea builds on the model of Frazier and Fodor [1978], though other aspects of their model are not assumed here.) As the sentence unfolds, the window shifts along in small steps (perhaps individual

words, or "chunks" containing small constituents; in reading these might correspond to saccades). Crucially for the current argument, it is assumed that the window size is greater than the step size, so that newly added words at one step remain in the window for the following step (or perhaps several steps). At each step, the newly added words must be identified and integrated with syntactic structures formed for the previous words, which are presumably still available (at least for the words within the window); as noted earlier, this may involve consideration of many analyses, combining possible interpretations of the new chunk with possible interpretations of the previous words still within the window. All this is assumed to occur fluently and effortlessly, part of the normal parsing process. Reanalyzing words outside the window should (according to this view) cause much more difficulty—the kind of "garden-path" effect widely discussed in psycholinguistics (Crain & Steedman, 1985; Lewis, 1998; Sturt, Pickering, & Crocker, 1999).

Under this view, the beginning of a sentence presents a unique situation: The first chunk is processed, consisting of the first word or few words, but there is nothing previous within the window with which the chunk must be syntactically integrated. (The same point might apply, to a lesser extent, to the following few chunks—until the window is completely full.) In general, then, we would expect syntactic processing at the very beginning of a sentence to be easier (requiring less computation) than elsewhere. It therefore makes sense that a language might evolve to allow additional syntactic possibilities at the beginning of the sentence; this causes additional processing complexity, but this is counterbalanced by the reduction in processing load due to the absence of previous syntactic context. (It seems reasonable to suppose that it is optimal for processing complexity to be roughly constant; this general idea has been proposed by Levy and Jaeger (2007), though they focus on a different aspect of processing complexity.) This offers a possible explanation for main clause phenomena. In simple terms, the reason that constructions such as participle preposing and topicalization are used mostly in main clauses is that such clauses are usually sentence-initial and processing is easier at that point, so that the sentence processor can handle more possibilities without being overwhelmed. It is worth restating that MCPs are marked, non-normative constructions (even in main clauses—evidence for this will be presented below), and they convey thoughts that can be expressed in other ways; for this reason, restricting their use in non-initial clauses does not greatly limit the language's communicative power. To put it another way, if a language were to limit the use of certain constructions sentence-internally, MCPs—non-normative constructions that are rare in any case and represent alternative forms of more common constructions—would be natural candidates for such as this.

An example will illustrate this reasoning. As noted above, a sentence beginning with *flying planes* could potentially continue in several different ways, involving different syntactic constructions. Imagine now that this phrase occurs sentence-internally, in a context that introduces additional ambiguity:

(11) I heard flying planes...

A verb such as *hear* can take a direct object (*I heard flying planes last night*), a sentential complement (*I heard flying planes can be quite dangerous*), or an object plus a bare

infinitive (*I heard flying planes roam the sky*). In the sentential complement case, the usual ambiguity of *flying planes* (plural noun or gerund) also arises. There are therefore at least four interpretations of this prefix. Now consider the possibility of a sentential complement that includes a topicalized NP (18a) or a preposed VP (18b):

(12) a. ? I heard flying planes he couldn't stand. (= I heard that he couldn't stand flying planes)
b. ? I heard flying planes above the battlefield were some of the nation's best pilots.

With (12a), there is again an ambiguity between the plural noun and gerund interpretations of *flying planes*. Thus, allowing main clause constructions in this embedded clause context adds at least another three possible interpretations to the prefix *I heard flying planes*—possibly creating an overwhelming processing load. Again, it may be that contextual information would always allow the intended interpretation to be identified; but the many possible interpretations would at least need to be *constructed*, in order to be evaluated, and this could require considerable processing work. (In computational terms, the problem is not just one of *evaluation*, but also of *search*.)

The view of sentence processing presented here is certainly oversimplified in some ways, particularly in its assumption of a sharply defined "window." It is more likely that words have varying levels of activation which decay as they recede in time, making them harder to reanalyze (Lewis & Vasishth, 2005). (One might say that inclusion in the window is a matter of "more-or-less" rather than "all-or-nothing.") And reanalysis difficulty is affected by other factors besides the sheer distance to the reanalysis point, such as the difference in complexity between the old parse and the new one (Gibson, 1991) and the degree of structural difference between them (Lewis, 1998; Sturt, Pickering, & Crocker, 2000). But these points do not weaken the current argument. It remains true that integrating new words with previous words requires some computational work—sometimes involving consideration of a large number of possibilities—and that this work is not required for words at the very beginning of the sentence. Note also that, even if there are relatively few analyses (perhaps only one) that can be formed with the previous context, the processor must still *search* for analyses, and this presumably involves some computation—computation that is only required when prior context is present.

The idea that processing is easiest at the beginning of a sentence may be controversial. Indeed, one might argue the opposite—that processing gets easier as a sentence continues. This might occur because early elements make later ones more predictable, both semantically and syntactically; for example, a transitive verb indicates that there will be a direct object and also provides information about what it is likely to be. Levy (2008) has used this reasoning to explain the fact that adding elements in a verb-final clause can make the verb easier to process. However, my argument is not that sentence processing is easier *overall* at the beginning of the sentence, but rather, that some factors push in that direction. Levy may well be right that words become more predictable as a sentence goes on, and that this facilitates processing. But processing complexity may also be affected by the number of interpretations that must be considered, and this surely increases as a

sentence goes on (though again, it is contained by the "windowing" process). Perhaps processing is optimal when both of these aspects of complexity are kept at a moderate level. Notably, the restriction on sentence-internal MCPs hypothesized here is argued to affect the second factor—the number of interpretations under consideration—so it is natural that it comes into play later in the sentence, counteracting the proliferation of analyses that occurs as the sentence gets longer.

The argument is, then, that common assumptions about sentence processing lead naturally to the prediction that rare constructions will be avoided in sentence-internal contexts. The following corpus analyses of MCPs and openers are designed to test this prediction, focusing especially on specific predictions that distinguish the processing view of rare constructions from pragmatic and grammatical accounts.

3. A corpus study of main clause phenomena

3.1. MCP types

In what follows, I examine the use of main clause phenomena in two corpora of written English, comparing their frequency in sentence-initial and sentence-internal contexts. Below I list seven MCP types that are considered in my corpus analysis, with examples. This list is based on the list presented in Hooper and Thompson (1973); many of the example sentences below are theirs as well. I exclude some of their categories; these are listed in Part I of Appendix S1, with explanations for why I exclude them.

Participle preposing—preposing of a progressive or passive participle phrase with subject-auxiliary inversion.

(13) Standing next to me was the president of the company.

Preposing around *be* (with preposed elements other than participle phrases, again with subject-auxiliary inversion): here I include preposing of a prepositional (8a), adjectival (8b), or adverbial (8c) phrase.

(14) a. Among the attendees were many prominent business leaders.
b. More significant would be the development of a semantic theory.
c. Here is the cake. (occurs mainly with the demonstrative adverb *here*)

Negative inversion—preposing of a negatively marked adverbial or prepositional phrase such as *never* or *at no time*, with subject-auxiliary inversion.

(15) Never have I seen such a crowd.

Directional adverb preposing and prepositional phrase substitution. These two constructions involve preposing of an adverbial (16a) or prepositional (16b) phrase with an inverted verb that does not normally invert, such as *come*, usually a verb of location or motion. Following recent practice (Heycock, 2005), I group these two cases together and call them locative inversion.

(16) a. Here comes Fred.
b. On the wall hangs a portrait of Mao.

Subject replacement—infinitive phrases (17a) or finite clauses with *that* (17b) in subject position. I analyze these two cases separately, calling them subject-position infinitives and subject-position clauses, respectively.

(17) a. To read so many magazines is a waste of time.
b. That Henry forgot the key irritated Carmen.

Topicalization—preposing of a direct or prepositional object.

(18) This book you should read.

In the corpus test, I counted tokens of each of the seven MCP constructions listed above in four different clause environments, as follows (the examples show ordinary—not MCP—clauses):

Sentence-initial main clauses: *The dog ran.*
Embedded clauses (i.e., verb complement clauses), with or without a complementizer: *He said (that) the dog ran.*
Non-initial, top-level coordinate clauses (preceded by a coordinating conjunction such as *and* or *but*): *The cat jumped and the dog ran.* (I excluded coordinate structures within dependent clauses.)
Sentence-initial main clauses with premodifying adjunct phrases: adverbial phrases (*Quickly the dog ran*), prepositional phrases (*At the sound of the gun, the dog ran*), adjective phrases (*Furious at the cat, the dog ran*), participle phrases (*Holding the bone, the dog ran*), infinitive phrases (*To escape the fox, the dog ran*), or adverbial clauses (*When the cat jumped*, *the dog ran*).

This list of clause environments is far from exhaustive: Other types of finite clauses include main clauses after colons and semi-colons, adverbial clauses, noun complement clauses (*The fact that*...), and relative clauses. However, *all* tokens of MCPs in the corpus were extracted; 320 tokens were found, and a large majority of these (87%) were found to fall into one of the four environments listed above.

The processing model presented earlier asserts that main clause phenomena occurring in sentence-internal contexts will incur processing difficulty; we therefore predict that they will occur less frequently in such contexts than in sentence-initial contexts. The crucial comparison is between sentence-initial main clauses and non-initial main clauses—that is, non-initial (top-level) coordinate clauses and premodified main clauses. Grammatical and pragmatic accounts of MCP usage do not predict any difference in MCP usage between these environments. Indeed, Hooper and Thompson (1973, p. 466) affirm that "coordinately-joined sentences qualify as roots," as does Heycock (2005). The processing theory does predict a difference, since main clauses that are preceded by other material require integration with that preceding context.

As noted earlier, Hooper and Thompson (1973) propose that MCPs may occur in embedded clauses, but only when they express assertions; while this criterion is difficult to assess (Green, 1976; Heycock, 2005), it seems to follow from this theory that MCPs would be less common in embedded clauses than in main clauses, since at least some embedded clauses presumably do not involve assertions. The processing theory also predicts avoidance of MCPs in embedded clauses, since they are never sentence-initial. In this respect, the processing theory's predictions align with those of Hooper and Thompson's theory; the results of this aspect of the study are not decisive between the two theories. The inclusion of the embedded clause environment was more exploratory in nature, to see if any interesting patterns emerge. It would have been desirable to examine adverbial clauses as well, but limitations of the data suggested that this would not be fruitful; see Part II of Appendix S1 for details.

3.2. Corpus analysis

My initial corpus was the Wall Street Journal portion of the Penn Treebank (I will call this the *hand-parsed WSJ corpus*) (Marcus, Santorini, & Marcinkiewicz, 1994); this was then supplemented by a larger corpus, as explained below. The hand-parsed WSJ corpus consists of about 1 million words (48,000 sentences) of natural English text from the 1989 *Wall Street Journal*. The text is manually annotated with syntactic constituents and preterminals (parts of speech). The definitions of MCP types and environment types in terms of Penn Treebank notation are shown in Part II of Appendix S1.

In the corpus analysis, I counted the number of occurrences of each MCP in each clause environment; I also counted the total number of tokens of each environment type. These figures are shown in Table 1 ("WSJ1" is the hand-parsed WSJ corpus). The MCP types are shown in descending order of frequency in the sentence-initial environment. It can be seen that, in all cases, the number of tokens of each MCP is quite low; no MCP type occurs in even 0.5% of tokens of any kind of environment. This confirms the general assumption that MCPs are "marked" constructions—alternative forms of more common constructions. However, the low frequency of MCPs makes it difficult to estimate their relative frequency in different environments. Following a common practice, I decided that the number of tokens was sufficient only in cases where the expected count of the construction (based on its count in the sentence-initial environment) was greater than five in each environment type; only two MCP types—participle fronting and *be* inversion—met this criterion.

In order to obtain a larger number of tokens, I employed a larger corpus—a set of about 860,000 sentences from the 1987 *Wall Street Journal* (BLLIP 1987–89 WSJ Corpus Release 1, Linguistic Data Consortium), hereafter the *automatically parsed WSJ corpus*. This corpus was parsed automatically by a parser that was trained on the hand-parsed WSJ corpus and thus uses the same labeling conventions; about 91% of constituents are labeled accurately (Charniak & Johnson, 2005). Given this fairly high error rate, it seemed preferable to use the results from the hand-parsed corpus for the cases where there was sufficient data (participle preposing and *be* inversion); the larger corpus was only used for the other five MCP types. The data were also manually filtered to

Table 1
Frequency of main clause constructions in Wall Street Journal text

		Environment			
MCP Type	Corpus	Initial	Coord.	Emb.	Premod.
Pre-part.	WSJ1	83	0 (5.4)	1 (26.1)	0 (25.1)
Be-inv.	WSJ1	132	1 (8.5)	3 (41.6)	1 (40.0)
Locative	WSJ1	30	3 (1.9)	0 (9.4)	2 (9.1)
	WSJ2	273	17 (17.6)	3 (85.9)	12 (82.7)
Negative	WSJ1	13	2 (0.8)	0 (4.1)	0 (3.9)
	WSJ2	164	8 (10.6)	2 (51.6)	2 (49.7)
Inf-subj.	WSJ1	13	1 (0.8)	0 (4.1)	2 (3.9)
	WSJ2	202	4 (13.0)	29 (63.5)	17 (61.2)
Clause-subj.	WSJ1	6	0 (0.4)	0 (1.9)	0 (1.8)
	WSJ2	77	0 (5.0)	0 (24.2)	1 (23.3)
Topic.	WSJ1	6	0 (0.4)	1 (1.9)	0 (1.8)
	WSJ2	41	2 (2.6)	0 (12.9)	2 (12.4)
Total environment tokens	WSJ1	35,426	2,285	11,153	10,731
	WSJ2 (est.)	630,229	40,650	198,412	190,904

Notes. WSJ1 = hand-parsed WSJ corpus, WSJ2 = automatically parsed WSJ corpus, initial = initial main clause, coord. = non-initial coordinate clause, emb. = embedded clause, premod. = premodified main clause, pre-part. = preposed participle, be-inv. = inversion around *be*, locative = locative inversion, negative = negative inversion, inf-subj. = subject-position infinitive, clause-subj. = subject-position clause, topic. = topicalization. Numbers in parentheses indicate expected count in each non-initial environment, based on initial environment count.

remove false positives (see Part II of Appendix S1 for details). The results are shown in Table I ("WSJ2"). The counts of environment tokens for the larger corpus were estimated from the smaller corpus. The counts used in the following analysis—either from WSJ1 or WSJ2—are surrounded by bold lines.

Numbers in parentheses in Table 1 show the expected count of each MCP type in each non-initial environment, based on its observed count in the sentence-initial environment. It can be seen that, in all 21 of the cases considered (7 MCP types × 3 non-initial environment types), the frequency of MCPs in non-initial environments is lower than the expected frequency. Fig. 1 shows the ratio of MCP frequency to expected frequency for each MCP type and non-initial environment. For each MCP type and each non-initial environment type, I performed a chi-square test to see if the proportion of environment tokens containing the MCP was significantly different from the proportion of sentence-initial main clauses containing it; the results are shown in Fig. 1. In all but three cases, the difference is significant: The exceptions are locative inversion, negative inversion, and topicalization in non-initial coordinates. Also of interest is the case of subject-position infinitive phrases; in this case, although the frequency of the MCP is significantly

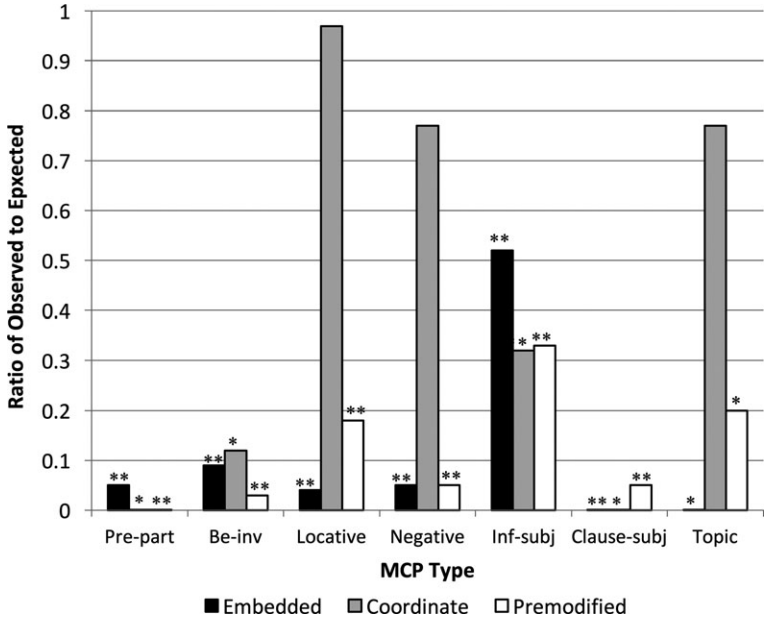

Fig. 1. Ratio of observed frequency to expected frequency for main clause constructions in non-initial environments. *Observed count is significantly lower than expected count, $p < .05$; **observed count is significantly lower than expected count, $p < .0001$. See Table 1 for abbreviations.

below the expected value in all three environments, its ratio to the expected value in the embedded-clause environment is over 1:2, much higher than that of other MCPs in this environment. Over the four environments considered, the overall probability of an MCP occurring (the total count of MCPs in the environment divided by the count of the environment, with counts adjusted to combine data from the two corpora) is highest for main clauses (0.0073), non-initial coordinate clauses (0.0012), embedded clauses (0.00053), and then premodified clauses (0.00027).

The fact that MCPs occur less often in sentence-internal clauses than in sentence-initial clauses in all of the cases considered—and significantly less often in 18 of the 21 cases—provides support for the processing view of MCP usage. Particularly notable is the fact that this pattern is found with non-initial coordinate clauses and premodified main clauses, since this is not predicted by pragmatic or grammatical accounts. However, the current test is not a completely convincing argument for the processing account, for several reasons. First, the degree of avoidance of sentence-internal MCPs varies greatly depending on the MCP and the environment. Certainly, other factors besides the current processing theory—factors currently unknown—will be needed to explain the distribution of MCPs; it seems possible that these factors, once known, will explain MCP usage completely without the need for any general principle of sentence-internal avoidance.

The processing account of MCPs is also open to criticism, particularly in the case of premodified clauses. MCPs sometimes provide information that could also be placed in a

premodifying phrase; for example, a fronted-participle sentence such as *Standing next to me was the president of the company* could be rephrased as *Standing next to me, the president of the company*... (some other predicate would be needed). To some extent, then, MCPs may remove the need for premodifying phrases. In general, premodifying phrases are thought to serve a different pragmatic function from fronted arguments—premodifying phrases are associated with *thematicization*, or "setting the scene" (Halliday, 2004; Hasselgård, 2010), and fronted arguments with defining the primary topic of the clause (Hooper & Thompson, 1973). Still, there may be cases where MCPs and premodifiers are complementary in their distribution. Another problem with the current test is that it relies on automatically parsed data which are only about 91% accurate. Since these data were produced by a probabilistic parsing algorithm trained on statistical data, one might expect its accuracy to be lower on rare syntactic constructions (since it will have seen fewer tokens of such constructions in training). A final limitation of the current test is that it does not allow testing of a key prediction of the processing view of MCPs: namely, that the avoidance of MCPs in sentence-internal environments should be strongest for MCP types that are especially rare. While some MCP types are more common than others, the very small number of tokens of all of the MCP types in non-initial environments makes it impossible to determine with any confidence whether the sentence-initial bias is stronger for the rarer MCP types.

Taken together, these concerns raise doubts about the processing-based explanation of MCP usage. The explanation would be more convincing if the sentence-initial bias for rare constructions could be demonstrated in other contexts. The following section explores this possibility.

4. A corpus analysis of openers

Another opportunity for testing the theory of sentence-initial bias for rare constructions is provided by premodifying adjunct phrases or "openers." Like main clause phenomena, openers occur at the beginning of a finite clause; in a sentence-initial main clause, they will normally occur right at the beginning of the sentence. Openers take a variety of syntactic forms; eight of these syntactic types will be considered here. Unlike MCPs, openers in general are quite common; the corpus analysis presented in the previous section—in which clauses with premodifying adjunct phrases were treated as an environment—found over 10,000 tokens of them in the hand-parsed WSJ corpus, or roughly one in every five sentences. Thus, the current view does not predict a strong avoidance of openers in non-initial contexts. However, the syntactic types of opener vary greatly in their frequency, and some are quite rare; the current theory predicts that these types should be particularly infrequent in sentence-internal contexts (relative to their frequency in sentence-initial contexts).

There has been little corpus research on premodifying adjunct phrases in English. Hasselgård (2010) offers corpus data on "adjunct adverbial" phrases in the clause-initial position (and other positions); these correspond roughly with premodifying phrases as defined

here. However, Hasselgård defines adjunct adverbials in semantic/pragmatic terms, rather than syntactically. Some main clause constructions (such as fronted prepositional phrases with subject-verb inversion) are considered adjunct adverbials in Hasselgård's study; other kinds of phrases treated as openers here—such as participle and adjectival phrases—are excluded. While Hasselgård's semantic/pragmatic perspective is certainly useful, our focus here is on syntactic factors that affect sentence production, so it makes sense to define constructions in syntactic terms.

4.1. Types of opener

I define an opener to be anything within a finite clause that precedes the subject (clauses with subject-verb inversion are excluded). Eight types of opener are considered, with examples; in parentheses is the type of word that heads the phrase.

Prepositional phrase (preposition): In the park, John chased the dog.
Adverbial phrase (adverb): Quickly, John chased the dog.
Adverbial clause (subordinating conjunction): While Mary watched, John chased the dog.
Temporal noun phrase (noun): Last week, John chased the dog.
Participle phrase (participle): Holding a stick, John chased the dog.
Infinitive phrase (infinitive marker *to*): To entertain Mary, John chased the dog.
Adjective phrase modifying the subject (adjective): Furious, John chased the dog.
Predicate noun phrase (noun): A fast runner, John chased the dog.

Openers that do not fall into one of these eight categories are extremely rare (see Part III of Appendix S1 for details). While this particular way of categorizing openers is perhaps open to debate, it seems to conform fairly well with conventional wisdom; it also finds support in the Penn Treebank itself, which uses distinct syntactic labeling conventions for each of the eight types (see Part III of Appendix S1). Note also that each of the eight types is headed by a different syntactic category. The one exception to this is predicate noun phrases and temporal noun phrases: The two types are both noun phrases, headed by nouns, and thus, one might argue, syntactically the same. However, inspection of temporal noun phrases in opener contexts shows that the vast majority of tokens fall into a small number of idiomatic phrases, such as *today/yesterday* or *last/this/next week/month/year* (these phrases alone account for 54.3% of all tokens), some of which are only marginally grammatical as ordinary noun phrases (such as *today* and *last week*). (Predicate and temporal noun phrases are also annotated quite differently in the Penn Treebank.) It seems plausible that there are special mechanisms for recognizing and parsing such phrases.

Intuition suggests that some of these opener types are much more common than others; prepositional openers (*In the park, John chased the dog*) seem absolutely routine, while predicative noun-phrase openers (*A fast runner, John chased the dog*) feel much more marked. And rare opener types in non-initial environments seem quite strange (a comma after *but* might help slightly):

(19) ? John was tired, but a fast runner, he chased the dog.

The aims of the corpus analysis were, first, to determine the frequency of each opener type in sentence-initial environments, and second, to determine whether the less frequent opener types were particularly avoided in sentence-internal contexts. Three environments were examined, all of which were considered in the MCP study as well: the sentence-initial, main clause environment; the non-initial, top-level coordinate clause environment; and the embedded clause environment. As with the MCP study, the crucial comparison is between the sentence-initial environment and the non-initial coordinate environment. No other theory besides the current one seems to predict a difference in the frequency of specific opener types between these two environments. (One possible exception is the theory of dependency length minimization; we will consider this in Section 5.) As for embedded clauses, these are fundamentally different from main clauses—syntactically, semantically, and pragmatically—and differences in opener frequency between these two environments might be due to a variety of factors. Still—as with the MCP test presented earlier—it seemed interesting to include embedded clauses in the analysis, in an exploratory way, partly because they are very common and thus yield a large amount of data. If the avoidance of rare opener types is stronger in embedded clauses than in main clauses, this could be taken as adding support to the processing theory—though only weakly, perhaps, given the many other factors that might be involved.

4.2. Corpus analysis

As with the MCP analysis, the current corpus study uses the hand-parsed WSJ corpus, supplemented by the automatically parsed WSJ corpus in cases where counts in the hand-parsed corpus are insufficient. In this case, however, the syntactic annotations in the automatically parsed WSJ corpus—which, as noted earlier, have a substantial error rate—were not used. Rather, I identified openers in the automatically parsed corpus by searching for surface strings of words and part-of-speech tags (these tags are more than 99% correct). For example, infinitive-phrase openers in the coordinate environment were identified by searching for the pattern "CC to . . .," where CC is a coordinating conjunction (usually *and* or *but*); tokens found by this search were then hand-filtered to exclude false positives. This strategy required some limiting of the search. Most notably, in the case of NP-predicate openers, it was not practical to search for a CC followed by any word that could possibly begin a noun phrase. Since more than half of the NP-predicate tokens in the hand-parsed corpus begin with the word *a* (or *an*), the search was limited to these (though the overall count of the NP-predicate type in sentence-initial environments, as reflected in the hand-parsed corpus, was retained as a measure of its frequency, as explained below). See Part III of Appendix S1 for more details.

Table 2 shows the counts of each of the eight opener types in each of the three environments. Again, bold lines in the table indicate the numbers used in the following statistical analysis. For the three rarest opener types—infinitive phrases, adjectival phrases, and predicate-NP phrases—counts from the automatically parsed corpus were used for non-

Table 2
Frequency of opener types in Wall Street Journal text

Environment		Initial		Coord.			Emb.		
Opener Type	Corpus	Count	Prob.	Count	Prob.	P_c/P_i	Count	Prob.	P_e/P_i
PP	WSJ1	4563	.10624	89	.03840	0.361	194	.04311	0.406
Adverbial	WSJ1	1984	.04619	54	.02330	0.504	16	.00356	0.077
Adv. clause	WSJ1	1322	.03078	48	.02071	0.673	165	.03667	1.191
Temporal NP	WSJ1	432	.01006	9	.00388	0.386	10	.00222	0.221
Participle	WSJ1	341	.00794	3	.00129	0.163	3	.00067	0.084
Infinitive	WSJ1	103	.00240	3	.00129	0.101	1	.00022	0.013
	WSJ2			10	.00024		67	.00084	
Adjective	WSJ1	41	.00095	0	0	0.025	0	0	0.030
	WSJ2			1	.000024		1	.000012	
Predicate NP	WSJ1	27	.00063	0	0		0		
Predicate NP	WSJ1	17	.00040	0	0	0	0	0	0
with *a/an*	WSJ2			0	0		0	0	
Env. count	WSJ1	42,949		2,318			4,500		
	WSJ2 (est.)			41,260			80,100		

Notes. WSJ1 = hand-parsed WSJ corpus, WSJ2 = automatically parsed WSJ corpus, initial = sentence-initial environment (env.), coord. = non-initial coordinate env., emb. = embedded-clause env., P_i = prob. in sentence-initial env., P_c = probability in non-initial coordinate env., P_e = probability in embedded clause env.

initial environments (counts of the non-initial environments were estimated from the hand-parsed corpus). Our focus is on the frequency of the eight types in relation to one another, and how this varies across environments. The leftmost column of numbers shows that, as predicted, the eight types vary greatly in frequency in sentence-initial environments (they are shown in descending order of frequency); prepositional openers are nearly 200 times more frequent than predicate-NP openers. The second column of numbers shows the conditional probability of each type in the sentence-initial environment: its count divided by the overall count of the environment. What is of particular interest is the probability of each type in the coordinate environment, relative to its probability in the sentence-initial environment; this ratio is shown by the column "P_c/P_i." It can be seen that the ratio decreases sharply as the sentence-initial probability of the type decreases: that is to say, there is a particular avoidance of the rarer opener types in non-initial contexts. This is exactly as predicted by the current processing theory. Fig. 2 shows the log of the ratio between coordinate-environment probability and sentence-initial probability for each opener type, plotted against the log probability of the type in sentence-initial contexts. The correlation between the two variables is positive and significant, $r = .88$, $p = .003$.

In Fig. 2, the log probability ratio for NP-predicate openers is defined by the counts of tokens beginning with *a* or *an*, but the sentence-initial probability of the type is defined

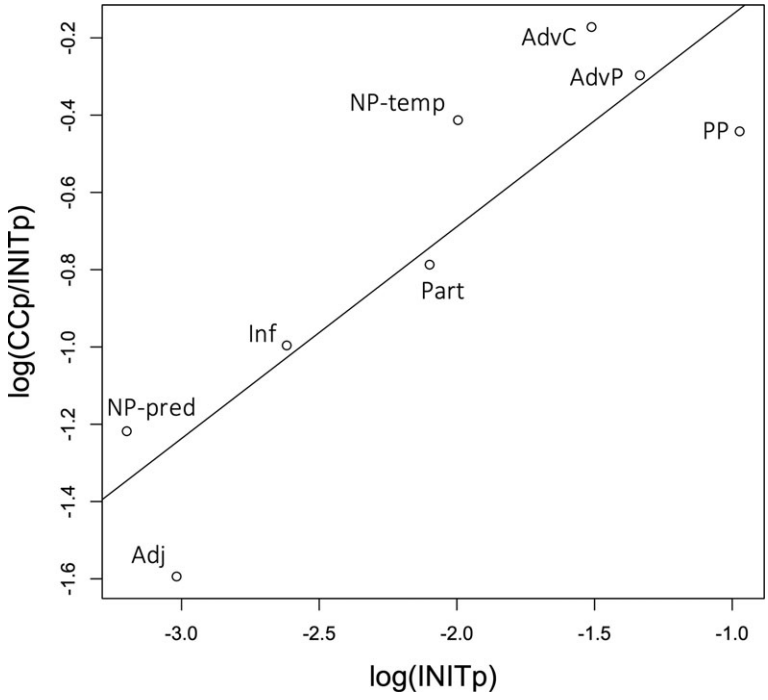

Fig. 2. Log (base 10) ratio of coordinate-environment probability to sentence-initial probability, log(CCp/INITp), plotted against log of sentence-initial probability, log(INITp), for the eight opener types. PP, prepositional phrase; AdvP, adverbial phrase; AdvC, adverbial clause; NP-temp, temporal NP; Part, participle phrase; Inf, infinitive phrase; Adj, adjectival phrase; NP-pred, predicate NP.

by the count of all tokens. The actual count for NP-predicate openers in the coordinate environment is zero, but this is adjusted to one in the calculation of the log probability ratio. In part, this is to facilitate the correlation analysis (since a probability of zero would yield an infinite log probability ratio), but there is a more principled reason for it as well. We wish to use the count that gives the best estimate of the log probability of the construction, given the coordinate environment. There is no reason to assume that the actual probability of the construction is zero; and if it is nonzero, a count of one yields a closer estimate to the actual log probability than a count of zero.

Table 2 also shows the same figures for the embedded clause environment (see also Fig. 3). The patterns observed with regard to the coordinate environment are apparent, though less strong. In general, the probabilities of opener types are lower in embedded-clause environments than in initial environments, and their probabilities in embedded-clause environments tend to decrease as their initial-clause probabilities decrease. But there are exceptions: In particular, adverbial-clause openers are much more frequent than adverbial-phrase openers, whereas the reverse is true in initial contexts; and adverbial-clause openers are slightly more probable in embedded clause environments than in initial environments. Still, the two least common opener types in the initial environment are also

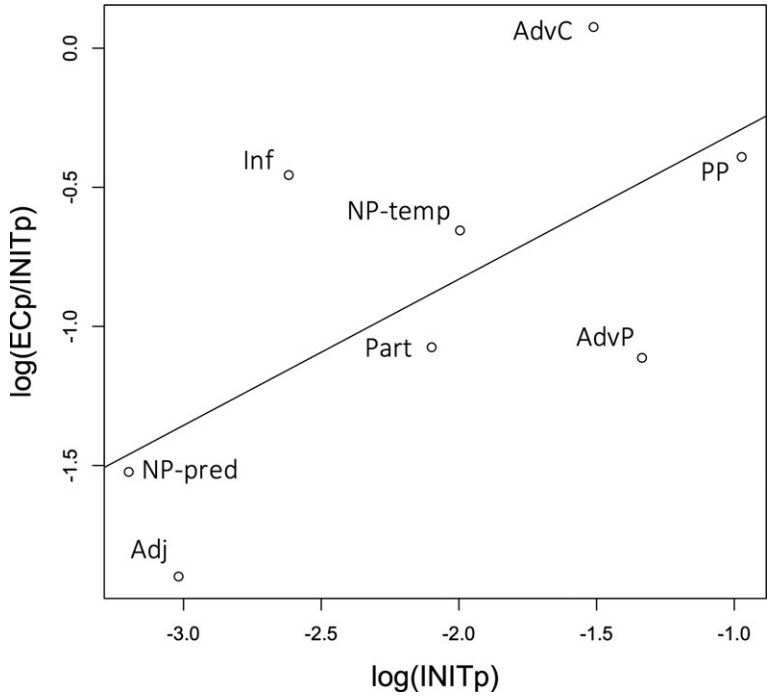

Fig. 3. Log (base 10) ratio of embedded-clause probability to sentence-initial probability, log(ECp/INITp), plotted against log of sentence-initial probability, log(INITp), for the eight opener types. See Fig. 2 for abbreviations.

the least common types in the embedded environment, and also show the lowest ratio of embedded-clause probability to sentence-initial probability. The correlation between the log of initial probability and the log ratio between embedded and initial probability is not significant, $r = .65$, $p = .08$.

5. Discussion

Previous accounts of main clause phenomena have explained their distribution in grammatical or pragmatic terms. The current paper suggests an alternative explanation, based on processing considerations: MCPs are avoided sentence-internally because they exacerbate the combinatorial explosion of possible syntactic analyses that arises in sentence-internal contexts. While both grammatical/pragmatic and processing accounts of MCPs predict their avoidance in embedded clauses, only processing accounts predict their avoidance in non-initial main clauses: premodified main clauses and non-initial, top-level coordinate clauses. A corpus analysis shows that, indeed, MCPs are less common in non-initial main clauses than in sentence-initial clauses: This emerges consistently for seven different MCP types in both premodified and non-initial coordinate clauses (and is

statistically significant in 11 of the 14 cases examined). This test is not entirely convincing as evidence for the processing theory, however, for several reasons: (a) There is enormous variation in the strength of the sentence-initial bias across MCPs which the processing theory cannot explain; (b) there is another possible explanation for the bias in the case of premodified clauses (the MCP may sometimes serve the same function that a premodifier would); and (c) the test relies on automatically parsed data that have a significant error rate.

To further test the prediction of sentence-initial bias in rare constructions, another syntactic phenomenon was examined: premodifying adverbial phrases ("openers"). While openers are common in general, some syntactic types of opener are far more common than others; the prediction was that the rarer types of opener would be more strongly avoided in sentence-internal contexts. A corpus analysis of eight opener types shows that, indeed, the sentence-initial bias of opener types (reflected in the ratio between their probability in non-initial coordinate clauses and their probability in sentence-initial main clauses) increases as their overall frequency decreases, a statistically significant trend. Unlike the MCP test, the opener test does not rely on automatically parsed data, and it also yields a more convincing pattern of results. While the four most common opener types do not show a pattern of increasing sentence-initial bias with decreasing frequency, these types are arguably so common that no difference in sentence-initial bias would be expected; the four less common types all show stronger sentence-initial bias than the four more common types, and among the four less common types, the pattern of increasing sentence-initial bias with decreasing frequency is quite consistent (see Table 2, "P_c/P_i"). Thus, the current processing theory is rather successful in explaining the pattern of sentence-initial bias in openers; this gives it added credence as an explanation for sentence-initial bias in MCPs as well.

In both the MCP and opener tests, the embedded clause environment was examined as well as non-initial main clause environments. This was done largely for exploratory reasons, simply to see if any interesting patterns emerged. It was expected that rare constructions would be less common in embedded clauses than in initial main clauses (in the case of MCPs, both the processing and pragmatic/grammatical theories predict this), and this is indeed the case. The fact that MCPs are (overall) less common in embedded clauses than in non-initial coordinate clauses suggests that syntactic or pragmatic factors, as well as processing factors, may exert pressure against them. (They are more common in embedded clauses than in premodified main clauses, which argues against this view—but, as noted earlier, there may be other reasons for their avoidance in premodified main clauses.) In the case of openers, the embedded-clause environment shows some tendency for sentence-initial bias to increase as opener-type frequency decreases, but the trend is less regular than with non-initial coordinate clauses and not significant.

In all of the tests reported here, there are significant differences in frequency between rare constructions that are not explained by the processing theory. In the first place, there are large differences in overall frequency between constructions, regardless of environment; some MCP types and opener types are much more common than others. No doubt this is partly due to the frequency of the underlying semantic structures—the messages that people want to convey—which is beyond our scope here. There are also differences

in the strength of the sentence-initial bias among constructions that are not explained by the processing theory. While these, too, are largely mysterious, a few observations may be made. In the case of openers, both adverbial clauses and infinitive phrases are much more common in the embedded-clause environment than other opener types of similar overall frequency. I suggest this may be because they often express causal relationships, as in these sentences from the hand-parsed WSJ corpus (the first with an adverbial clause, the second with an infinitive phrase); the embedded clauses are italicized.

(20) a. Lawyers worry that *if they provide information about clients, that data could quickly end up in the hands of prosecutors.*
b. Meanwhile, Shearson Lehman's Mr. Devario said that*, to stay competitive, the U.S. paper industry needs to catch up with the European industry.*

In newspaper text—at least, in the "hard news" sections that predominate in the corpora used here—writers may be reluctant to assert causal relationships, since that would go beyond the purely factual perspective that is supposed to characterize such text; instead, they tend to attribute them to other people, often using an embedded clause (as in (20) above). With regard to MCPs, a notable outlier is the high frequency of subject-position infinitives in embedded clauses, seen in sentences such as this one:

(21) In his personal diary, Hayes wrote *that to abandon principle in the face of this congressional tactic would be to violate a public trust.*

This construction often carries a tone of moral approbation, as is clear in the sentence above. Again, it is not surprising that newspaper reporters tend to attribute such sentiments to others rather than stating them directly.

In the case of openers, an alternative explanation comes to mind for the patterns observed here: the avoidance of long dependencies. Dependency length minimization (DLM) has been found to be an important principle in language production, affecting both grammar and syntactic choice (Ferrer i Cancho, 2006; Futrell, Mahowald, & Gibson, 2015; Gildea & Temperley, 2010; Hawkins, 1994; Temperley, 2007). In the case of both non-initial coordinate clauses and embedded clauses, it is generally assumed that the head of the clause (the finite verb) makes a dependency connection to an external head to the left. The presence of a premodifying phrase will lengthen this dependency; this view therefore predicts that openers should generally be less common sentence-internally than sentence-initially, as they are. (In the dataset used here, 20.5% of sentence-initial clauses have openers, compared to 8.8% of non-initial coordinate clauses and 8.7% of embedded clauses.) One might argue that, with constructions like predicate-NP and adjectival openers, the opener actually connects to the subject NP rather than to the finite verb; but even then, the finite verb presumably connects back to the head (or some other element) of the previous clause, and an opener will lengthen this dependency. DLM makes the further prediction that the sentence-initial bias should be stronger for longer types of opener, since these will cause greater lengthening of the crossing dependency. The average length (number of words) of each opener type, across all three environments, is shown in Table 3. The prediction of DLM fails spectacularly. The longest opener type is adverbial

Table 3
Average word length of opener types in the hand-parsed WSJ corpus

Opener Type	Count	Length
PP	4,846	5.33
Adverbial	2,054	1.48
Adv-clause	1,535	10.50
Temporal NP	451	2.21
Participle	347	9.20
Infinitive	107	6.79
Adjective	41	5.66
Predicate NP	27	7.74

clauses, but this is also the type that shows the *least* sentence-initial bias (see Table 2): It is more common in embedded clauses than in initial clauses, and more than two-thirds as common in non-initial coordinate clauses as in initial clauses. It seems clear, then, that DLM is of little explanatory value here.

One might wonder if the sentence-initial bias of rare constructions could be explained in terms of their discourse function. The discourse function of MCPs has been quite widely discussed; by general agreement, preposing an element such as a prepositional phrase or participle phrase *topicalizes* it in the pragmatic sense, making it the topic of the clause. (This is not to be confused with topicalization in the syntactic sense, which is usually restricted to NPs.) The topic of the clause normally represents given information; the following subject is then a comment on that topic, representing new information. Seen in this way, the use of MCPs follows the widely observed pattern that given discourse elements tend to precedes new ones (Arnold et al., 2000; Bock, 1977). This explanation has been put forth for a variety of the MCP constructions considered here, including participle fronting, inversion around be, and locative inversion (Green, 1980); infinitive and clause fronting (Miller, 2001); and NP topicalization (Prince, 1984). Consider sentence (2a) above, reprinted here:

(22) Carrying the flag was one of the town's leading citizens.

By implication, the reader already knows that someone was carrying the flag; the news is that it was one of the town's leading citizens. Similar arguments have been made with regard to premodifying phrases, though the discourse function of a premodifying phrase is usually described as thematicization rather than topicalization, and the association between discourse function and informational status is weaker; thematic discourse elements often represent given information, but by no means always (Halliday, 2004; Hasselgård, 2010).

While this pragmatic account might explain the *clause*-initial placement of these constructions, what needs to be explained is the fact that they also tend to be *sentence*-initial. Of particular importance is the distinction between sentence-initial clauses and non-initial coordinate clauses: A pragmatic account would need to explain why both MCPs and openers are biased toward the former. Discourse-based theories of MCPs have not generally

distinguished between the two clause types, and some have explicitly treated them as equivalent (Heycock, 2005; Hooper & Thompson, 1973). In addition, a pragmatic account would need to account for the fact that the sentence-initial bias is stronger for rarer opener types. At present, then, it is not clear how the phenomena presented here could be explained in purely pragmatic terms (though one possible explanation along these lines will be considered below).

The current argument is that the avoidance of MCP constructions and rare opener types in non-initial contexts facilitates processing, because it means that they do not have to be considered in such contexts. One problem for this argument is that these rare constructions are not *completely* avoided in non-initial contexts; sometimes they do occur. There are two possible responses to this objection. One possible response relates to the communicative function of these constructions. As noted earlier, the use of MCPs has generally been explained in pragmatic terms, as a way of topicalizing a certain discourse element. It may be that the pragmatic reasons for using an MCP in a sentence-internal context are occasionally so strong that they override the processing pressures against this usage. A similar argument could be made with regard to rare opener types. There is a clear parallel here with other cases of syntactic choice, such as dative alternation. There, too, pragmatic factors play a role in the choice—especially the preference for "given-before-new" ordering—but there are also patterns of usage that seem clearly due to processing factors, such as the preference for "short-long" ordering independent of discourse status (Arnold et al., 2000). In this sense, the view of MCPs and rare opener types that emerges here—as influenced by pragmatic factors but also by independent processing factors—accords well with research on other aspects of syntactic choice.

Another possible explanation for the existence of non-initial MCPs and rare openers comes from a processing viewpoint. While these rare constructions generally seem to be avoided in sentence-internal contexts, there may be certain syntactic contexts in which they are acceptable. According to the current theory, we would expect these contexts to be ones in which the rare construction is easily identifiable to the comprehender. One possible factor here is repetition (also known as priming or parallelism). It is well known that once a syntactic construction is used in a discourse, it tends to be used again—even within the same sentence (Dubey, Keller, & Sturt, 2008; Pickering & Ferreira, 2008). This effect is stronger for rare constructions than for common ones (Reitter, Keller, & Moore, 2011), and it is especially strong within coordinate constructions, in that the second phrase of a coordinate construction is likely to syntactically match the first (Dubey et al., 2008; Temperley & Gildea, 2015). Parallelism is also a factor in comprehension: Constructions that have been primed by the preceding discourse are more easily processed (Frazier, Munn, & Clifton, 2000; Pickering & Ferreira, 2008). This factor may account for some of the uses of sentence-internal rare constructions in the current corpus. Consider sentence (23) below, a coordinate construction in which NP topicalization, a very low-frequency type of MCP, occurs in both the first and second clauses. (The latter clause is one of only two tokens of NP topicalization in a non-initial coordinate.)

(23) "Half of it you don't do, and the other half you get help for," he says.

Perhaps the NP topicalization in the first clause in some way primes this construction so that it is more readily available in the second clause. A similar point could be made about sentence (24) below—the only token found of an adjectival opener in a coordinate environment:

(24) This is not a big voice, but, pliant and perfectly projected, it just floated into Sybil's maw.

In this case, the use of a coordinate structure *within* the adjectival phrase may facilitate the processing of this extremely rare syntactic usage.

The previous discussion of the pragmatic and processing-based factors involved in sentence production has been non-committal as to the actual procedure whereby these factors are brought to bear. While sentence production has been extensively studied (Bock & Levelt, 1994; Dell, 1986; Hartsuiker & Westenberg, 2000; Kempen & Hoenkamp, 1987), research in this area has focused mainly on speech and has been mostly concerned with the internal construction of clauses, as opposed to the larger, multi-clause constructions that are at issue here. With regard to the production of complex written sentences, one possibility is that clauses (represented in an abstract conceptual form) are first assigned positions in a sentence, and then constructed syntactically. Another possibility is that the syntactic structure of each clause is determined first, and then their positions in the sentence. (Of course, each of these accounts is, at best, a huge oversimplification the actual production process, but one or the other may have more resemblance to that process.) The preference for sentence-initial placement of rare constructions could arise under either of these accounts—either by favoring rare constructions in clauses that are already sentence-initial (under the first account), or by favoring sentence-initial placement of clauses containing rare constructions (under the second account). Pragmatic constraints (such as the fronting of "given" discourse entities within the clause) could occur under either scenario as well. Yet a third scenario should be considered, however, which is that the sentential arrangement of clauses and their internal structure are to some extent independent, but both are affected by pragmatic considerations. For example, in constructing a sentence that contains a mention of a previous discourse item, producers may be motivated both to put the clause containing the given element in a sentence-initial position, and to front that item within the clause. In that case, the sentence-initial placement of MCPs might arise as a by-product of these two independent pragmatic mechanisms. There is, at present, little empirical basis to decide between these alternative accounts, and the possibility of a purely pragmatic account of MCP usage cannot be ruled out. It should be emphasized again, however, that the association between pragmatic function and syntactic position is weaker in the case of openers than in the case of MCPs, and also that a purely pragmatic account cannot explain the correlation between opener frequency and sentence-initial bias that has been observed here.

The processing theory of rare constructions also makes a prediction about the *comprehension* of such constructions. If (as the theory predicts) MCPs and rare opener types are

not generally considered in the processing of sentence-internal clauses, then they should cause processing difficulty when they occur (unless, again, there is some special circumstance that facilitates their processing, such as parallelism). As suggested earlier, this can be viewed as a kind of "garden-path" phenomenon, in which the normal parsing process fails. Garden-path situations may result in a complete parse failure—the sentence is simply not understood—or they may trigger reanalysis of the portion of the sentence outside the processing "window," requiring unusual time and effort. Perhaps garden-path situations also cause the comprehender to consider constructions within the processing window that were not considered in the initial analysis, such as MCPs in non-initial contexts; this would allow the intended interpretation to be found, but (again) only with unusual effort. Inspecting the non-initial MCP tokens, there are several sentences that, indeed, I find somewhat difficult to parse, such as these two:

(25) a. They said contributing to the downward drift was the fact that many professional traders had chosen to square positions ahead of the weekend. [Participle preposing in an embedded clause]

b. To be sure, with a landfill comes the risk of running afoul of ever-tightening federal and state environmental regulations. [Locative inversion in a premodified clause]

(Compare the difficulty of (25a) to the same sentence with the embedded clause in main clause position: *Contributing to the downward drift...*) And this is a situation where we *expect* a non-initial MCP! In context, of course, the preceding discourse may well facilitate the processing of these sentences in some way. And in any case, introspective evidence such as this is hardly decisive; experiments would be needed to determine whether non-initial MCPs really do cause processing difficulty.

The Wall Street Journal corpora used in the current study are stylistically quite diverse—including hard news, movie reviews, sports writing, letters to the editor, and many other things—so they represent a fairly broad spectrum of written text. However, it would certainly be desirable to extend the study to other kinds of production data. The challenge is to find adequately large annotated corpora for analysis, given the rarity of MCP constructions and some opener types. Speech would be another natural area for exploration; here, too, the challenge is to find suitable data for analysis. The syntactic structure of speech tends to be simpler than that of writing; one indication of that is average sentence length, which is much lower in speech (8.2 words in the widely used Switchboard corpus of spontaneous English speech) than in writing (20.9 in the hand-parsed WSJ corpus). For this reason, it seems likely that the use of MCPs is more limited in speech. As an initial test, I examined the frequency of preposing around *be*—the most common MCP type in the hand-parsed WSJ corpus—in the hand-parsed Switchboard corpus. This corpus contains about 900,000 words from naturally occurring telephone conversations (about the same size as the hand-parsed WSJ corpus). Only 21 tokens of the construction were found, as opposed to 138 in the hand-parsed WSJ corpus. It would be difficult to draw any conclusions from so little data.

Main clause phenomena have been widely studied in other languages (see, for example, the essays in Aelbrecht et al., 2012). The processing account of MCPs presented here could well be applied to other languages, with the same prediction: They should be more common in sentence-initial environments than in other environments. To explore this in any depth is beyond our scope here. It is worth noting, however, that a number of MCPs in other languages appear to be completely restricted to main clauses—not just disfavored but prohibited in other contexts (Aelbrecht et al., 2012; Hill, 2007; Miyagawa, 2012). In some cases, there are convincing pragmatic arguments for this restriction, but this does not rule out a role for processing factors as well. It would be interesting to see if these phenomena are restricted to sentence-initial main clauses, or distributionally skewed toward such contexts.

What implications might the current study have beyond sentence processing? Possible connections might arise with other cognitive domains that involve hierarchical structures similar to those in language—especially if these hierarchical structures must be extracted by the perceiver from a linear input and are underdetermined by that input. In that case, it seems generally true that there will be fewer analyses to consider in "internal" rather than "initial" contexts; following the reasoning of the current study, this should result in greater processing difficulty in internal contexts, and in compensatory production mechanisms like those observed here. One possible domain of application for this idea is discourse structure. Several theories of discourse structure assume that discourse elements (usually clauses) form hierarchical structures similar to those in syntax (Hirao, Yoshida, Nishino, Yasuda, & Nagata, 2013; Mann & Thompson, 1988; Van Kuppevelt, 1995). It seems intuitive that there might be more ways of connecting elements in the middle of the discourse than at the beginning, potentially creating processing difficulty, though the higher predictability of non-initial elements (given the prior context) might counteract this—a point that was made earlier with regard to sentence processing as well.

Another domain that deserves consideration is music. It has been suggested that music is mentally represented using dependency trees resembling those found in language, with notes or chords forming head-dependent relations with other notes and chords in a recursive fashion (Lerdahl & Jackendoff, 1983; Patel, 2003). If this is the case, then one can imagine that, in the middle of a piece of music (or a section of a piece), a combinatorial explosion of possible analyses might arise; at the beginning, there would be fewer possible analyses. This then leads to the same prediction put forth here for language: We would expect to see greater limits on the grammatical (or perhaps, stylistically acceptable) possibilities in the middle of a piece, and greater freedom at the beginning. I will not explore this further here but offer it as a question for future investigation. It should be noted that the dependency-based, "combinatorial" view of musical structure is, at this point, quite conjectural and far from universally accepted (which is not the case for sentence structure in language) (for a skeptical viewpoint, see Temperley, 2011). Thus, investigating predictions that follow from this view might be one way of testing its validity.

Returning to sentence processing, this study adds to a growing body of research showing ways in which language has evolved to facilitate comprehension (Gildea & Temperley, 2010; Hawkins, 1994; Levy & Jaeger, 2007; Piantadosi, Tily, & Gibson, 2011). It relates most closely, perhaps, to the topic of ambiguity avoidance. A number of studies have explored whether syntactic choices are influenced by a preference to avoid ambiguity. For the most part, the conclusions have been negative: In situations of choice such as dative alternation and optional complementizer inclusion, ambiguity avoidance does not appear to be a major factor (Arnold et al., 2000; Jaeger, 2010; but see Fedzechkina, Jaeger, & Newport, 2012). This has been attributed to people's remarkable ability to bring context to bear in sentence processing, allowing them to eliminate unintended interpretations and zero in on the correct one and making ambiguity avoidance strategies unnecessary (Wasow, 2015). What sets the current study apart from this prior work is the idea that unintended interpretations may impede processing even if they would not lead to actual misunderstanding, because they complicate—and thus slow down—the search process. We would expect language to evolve mechanisms for reducing this danger; the avoidance of main clause constructions and rare opener types in non-initial main clauses appears to offer a compelling example of this.

References

Aelbrecht, L., Haegeman, L., & Nye, R. (2012). Main clause phenomena and the privilege of the root. In L. Aelbrecht, L. Haegeman, & R. Nye (Eds.), *Main clause phenomena: New horizons* (pp. 1–19). Amsterdam: John Benjamins.

Arnold, J., Wasow, T., Losongco, A., & Ginstrom, R. (2000). Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering. *Language*, *76* (1), 30–55.

Bock, K. (1977). The effect of a pragmatic presupposition on syntactic structure in question answering. *Journal of Memory and Language*, *16*(6), 723–734.

Bock, K., & Levelt, W. (1994). Language production: Grammatical encoding. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 945–984). New York: Academic.

Brants, T., & Crocker, M. (2000). Probabilistic parsing and psychological plausibility. In *Proceedings of the International Conference on Computational Linguistics (COLING 2000)* (pp. 111–117). Saarbrücken, Germany: ACL.

Bresnan, J., Cueni, A., Nikitina, T., & Baayen, H. (2007). Predicting the dative alternation. In G. Boume, I. Kraemer, & J. Zwarts (Eds.), *Cognitive foundations of interpretation* (pp. 69–94). Amsterdam: Royal Netherlands Academy of Science.

Charniak, E., & Johnson, M. (2005). Coarse-to-fine n-best parsing and MaxEnt discriminative reranking. In K. Knight, H. Ng, & K. Oflazer (Eds.), *Proceedings of the 43rd annual meeting of the assocation for computational linguistics* (pp. 173–180). Ann Arbor, MI: Association for Computational Linguistics.

Church, K., & Patil, R. (1982). Coping with syntactic ambiguity or how to put the block in the box on the table. *American Journal of Computational Linguistics*, *8*(3–4), 139–149.

Collins, M. (2003). Head-driven statistical models for natural language parsing. *Computational Linguistics*, *29*(4), 583–637.

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

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*(3), 283–321.

den Besten, H. (1983). On the interaction of root transformations and lexical deletive rules. In W. Abraham (Ed.), *On the formal syntax of the Westgermania* (pp. 47–131). Amsterdam: John Benjamins.

Dubey, A., Keller, F., & Sturt, P. (2008). A probabilistic corpus-based model of syntactic parallelism. *Cognition*, *109*(3), 326–344.

Dubey, A., Keller, F., & Sturt, P. (2013). Probabilistic modeling of discourse-aware sentence processing. *Topics in Cognitive Science*, *5*(3), 425–451.

Emonds, J. (1969). Root and structure-preserving transformations. Doctoral dissertation, MIT, Cambridge, MA.

Emonds, J. (1976). *A transformational approach to English syntax*. New York: Academic Press.

Fedzechkina, M., Jaeger, F., & Newport, E. (2012). Language learners restructure their input to facilitate efficient communication. *Proceedings of the National Academy of Sciences*, *109*(44), 17897–17902.

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

Frank, S. (2009). Surprisal-based comparison between a symbolic and a connectionist model of sentence processing. In N. Taatgen & H. Van Rijn (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 1139–1144). Austin, TX: Cognitive Science Society.

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

Frazier, L., Munn, A., & Clifton, C. (2000). Processing coordinate structures. *Journal of Psycholinguistic Research*, *29*(4), 343–370.

Futrell, R., Mahowald, K., & Gibson, E. (2015). Large-scale evidence of dependency minimization in 37 languages. *Proceedings of the National Academy of Sciences*, *112*(33), 10336–10341.

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

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In Y. Miyashita, A. Mirantz, & W. O'Neil (Eds.), *Image, language, brain* (pp. 95–126). Cambridge, MA: MIT Press.

Gildea, D., & Temperley, D. (2010). Do grammars minimize dependency length? *Cognitive Science*, *34*(2), 286–310.

Green, G. (1976). Main clause phenomena in subordinate clauses. *Language*, *52*(2), 382–397.

Green, G. (1980). Some wherefores of English inversions. *Language*, *56*(3), 582–601.

Haegeman, L. (2010). The internal syntax of adverbial clauses. *Lingua*, *120*(3), 628–648.

Hale, J. (2006). Uncertainty about the rest of the sentence. *Cognitive Science*, *30*(4), 643–672.

Halliday, M. (2004). *An introduction to functional grammar* (3rd ed.), rev. by C. Matthiessen. London: Edward Arnold.

Hartsuiker, R., & Westenberg, C. (2000). Word order priming in written and spoken sentence production. *Cognition*, *75*(5), B27–B39.

Hasselgård, H. (2010). *Adjunct adverbials in English*. Cambridge, UK: Cambridge University Press.

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

Heycock, C. (2005). Embedded root phenomena. In M. Everaert & H. van Riemsdijk (Eds.), *The Blackwell companion to syntax* (pp. 174–209). Oxford, UK: Blackwell.

Hill, V. (2007). Vocatives and the pragmatics–syntax interface. *Lingua*, *117*(12), 2077–2105.

Hirao, T., Yoshida, Y., Nishino, M., Yasuda, N., & Nagata, M. (2013). Single-document summarization as a tree knapsack problem. In T. Baldwin & A. Korhonen (Eds.), *Proceedings of the 2013 conference on empirical methods in natural language processing* (pp. 1515–1520). Seattle, WA: Association for Computational Linguistics.

Hooper, J., & Thompson, S. A. (1973). On the applicability of root transformations. *Linguistic Inquiry*, *4*(4), 465–497.

Jaeger, F. (2010). Redundancy and reduction: Speakers manage information density. *Cognitive Psychology*, *61*(1), 23–62.

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

Kempen, G., & Hoenkamp, E. (1987). An incremental procedural grammar for sentence formulation. *Cognitive Science*, *11*(2), 201–258.

Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, MA: MIT Press.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177.

Levy, R., & Jaeger, T. (2007). Speakers optimize information density through syntactic reduction. In B. Schlökopf, J. Platt, & T. Hoffman (Eds.), *Advances in neural information processing systems (NIPS) 19* (pp. 849–856). Cambridge, MA: MIT Press.

Lewis, R. (1998). Reanalysis and limited repair parsing: Leaping off the garden path. In J. Fodor & F. Ferreira (Eds.), *Reanalysis in sentence processing* (pp. 247–285). Dordrecht, the Netherlands: Kluwer.

Lewis, R. (2000). Falsifying serial and parallel parsing models: Empirical conundrums and an overlooked paradigm. *Journal of Psycholinguistic Research*, *29*(2), 241–248.

Lewis, R., & Vasishth, S. (2005). An activation model of parsing as skilled memory retrieval. *Cognitive Science*, *29*(3), 375–419.

Mann, W., & Thompson, S. (1988). Rhetorical structure theory: Toward a functional theory of text organization. *Text-Interdisciplinary Journal for the Study of Discourse*, *8*(3), 243–281.

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.

Miller, P. (2001). Discourse constraints on (non) extraposition from subject in English. *Linguistics*, *39*(4), 683–701.

Miyagawa, S. (2012). Agreements that occur mainly in the main clause. In L. Aelbrecht, L. Haegeman, & R. Nye (Eds.), *Main clause phenomena: New horizons* (pp. 79–111). Amsterdam: John Benjamins.

Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience*, *6*(7), 674–681.

Piantadosi, S., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences*, *108*(9), 3526–3529.

Pickering, M., & Ferreira, V. (2008). Structural priming: A critical review. *Psychological Bulletin*, *134*(3), 427–459.

Prince, E. (1984). Topicalization and left-dislocation: A functional analysis. *Annals of the New York Academy of Sciences*, *433*(1), 213–225.

Reitter, D., Keller, F., & Moore, J. (2011). A computational cognitive model of syntactic priming. *Cognitive Science*, *35*(4), 587–637.

Stevenson, S., & Smolensky, P. (2006). Optimality in sentence processing. In P. Smolensky & G. Legendre (Eds.), *Toward a calculus of the mind/brain: Neural network theory, optimality, and universal grammar* (pp. 307–338). Cambridge, MA: MIT Press.

Stolcke, A. (1995). An efficient probabilistic context-free parsing algorithm that computes prefix probabilities. *Computational Linguistics*, *21*(2), 165–201.

Sturt, P., Pickering, M., & Crocker, M. (1999). Structural change and reanalysis difficulty in language comprehension. *Journal of Memory and Language*, *40*(1), 136–150.

Sturt, P., Pickering, M., & Crocker, M. (2000). Search strategies in syntactic reanalysis. *Journal of Psycholinguistic Research*, *29*(2), 183–194.

Tabor, W., & Hutchins, S. (2004). Evidence for self-organized sentence processing: Digging-in effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(92), 431–450.

Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632–1634.

Temperley, D. (2007). Minimization of dependency length in written English. *Cognition*, *105*(2), 300–333.

Temperley, D. (2011). Composition, perception, and Schenkerian theory. *Music Theory Spectrum*, *33*(2), 146–168.

Temperley, D., & Gildea, D. (2015). Information density and syntactic repetition. *Cognitive Science*, *39*(8), 1802–1823.

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

Van Kuppevelt, J. (1995). Discourse structure, topicality and questioning. *Journal of Linguistics*, *31*(1), 109–147.

Wasow, T. (2015). Ambiguity avoidance is overrated. In S. Winkler (Ed.), *Ambiguity: Language and communication* (pp. 29–47). Berlin: DeGruyter.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article:

Appendix S1. Supplementary material