

**The time-course of lexical and structural processes
in sentence comprehension**

Harry Tily^{1*}, Evelina Fedorenko^{2*} & Edward Gibson²

¹Stanford University

²Massachusetts Institute of Technology

Send correspondence to the first or second author:

Harry Tily, Stanford University, Department of Linguistics, Margaret Jacks Hall,
Building 460, Stanford, CA 94305-2150

Evelina Fedorenko, MIT, Department of Brain & Cognitive Sciences, 46-3037F,
Cambridge, MA 02139

Emails: hjt@stanford.edu, evelina9@mit.edu

Manuscript dated May 29, 2009; comments welcome

Please do not cite or quote without permission

* These authors contributed equally to this work.

Abstract

Online sentence comprehension involves multiple types of cognitive processes: lexical processes such as lexical access, which call on the user's knowledge of the meaning of words in the language, and structural processes such as the integration of incoming words into an emerging representation. In this article, we investigate the temporal dynamics of lexical access and syntactic integration. Unlike much previous work which has relied on temporary ambiguity to investigate this question, we manipulate the frequency of the verb in unambiguous structures involving a well-studied syntactic complexity manipulation (subject- vs. object-extracted clefts). The results demonstrate that whereas for high-frequency verbs, the difficulty of reading a more structurally complex object-extracted cleft structure relative to a less structurally complex subject-extracted cleft structure is largely experienced in the cleft region, this difficulty is largely experienced in the post-cleft region for low-frequency verbs. We interpret these results as evidence that some stages of structural processing follow lexical processing. Furthermore, we find evidence that structural processing may be delayed if lexical processing is costly, and that the delay is proportional to the difficulty of the lexical process. Implications of these results for contemporary accounts of sentence comprehension are discussed.

Introduction

Comprehension is informed by knowledge about what kinds of events are likely in the world; about the structures that are available, probable, or simple in the language being spoken; and about the meaning and idiosyncratic behavior of individual words and constructions. Understanding how and when language users combine different types of knowledge in incremental online sentence comprehension is a central research agenda in psycholinguistics.

In influential work in the 1980s, Frazier and colleagues developed a series of "syntax first" serial models of parsing, proposing that top-down structural principles guide structural choices, and that syntactic structures are created and selected with relatively little regard for the particular lexical items encountered in the sentence (e.g., Fodor, Bever & Garrett, 1974; Frazier, 1979; Frazier & Rayner, 1982; Frazier 1987). Knowledge of likely thematic relationships between the words would only be used to help resolve ambiguity in a later reanalysis stage if the initially assumed parse turned out to be wrong (Rayner, Carlson & Frazier, 1983; Ferreira & Clifton, 1986; Mitchell, 1987; Ferreira & Henderson, 1990). The finding that preferences for one structure over another can be modulated by semantic plausibility and specific lexical information even early on in the parse (e.g., Ford, Bresnan & Kaplan, 1982; Marslen-Wilson & Tyler, 1987) motivated an alternative family of lexical constraint-based competition models. These are characterized by their ability to let multiple sources of information – including world knowledge and lexical idiosyncrasy – guide structural decisions from the outset (e.g., Trueswell, Tanenhaus & Kello, 1993; Trueswell, Tanenhaus & Garnsey, 1994; MacDonald, Pearlmutter & Seidenberg, 1994; Garnsey, Pearlmutter, Meyers & Lotocky,

1997). In lexical constraint-based models, syntactic knowledge is associated with lexical items, and preferences for one syntactic structure over another are driven by the specific words encountered. Though all contemporary accounts of syntactic comprehension assume some role for both structural and lexical knowledge, they differ in whether lexical information is given a primary role, driving structural choices (e.g., MacDonald et al., 1994; Trueswell, 1996); a secondary role, in reanalysis of initially misparsed structures (e.g., Traxler & Pickering, 1996); or weighted in some combination with syntactic knowledge where neither is dominant (e.g., Tabor, Juliano & Tanenhaus, 1997; Gibson, 2006; Levy, 2008).

Much of the evidence for one of these models over another comes from studies on *temporary ambiguity*, using sentences which are initially compatible with two possible syntactic structures. Researchers manipulate the sources of information available to comprehenders in this early ambiguous input, shedding light on the kinds of constraints that influence comprehenders' choices of syntactic structures early on (e.g., Bever, 1970; Frazier & Rayner, 1982; Crain & Steedman, 1985; Ferreira & Clifton, 1986; Trueswell, Tanenhaus & Garnsey, 1994). Although ambiguity and temporary ambiguity have proved useful tools for comparing models of comprehension, most results can be explained by some member of any of the families of models discussed above (see Van Gompel, Pickering & Traxler (2000, 2001), for discussion). An alternative way to investigate the influences of multiple constraints on interpretation choices is by looking directly at the time-course over which effects are observed in online processing. This can be done even in unambiguous sentences. Because the "first-stage" processes of serial models are purely structural, there is no reason to expect detailed lexical processing

immediately when a word is encountered: lexical semantics are only required upon "second-stage" reanalysis or final interpretation, and so could be retrieved at any time. In contrast, most constraint-based models tacitly or explicitly assume that lexical and syntactic information is available simultaneously, or "interactively" (e.g., McRae et al., 1998; Spivey & Tanenhaus, 1998). Indeed, it has been observed that the use of lexical constraints in the processor represents a more optimal comprehension strategy, and a fully optimal solution would be able to take into account multiple sources simultaneously (see McClelland, 1987). Finally, certain models (e.g., Kintsch, 1988; Boland, 1997; Boland & Blodgett, 2001) assume that lexical information is retrieved prior to selection of a syntactic structure.

Some insight into the temporal ordering of lexical and structural influences in comprehension comes from studies of reading using eye-tracking, by virtue of its fine temporal resolution and the possibility of using multiple types of measure (first fixation time, total fixation time, etc.) to tease apart staged effects. For instance, Clifton, Traxler, Mohamed, Williams, Morris & Rayner (2003) investigated reading times for the disambiguating region in the main clause / reduced relative ambiguity shown in (1).

- (1) a. The witness examined by the lawyer ...
b. The evidence examined by the lawyer ...

In accord with Trueswell et al. (1994), total first-pass reading times for the disambiguating region ("by the lawyer") suggested that lexical information – the animacy of the subject (1a vs. 1b) – determines whether the comprehender assumes a correct

reduced relative or an incorrect main clause analysis ("the witness examined someone"). However, go-past times (which include the duration of regressive eye movements made until the reader continues to the right) showed evidence of difficulty in both conditions relative to an unambiguous baseline ("The evidence that was examined..."). Clifton et al. proposed that go-past times reflected a purely structural preference for the main clause analysis that could not be modulated by lexical semantics and resulted from a separate first-stage process. Yet in a broad survey of other eye-tracking results, Clifton, Staub & Rayner (2007) noted that few other studies have reported a similar discrepancy between the two measures. This may indicate that if structural preferences and lexical information act at independent processing stages, in many structures those stages occur too close together to be teased apart by these methods.

Boland and colleagues (e.g., Boland, 1997; Boland & Blodgett, 2001) have carried out several studies investigating the relative sequencing of comprehension effects attributable to lexical information and discourse congruency over longer regions. Boland and colleagues used words ambiguous between a noun and a verb (2c vs. 2d), in contexts that were biased towards one or the other interpretation (2a vs. 2b).

- (2) a. ... Kate looked at all of Jimmy's pets. (noun-biasing context)
- b. ... Kate watched everything that Jimmy did. (verb-biasing context)
- c. She saw his duck [and chickens ...] (noun target)
- d. She saw him duck [and stumble ...] (verb target)

Each of these ambiguous words has some bias towards a noun or verb interpretation (e.g.,

"duck" is more frequently a noun). An effect of this lexical information was observed on reading times at the word itself: here, "his duck" was read quickly relative to "him duck", suggesting that lexical information about the relative frequency of a homograph pair influences syntactic processing early on (e.g., Tabossi, Colombo & Job, 1987). In contrast, the discourse effect due to a matching or mismatching context also influences syntactic processing (e.g., noun targets are read faster in noun-biasing contexts), but this effect is not observed until one or two words later. Boland & Blodgett (2001) interpret these results as evidence that lexical information is incorporated immediately when the word is read, while discourse-level constraints are only incorporated at a later stage of processing. Consequently, Boland & Blodgett's model holds that structural decisions are influenced by lexical information from the outset, while discourse-level constraints only have an influence at a later stage of processing: there are two, temporally ordered, processes involved.

A number of other findings also indicate that lexical processes are not influenced by context. Tabor et al. (1997) and Gibson (2006) suggest that comprehension of an ambiguous word is initially influenced by the presence of a homophone (the complementizer reading of "that") even in a context that rules out that reading ("the lawyer visited that skilled surgeon"). Fedorenko, Piantadosi, Frank & Gibson (submitted) generalize this result to a larger set of words that are ambiguous between a noun and a verb (e.g., "chair"), showing that the frequency of the member of the homophone pair not compatible with the context nevertheless influences reading times and lexical decision latencies. Taken together, these results provide evidence that some or all lexical retrieval processes precede any structural decisions that are informed by

discourse context, and therefore are not facilitated by contextual information.

In this article, we present further evidence that some stages of lexical processing precede structural processing. Furthermore, we find evidence that structural processing may be delayed if lexical processing is costly, and that the delay is proportional to the difficulty of the lexical process. Rather than manipulating discourse-level or syntactic contexts in which the target ambiguous lexical item appears in order to investigate how structural and lexical factors interact, as Boland and colleagues or Tabor et al. do, we consider a contrast between two unambiguous structures, in which the processing difficulty in the more complex structure is commonly assumed to reflect a purely syntactic difference. We then manipulate a lexical property of one of the critical words in these structures. In particular, we consider the well-studied asymmetry between subject- and object-extracted cleft sentences (henceforth SCs and OCs), such as those in (3).

- (3) a. It was the barber that saw the lawyer in the parking lot. (SC)
b. It was the barber that the lawyer saw in the parking lot. (OC)

Object-extracted cleft sentences and closely related object-extracted relative clauses have been shown to incur greater comprehension difficulty at the verb (“saw” in (3)) than their subject-extracted counterparts (e.g., Wanner & Maratsos, 1978; Ford, 1983; King & Just, 1991; Gibson, 1998; Gordon, Hendrick, & Johnson, 2001; Gordon, Hendrick, & Levine, 2002; Traxler, Morris, & Seely, 2002; Traxler, Williams, Blozis & Morris, 2005; Gennari & MacDonald, 2008). Multiple reasons for this effect have been posited, including a structural preference for initially assuming subject-extracted structures (e.g., Clifton &

Frazier, 1989), a greater memory demand in object-extracted structures due to the greater distance between the verb and object or necessity of incorporating two arguments at the verb (e.g., Ford, 1983; Gibson, 1998) or to interference between the object and subject noun phrases (e.g., Gordon et al., 2001), putatively higher frequency of subject-extracted structures in the sentences tested (e.g., Hale, 2001; Gennari & MacDonald, 2008), greater conceptual ease of identifying with an agent rather than patient viewpoint (MacWhinney & Pleh, 1988), or the fact that the relative order of arguments in the structure parallels the dominant SVO pattern of main clauses in the language (Sheldon, 1974; MacDonald & Christiansen, 2002). Most accounts of this difficulty share at least the assumption that it reflects a structure-building or selection process: there are two different structures, and difficulty arises due to building the more complex, infrequent, or otherwise disfavored structure, or having to later select or build from scratch a new object-extracted structure to replace an incorrectly assumed subject-extracted one.

As well as using the contrast shown in (3) to vary difficulty due to structural processes, we simultaneously investigate difficulty due to lexical retrieval. The speed of lexical retrieval is influenced by properties of the word including length, frequency and prior use (Marslen-Wilson, 1990), the presence of similar words in the lexicon (Luce, Pisoni & Goldinger, 1990) and semantic overlap with recently accessed words (Meyer & Schvaneveldt, 1971). These properties can be manipulated in order to make retrieval take more or less time. However, lexical retrieval is typically investigated using individual words presented out of context or in simple carrier phrases, while work on syntactic processing difficulty usually treats lexical retrieval only as a source of confound, explicitly controlling for correlates of lexical access difficulty (e.g., Ferreira & Clifton,

1986). Here, we simultaneously manipulate both a lexical property (the frequency of the verb) and a structural property (the contrast between SC and OC sentences), as shown in (4).

(4) a. High Frequency / Subject-extracted:

It was Vivian who lectured Terrence for always being late.

b. High Frequency / Object-extracted:

It was Vivian who Terrence lectured for always being late.

c. Low Frequency / Subject-extracted:

It was Vivian who chided Terrence for always being late.

d. Low Frequency / Object-extracted:

It was Vivian who Terrence chided for always being late.

Low frequency verbs (4c, 4d) are expected to cause more lexical retrieval difficulty (Marslen-Wilson, 1990), and OC sentences (4b, 4d) are expected to cause more structural processing difficulty. Logically, the structural and lexical effects could surface in either order. The simple presence of the verb is highly predictable given the context that precedes it, particularly in the object-extracted case (e.g., "It was Terrence who Vivian ...") even if the identity of the verb is not; therefore, an OC structure could be posited at this point even before the verb itself is read, or at least before it is identified with a specific lexical item and its associated semantics. Alternatively, the lexical verb could be fully processed before the syntactic structure is posited; a plausible strategy if the processor relies on probabilistic information associated with specific lexical items in

choosing a likely structure. An intermediate possibility is that the two processes could overlap temporally.

We expect both structural difficulty and difficulty due to lexical access of the verb to surface approximately around the point the verb is encountered (see Grodner & Gibson, 2005). However, the precise temporal ordering of the two types of difficulty should inform models of comprehension. Specifically, if structural difficulty precedes lexical access, it is hard to see how lexical information could be guiding the structural process. On the other hand, if lexical access precedes structural difficulty, it remains logically possible that lexical information, although available, is not being used to guide the structural process. In our experiment, if both main effects of structure and frequency are observed only at the same point around the verb, it would suggest that the lexical and structural processes are simultaneous or overlapping (or that the experimental procedure does not have a sufficiently fine temporal resolution to tell otherwise). Alternatively, if we see one of the two effects around the verb and the other on a following region, we could take this as evidence that the stages of lexical retrieval and syntactic processing are ordered such that one is complete or at least well underway before the other begins.

Experiment

Methods

Participants Forty-eight 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.

Design and materials We created 24 sets of cleft sentences like those in (4), crossing

subject-/object-extraction with verb frequency (high/low). Lexical frequencies for the verbs were extracted from the Google Web-1T corpus (Brants & Franz, 2006), which contains counts observed in 1 trillion words of text taken from the World Wide Web. The high-frequency verbs had a mean frequency of 5775000 / mean log frequency of 14.32, and the low-frequency verbs 97160 / 11.21. The frequencies were significantly different as determined by a paired-samples t-test in both raw ($t(23) = 2.36, p < .05$) and logged counts ($t(23) = 11.74, p < .001$). The pairs were chosen such that the meaning of the verbs was as similar as possible, and the two sets of verbs did not differ significantly in length, either by number of letters ($t(23) = -.36, p = .72$) or syllables ($t(23) = -1.00, p = .33$).

Clefts were used instead of relative clauses in order to allow the use of personal names instead of definite noun phrases. We reasoned that using personal names instead of definite noun phrases would minimize any potential differences in the plausibility of the high and low frequency conditions, given that different verbs must necessarily be used. The gender of the names was balanced: a quarter of the items (6 out of 24) had two male names, a quarter had two female names, a quarter had a male and a female name with the male name in the clefted position, and a quarter had a male and a female name with the female name in the clefted position (see Appendix for a complete list of experimental materials).

In addition to the target sentences, the experiment included 48 filler sentences. The length and syntactic complexity of the fillers was similar to that of the target sentences. The fillers were constructed to involve personal names to make them similar to the target sentences. The stimuli were pseudo-randomized separately for each

participant, with at least one filler sentence separating the target sentences. Each participant saw only one version of each sentence, following a Latin Square design.

After each trial, participants were asked a yes/no comprehension question to ensure that they had read the sentence for meaning. The question appeared on the screen and participants pressed one of two keys to respond.

Procedure Sentences were presented word-by-word in a self-paced moving window paradigm, a standard reading paradigm in which all but one word on the screen are replaced by dashes, and participants press a key to move to the subsequent word (Just, Carpenter & Wooley, 1980). The experiment was presented using the Linger experimental software developed by Doug Rohde (available online at <http://tedlab.mit.edu/~dr/Linger/>). Participants took approximately 20 minutes to complete the experiment.

Results

Our experiment followed a factorial design, crossing SC/OC with high frequency/low frequency of the verb.¹ We ran this analysis first, but because exploration and visualization of the data indicated that using a binary frequency variable rather than a continuous measure was obscuring an important pattern in the data, we also conducted an

¹ The initial selection of the materials was based on estimates from the Google search engine (searching for the past-tense forms of the target verbs and recording the number of hits), and items were constructed such that one member of the meaning-equivalent pair was higher in frequency than the other. In subsequent recoding, we found that for all but two of the verbs, all the high-frequency verbs were more frequent than all the low-frequency verbs in the Google Web-1T corpus. The two exceptions resulted from the fact that the highest low-frequency verb, from item 6, was higher than the lowest high-frequency verb, from item 11. Therefore, in the regression analyses where frequency was treated as a binary predictor, we recoded items 6 and 11, such that for item 6 both verbs in the pair were coded as high-frequency verbs, and for item 11 both verbs were coded as low-frequency verbs. All results were statistically similar with the original coding (or excluding these two items). However, it made more sense to us conceptually to recode the items such that all high-frequency verbs were higher than all low-frequency verbs. The statistical analysis used does not require a fully balanced design.

analysis where frequency was treated as a continuous variable. At the end of the Results section, we discuss some problems with binning continuous predictors.

All analyses reported here were conducted with the lme4 package (Bates et al., 2008) for the statistical language R (R Core Development Team, 2008). To choose final models, we included all main effects and interactions, and removed predictors stepwise until the credibility of any smaller model having generated the data in comparison to the fuller model was above $p=.05$ by the log-likelihood χ^2 test. We did not consider models that contravene the hierarchy principle (i.e. that include interactions without the corresponding main effects). Since degrees of freedom cannot be straightforwardly determined for the coefficients of multilevel regression models, p values were obtained using the function pvals.fnc in the R package languageR (Baayen, 2008). This function uses Markov Chain Monte Carlo simulation to approximate the posterior distribution over possible values of all of the parameters in the model given the data, and reports p values showing for each parameter the proportion of that distribution which crosses zero.

Comprehension question performance Across conditions, participants answered the comprehension question correctly 80.3% of the time, a rate comparable to previously reported results for similar sentences (e.g. Gordon et al., 2002). Table 1 presents the mean accuracies across the four conditions. A multilevel logistic regression with crossed random intercepts for participants and items indicated that questions were answered significantly more accurately for SCs than OCs ($\beta=.89$, $p<2e-7$). There was no significant effect of frequency and no interaction.

<u>Syntactic complexity</u>	<u>Verb frequency</u>	
	High	Low
Subject-extracted cleft	85.1 (2.1)	86.8 (2.0)
Object-extracted cleft	74.7 (2.5)	76.0 (2.6)

Table 1. Mean comprehension accuracies by condition, as percent correct (standard errors in parentheses).

Reaction times For the purposes of the analyses, each sentence was divided into the following regions: (1) the sentence beginning “It was”, (2) personal name and the relativizer “who”, (3) the first critical region consisting of the subject-/object-extracted cleft (e.g., “Terrence chided”/ “chided Terrence”), (4) the second critical region consisting of the two words after the cleft (e.g., “for always”), and (5) the rest of the sentence. We analyzed all the trials, regardless of whether the comprehension question was answered correctly. The statistical data patterns were very similar in the analysis of only the trials where the comprehension question was answered correctly. Reading times more than three standard deviations away from the mean for a region within condition were removed from the analyses, excluding 2.1% of the data. Figure 1 and Table 2 present the mean reading times per region across the four conditions.

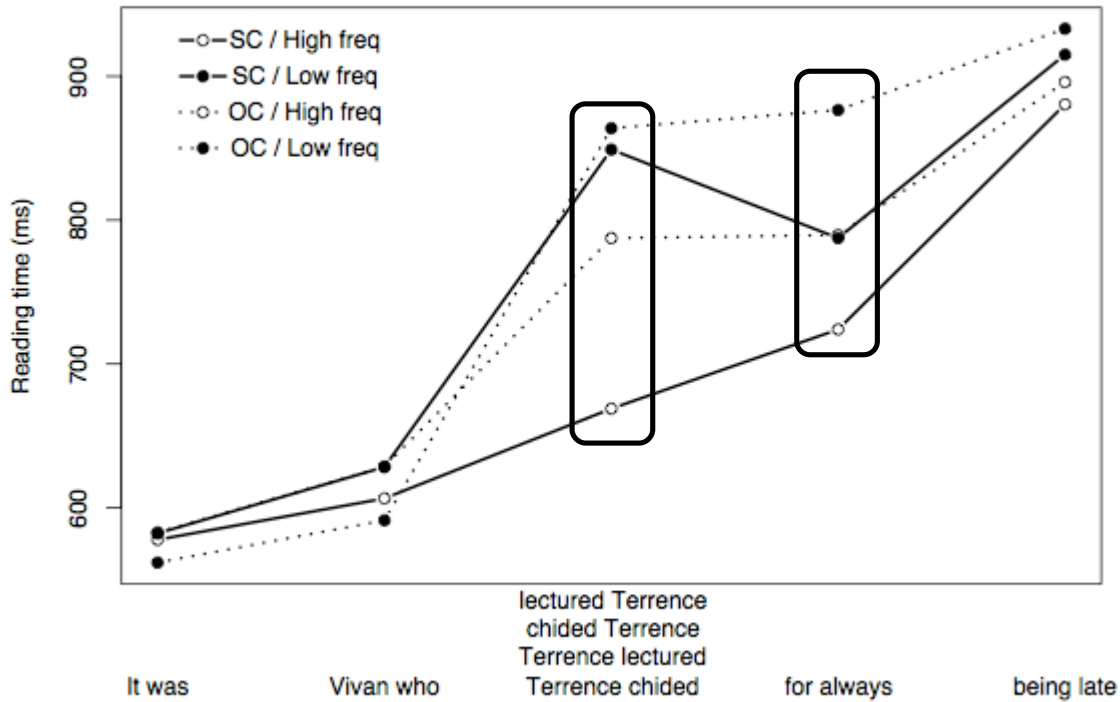


Figure 1: Per-region reading times by condition. The two critical regions are circled.

	<i>It was</i>	name <i>who</i>	cleft	post-cleft	rest
SC / High freq	577.83 (8.2)	606.56 (12.6)	668.94 (17.5)	724.03 (17.7)	880.43 (43.0)
SC / Low freq	582.23 (8.6)	628.39 (12.5)	849.02 (26.5)	787.49 (19.0)	915.00 (40.7)
OC / High freq	582.90 (8.6)	628.71 (13.0)	787.46 (26.4)	789.62 (24.2)	895.92 (41.6)
OC / Low freq	561.93 (7.1)	591.33 (11.5)	863.86 (34.0)	876.50 (27.4)	933.05 (50.9)

Table 2: Per-region reading times by condition. Standard errors in parentheses.

We fitted separate multilevel linear regressions of raw reading time with crossed random intercepts for participants and items to each sentence region. At the first critical region (the cleft), there was a significant main effect of frequency, which interacted with syntactic structure. The high-frequency conditions were read faster than the low-

frequency conditions, and the high-frequency SC condition was read faster than the high-frequency OC condition. That is, whereas we observe the standard extraction effect for the high-frequency conditions, we do not for the low-frequency conditions. These effects are displayed in Table 3 and Figure 2. The presence of all predictors in this and other models presented here was justified by log likelihood model comparison.

	β	Std. Error	t	p
Intercept (low-freq/OC)	872.08	49.73	17.537	<2e-16
structure=SC	-17.12	27.54	-0.622	0.5343
verb frequency=high	-77.86	27.77	-2.803	0.0051
structure*frequency	-94.79	39.12	-2.423	0.0156

Table 3: Final linear regression model for the reading times in the cleft region using a binary frequency predictor.

At the second critical region (post-cleft), there were significant main effects of both frequency and structure, such that the SC conditions were read faster than the OC conditions, and the high-frequency conditions were read faster than the low-frequency conditions. These effects are displayed in Table 4 and Figure 2. Although the interaction between the two variables was not significant ($\beta=45.30$, standard error=39.40, $p=.25$), the numerical trend was in the opposite direction from the interaction observed at the cleft region, such that the difference between the SC and OC conditions was greater for low-frequency than for high-frequency conditions.

	β	Std. Error	t	p
Intercept (low-freq/OC)	874.13	32.31	27.053	<2e-16
structure=SC	-78.31	19.62	-3.992	0.0001
verb frequency=high	-73.36	20.01	-3.666	0.0003

Table 4: Final linear regression model for the reading times in the post-cleft region using a binary frequency predictor.

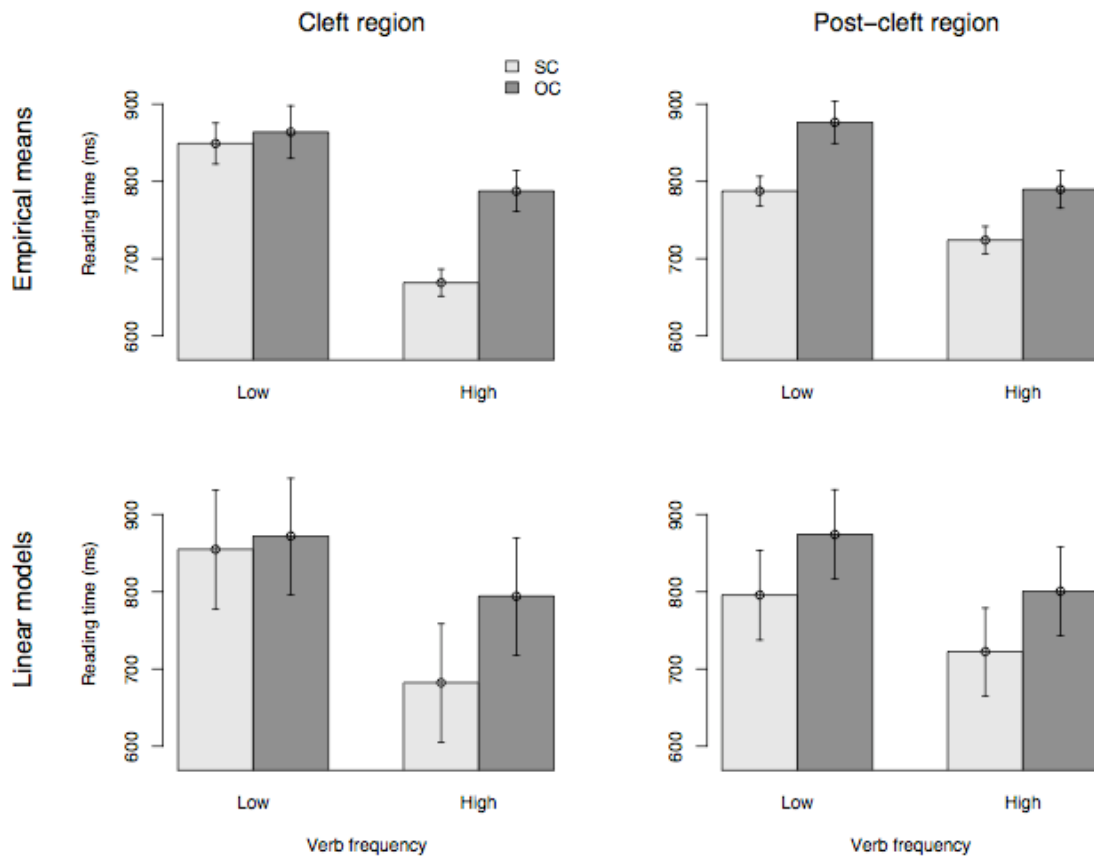


Figure 2: Empirical per-condition means (upper plots, error bars represent two standard errors of the mean) and fitted model predictions for the two linear regression models

(lower plots, error bars represent two standard deviations).

No effects were predicted prior to the first critical region because the materials are identical up to that point. Consistent with this prediction, at the first two-word region (“It was”), a multilevel linear regression revealed no main effects or interactions (the absolute value of all t-statistics <1.44 , $p_s > 0.10$). In the next two-word region (“name who”), however, there were unexpected effects of structure such that the OC conditions were read faster than the SC conditions ($\beta = 14.576$, standard error = 5.232, $p = 0.005$), frequency such that the high-frequency conditions were read slower than the low-frequency conditions ($\beta = 15.575$, standard error = 5.274, $p = 0.003$), and a subadditive interaction such that the SC/high-frequency condition was read somewhat faster ($\beta = -21.748$, standard error = 7.408, $p = 0.003$). Because the materials did not differ between conditions at this region, and the effect sizes are very small (i.e. the maximum difference between conditions is ~ 15 ms; cf. the differences of ~ 70 -180 ms for the crucial contrasts at the critical regions), we suspect the significance to be due to chance external factors. At the region after the second critical region, there were no main effects or interactions (the absolute values of all t-statistics <1.7 , $p_s > 0.08$).

We now present an analysis of the same data with frequency recoded as a continuous covariate using counts from the Web-1T corpus (Branz & Franz, 2006). This second analysis detects effects not apparent in the first analysis. The pairwise contrasts between our high- and low-frequency verb pairs with similar meanings are maintained due to the inclusion of per-item random intercepts in the models reported below, but the interpretation of the frequency variable now indicates differences in reading time due to

the *relative* frequency differences between the individual verbs. Following Gelman (2008), we standardize the continuous log frequency variable by subtracting the mean and dividing by two standard deviations. This allows an approximate comparison of the relative effect size of the binary and continuous predictors by looking at the relative sizes of the coefficients, and makes interpretation of its effect easier in the presence of an interaction term.

At the first critical region (the cleft), a multilevel linear regression with random intercepts for participants and items indicated significant main effects of both syntactic structure and standardized log frequency, such that higher frequency verbs and subject clefts both led to faster reading times. Additionally, the two predictors interacted such that higher frequency subject clefts were read faster still. These effects are shown in Table 5.

	β	Std. Error	t	p
Intercept (mean freq/OC)	833.09	47.96	17.371	<2e-16
structure=SC	-63.89	19.50	-3.276	0.0011
log verb frequency	-83.88	28.50	-2.944	0.0033
structure*frequency	-79.44	39.25	-2.024	0.0432

Table 5: Final linear regression model for the reading times in the cleft region using a continuous frequency predictor.

At the second critical region (post-cleft), a multilevel linear regression with

random intercepts for participants and items again indicated significant main effects of both syntactic structure and standardized log frequency, such that higher frequency verbs and subject clefts both led to faster reading times. Additionally, the two predictors interacted significantly, but this time in the opposite direction, such that the high-frequency SC condition was read more slowly. This interaction approximately counteracts the effect of structure for the higher frequency verbs, resulting in an overall effect of structure only in the lower frequency verbs. These effects are shown in Table 6.

	β	Std. Error	t	p
Intercept (mean freq/OC)	837.45	30.63	27.338	<2e-16
structure=SC	-78.26	19.57	-3.999	0.0001
log verb frequency	-126.66	28.54	-4.438	<2e-16
structure*frequency	84.47	39.20	2.155	0.0314

Table 6: Final linear regression model for the reading times in the post-cleft region using a continuous frequency predictor.

To summarize, we find expected effects of syntactic structure (subject clefts lead to faster reading times than object clefts) and verb frequency (higher frequency verbs lead to faster reading times than lower frequency verbs). However, these predictors interact: at the cleft region, the SC/OC asymmetry leads to a greater difference in reading times for higher frequency verbs. In contrast, at the post-cleft region, there is no evidence for such an interaction using a dichotomous frequency predictor, and with the

more sensitive continuous measure we in fact find the reverse pattern: structure exerts a greater effect on reading time for lower frequency verbs.

These results are consistent with a parsing model in which lexical access of the verb must precede the structural process that incurs difficulty in the object-extracted case (i.e., integrating distant dependents, positing a complex syntactic structure, or reanalyzing a subject-extracted clause as an object-extracted one). When the verb is high frequency, lexical retrieval is fast, and the structural operation is performed during the verb region; when the verb is low frequency, retrieval is slow, and the structural operation is delayed until the post-cleft region.

If lexical retrieval time is a continuous function of frequency, we should expect a corresponding continuous variation in the point at which the structural effect is observed. That is, we should see less of the structural effect observed on the spillover region and more on the cleft region as verb frequency increases. To investigate this possibility in more detail, we analyzed the data from the two critical regions in a single model, including region as a main effect and in interaction with the other predictors. The coefficients of the higher-order interaction term between structure, frequency, and region can then be interpreted as the difference in the location of the structural effect as a function of the frequency of the verb. As before, we include crossed random intercepts to account for variation due to participants and items.

	β	Std. Error	t	p
Intercept (mean freq/OC/cleft)	831.461	37.274	22.307	<2e-16
structure=SC	-66.730	21.050	-3.170	0.0015

log verb frequency	-87.939	30.604	-2.873	0.0041
region=post-cleft	8.398	21.029	0.399	0.6897
structure*frequency	-77.014	42.301	-1.821	0.0688
frequency*region	-34.613	42.258	-0.819	0.4128
structure*region	-8.255	29.753	-0.277	0.7815
structure*freq*region	155.944	59.536	2.619	0.0089

Table 7: Final linear regression model for reading time in the two critical regions, with region (cleft vs. post-cleft) added as a predictor in the model.

Since the second-order interaction term is significant, the hierarchy principle dictates that we should not remove either the non-significant first order interactions or the non-significant main effect (region). Rather, the entire group of predictors should be taken as significant as a group: the non-significance of these terms does not mean that they do not codetermine reading time, only that their importance is in conjunction with the significant interaction terms.

The main effect of frequency is significant in the expected direction: it shows that higher frequency verbs tend to lead to faster reading times, aggregating across both regions and both structures. Likewise, subject clefts are read faster than object clefts, as expected. The interaction terms cannot be interpreted individually, but must be taken as a group. For that reason, it is better to visualize the per-condition predictions. This is done in Figure 3, which shows the predictions that the model makes for reading time as a function of verb frequency for each of the four combinations of the two regions and two

syntactic structures. The top two plots show a LOESS curve fitted from the data: this gives a description of the empirical trends in the data. In particular, it can be seen that there is a greater difference between the SC and OC reading times for higher frequency verbs at the cleft region, and for lower frequency verbs at the post-cleft region. The bottom two panels show the predictions of the linear regression model described above. Visually comparing the top and bottom panels confirms that these linear estimates capture the major trends in the data well. The chief departure of the model predictions from the empirical data seems to be for the very high frequency verbs in the SC condition at the cleft region. This is also the sparsest region of the frequency spectrum in our stimuli, and although it is certainly possible that there may be a systematic nonlinear relationship between frequency and reading time for this condition, the data pattern is also compatible with the simpler hypothesis represented by our model, clouded slightly by statistical noise in results from one of the three highest frequency verbs tested.

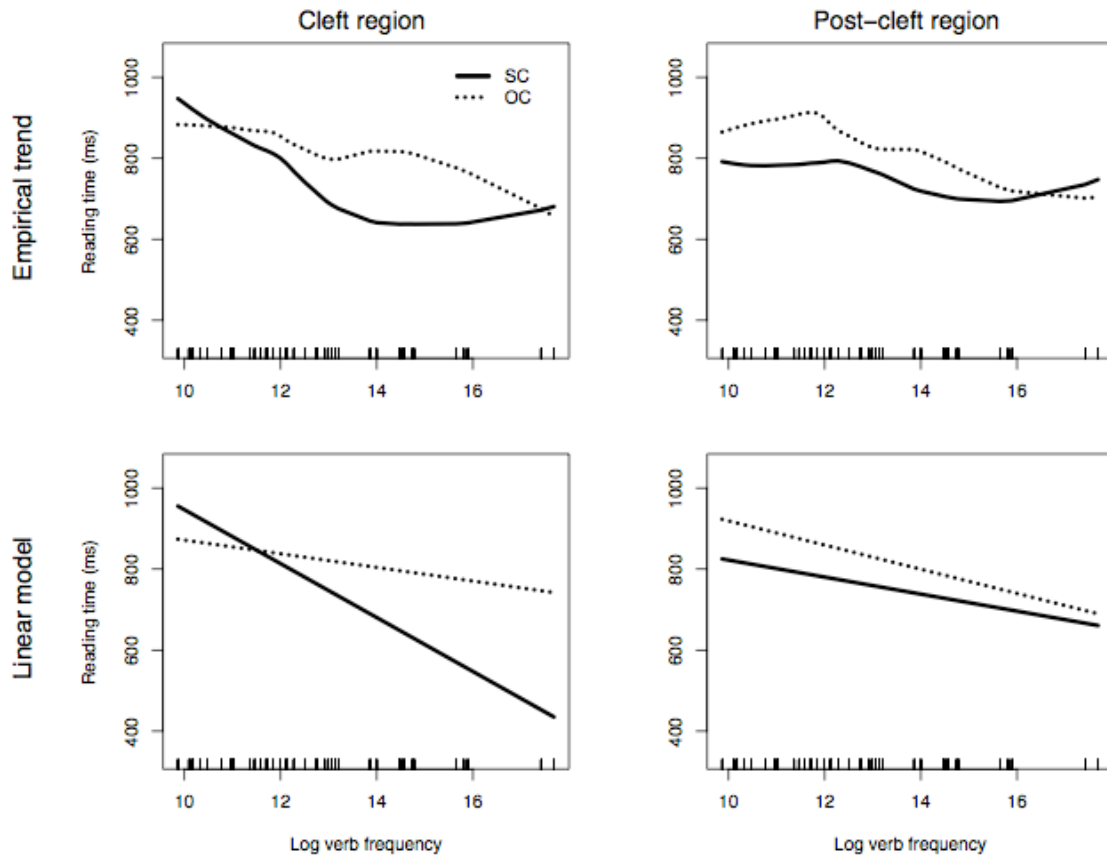


Figure 3: The model predictions (two lower panels) and the data (two top panels) for the two critical regions: cleft (the left panels) and post-cleft (the right panels).

To summarize these results, we presented a hierarchical linear regression model that predicts reading time difficulty in the cleft region and the post-cleft region as a function of the structure (SC vs. OC) and the frequency of the verb. The model identifies significant interactions between those variables which combine to make the following generalizations: whereas for high-frequency verbs, the difficulty of reading an OC relative to an SC is largely experienced in the cleft itself, for low-frequency verbs it is largely experienced in the post-cleft region. Moreover, including data from both regions

reveals that this is a gradient effect: as the frequency of the verb decreases, more of the difficulty is gradually shifted from the cleft region to the post-cleft region.

We also draw attention to a well-studied but often ignored methodological point. Coding an underlyingly continuous variable such as frequency as a binary factor may dramatically reduce the power of subsequent statistical tests. In our case, switching from a binned factor to the log frequency covariate revealed a strongly significant interaction that had not been apparent before. Dichotomization of a continuous measure is in general not advisable if there is a smooth function which could characterize its relationship with the dependent variable instead: for lexical frequency, which has well-studied logarithmic and linear effects on many behavioral measures, this is certainly the case. Binning the frequency estimates leads to a loss of information about frequency differences between verbs within each group, and hence a loss of statistical power (see e.g., Cohen, 1983).

Discussion

The experiment described here was designed to investigate the temporal ordering of structural and lexical processes in online language comprehension. Although it is logically possible that the retrieval of a verb's lexical entry could occur before, during, or after the creation of the appropriate structure, our results argue for the first of these possibilities. In particular, we find that differences in reading times which correspond to the different syntactic structures tested are experienced later for low-frequency verbs than for high-frequency verbs. We interpret this as a full or partial temporal staging of these processes: in the type of sentence we investigated, structural processes are only invoked once the verb's lexical entry has been retrieved. Factors such as frequency which influence the time taken in retrieval therefore cause any structural processes to take place correspondingly later.

We should note that although the "structural" difficulty we observed is plausibly syntactic, as it corresponds to the SC/OC manipulation, it could conceivably reflect the *interpretation* of the verb-argument dependency structure rather than its *construction*. For this reason, we refer to the SC/OC asymmetry as "structural" rather than simply "syntactic". If there are separate processing stages to (a) posit an abstract syntactic structure, and (b) interpret it, it is possible that the first of these processes happens concurrently with lexical retrieval, and only the second is delayed by a low frequency verb. That said, it is informative that we saw no evidence for any difference between the structural conditions at the cleft region for the lowest end of the frequency spectrum (this can be seen most clearly in the plots in Figures 1, 2 and 3) meaning that any hypothetical process occurring concurrently with lexical retrieval would have to incur no observable

processing cost. Since the cleft region is the point at which the SC/OC effect is observed in the sizable literature discussed in the Introduction, it seems likely that the effect that we find can be delayed by lexical access is this same previously reported effect, and not some separate process. If this is the case, we expect this "delayed" structural effect to be modulated by other factors that have been shown to influence the size of the extraction effect, such as noun animacy (e.g., Mak, Vonk, & Schriefers 2002; Traxler et al., 2002) or interference from other elements in memory (e.g., Gordon et al., 2001; Gordon et al., 2002; Fedorenko, Gibson & Rohde, 2006).

As discussed in the Introduction, the strict ordering of the two processes is compatible with most constraint-based competition theories, but not explicitly predicted by them. Expositions of the constraint-based approach have chiefly focused on showing the importance of lexical information in guiding syntactic decisions, which runs contrary to the predictions of syntax-first/reanalysis models. Constraint-based theories do not typically discuss whether the reverse flow of information, from syntactic to lexical processing is also possible, but if a guiding principle of such theories is that comprehenders make approximately optimal use of all available information (in the sense of McClelland, 1987) then one would expect this to be the case. For instance, under expectation-based comprehension theories, syntactic structures may be tentatively posited well in advance, allowing accelerated processing of later words associated with those structures. However, the results presented here argue the opposite: some structural processes are constrained to occur after lexical retrieval of the words they incorporate. This suggests that comprehenders might not be able to make optimal use of some structural properties during lexical processing.

In related work, Tabor et al. (1997) and Gibson (2006) showed that lexical retrieval can be slowed down by the presence of homophonous lexical items that could have been ruled out by the syntactic context. This suggests some independence of lexical and structural processes: lexical retrieval does not appear to make full use of syntactic information, a counter-intuitive finding if one expects the comprehender to make optimal use of all available information. Whereas our experiment does not test this hypothesis, our results yield a very compatible conclusion. In fact, they suggest a possible explanation for the previous results: it may be that lexical retrieval in Tabor et al.'s and Gibson's studies fails to make use of available syntactic information because the appropriate structural decisions have not yet been made. We leave the exploration of this possibility to future research.

The work of Boland and colleagues (Boland, 1997; Boland et al. 2001) indicates that lexical information is immediately available to guide structural choices, while discourse information only becomes available at a later stage. Our results are compatible with this account, as they suggest that lexical retrieval indeed precedes structural processing associated with the difference between object- and subject-extracted cleft structures. While our results do not impact Boland's theory according to which lexical and discourse information affect multiple specific processing stages in comprehension, they pave the way for a fuller and more detailed model that elucidates the temporal relationship between stages of online processing.

Garden-path models assume that lexical information is only useful in determining syntactic structure if the currently assumed structure turns out to be incorrect and lexically specific knowledge can be used to inform the reanalysis process. In the case of

the cleft sentences we studied, models based on the garden-path theory hold that the subject-extracted structure is assumed initially, and the increased difficulty in the object-extracted case is due to reanalysis. However, previous experimental work has uniformly found that in object-extracted relative clauses and cleft sentences generally, difficulty is experienced at the embedded verb and not before, even though the subject-extracted structure can be ruled out at the beginning of the embedded subject NP (e.g., at "the lawyer" in (3b)). One possible explanation of this fact is that difficulty surfaces later than expected by the theory, the "spill-over" effect often observed in reading time studies. However, Grodner & Gibson (2005) manipulated the amount of material preceding the embedded verb in the object-extracted structure by adding modifiers to the embedded subject (e.g., including a prepositional phrase after the embedded subject, as in "The administrator who the nurse from the clinic supervised ...") and found that regardless of the length of the "spill-over" region after the embedded subject, the chief locus of difficulty was always the embedded verb. Turning to the results presented here, the fact that the structural process appears to be ordered after lexical retrieval of the verb is incompatible with a reanalysis-based hypothesis. If reanalysis occurs at the embedded subject noun phrase, there is no reason to expect reanalysis difficulty to be postponed until after the verb specifically, and that is the point at which difficulty is experienced in this experiment. An alternative account in which difficulty was due to reanalysis, but reanalysis could not proceed until the verb had been fully processed would account for these data well. Given that lexical information is assumed to aid reanalysis, waiting for the lexical information given by the verb before attempting reanalysis might be a reasonable strategy. Nevertheless, we know of no specific proposal along these lines.

Given recent advances made in the theory of expectation-driven processing (e.g., Hale, 2001; Levy, 2008), it is worth discussing the implications of these results for that family of theories, as well. Superficially, our findings seem incompatible with expectation-driven accounts: classic results such as those of Konieczny (2000) show that the presence of preverbal dependents may decrease reading time at the verb, as they increase readers' level of expectation for an upcoming verb. Under the "surprisal" account spelled out in Hale (2001), these results are expected because comprehenders posit upcoming syntactic structure probabilistically in proportion to their probability given words seen so far: more predictable structures lead to faster reading times because the effort needed to construct them was expended earlier in the sentence. Of course, this could not be possible if syntactic processes are delayed until after lexical retrieval: any syntactic effects could only be observed at or after the words involved. This strikes a chord with another finding which is not well accounted for by expectation-driven theories: the object-extracted structure can be predicted before the verb, and yet difficulty is not experienced until the verb (see Grodner & Gibson, 2005). That locus of difficulty is to be expected if the relevant syntactic process cannot begin until the verb's lexical entry is retrieved, in which case an expectation-based approach cannot account for any differential difficulty between the two structures. Of course, this cannot be the whole story: expectation provides a natural account of findings for a range of sentence types in multiple languages (see Levy, 2008 for a survey). A possible refinement might build on the idea that multiple syntactic processes are invoked during online comprehension, and that the specific process which leads to the well-studied difficulty in comprehending object-extracted relative clauses and clefts is (a) not significantly boosted by high

predictability of the verb, and (b) not invoked until after the verb's lexical entry has been retrieved. We know of no specific proposals along these lines to date.

In summary, the results presented here suggest that certain structural processes – specifically, the processes responsible for difficulty in object-extracted clefts or relative clauses – are temporally constrained to begin only after lexical retrieval of the verb is complete. This novel finding is not directly predicted by any existing account of online sentence comprehension, though with certain modifications, any of the accounts that are dominant in the sentence processing literature could be adapted to explain these results. However, theories in which difficulty is anticipated early on in the relative clause (expectation-based and reanalysis-based theories) are less well equipped to incorporate these findings. By looking at the time-course over which online effects are experienced, we were able to better isolate two component processes involved in online sentence comprehension, and better understand the flow of information between them. Investigating the time-course of stages of processing is a potentially rich source of data for evaluating models of language comprehension, and should inform the design of future models accordingly.

References

- Bates, D., Maechler, M. and Dai, B. (2008). lme4: Linear mixed-effects models using Eigen and Eigenfaces. R package version 0.999375-27. <http://lme4.r-forge.r-project.org/>
- Bever, T., (1970). The cognitive basis for linguistic structures. In J.R. Hayes (ed.). *Cognition and the development of Language*. Wiley, New York, 279-362.
- Boland, J.E. & Blodgett, A. (2001) Understanding the constraints on syntactic generation: Lexical bias and discourse congruency effects on eye movements. *Journal of Memory and Language*, 45, 391-411.
- Boland, J.E. (1997). Resolving syntactic category ambiguities in discourse context: Probabilistic and discourse constraints. *Journal of Memory and Language*, 36, 588-615.
- Brants, T. & Franz, A. (2006) Web 1T 5-gram Version 1. Philadelphia: Linguistic Data Consortium.
- Clifton, C., & Frazier, L. (1989). Comprehending sentences with long-distance dependencies. In G.N. Carlson & M.K. Tanenhaus (eds.), *Linguistic structure in language processing* (pp. 273-317). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Clifton, C., Staub, A. & Rayner, K. (2007). Eye-movements in reading words and sentences. In R.P.G. van Gompel, M.H. Fischer, W.S. Murray and R.L. Hill (eds.) *Eye Movements: A Window on Mind and Brain*. Oxford: Elsevier. 341-471.
- Clifton, C.J., Traxler, M., Mohamed, M.T., Williams, R.S., Morris, R.K., & Rayner, K. (2003). The use of thematic role information in parsing: Syntactic processing autonomy revisited. *Journal of Memory and Language*, 49, 317-334.
- Cohen, J. (1983) The cost of dichotomization. *Applied Psychological Measurement*, 7, 249-253.
- Crain, S. & Steedman, M. (1985). On not being led up the garden path: the use of context by the psychological parser. In: Dowty D., Karttunen L. & Zwicky A. (eds.). *Natural Language Processing: Psychological, Computational and Theoretical Perspectives*, 320-358. Cambridge, UK: Cambridge University Press.
- Fedorenko, E., Gibson, E., & Rohde, D. (2006). The nature of working memory capacity in sentence comprehension: Evidence against domain-specific working memory resources. *Journal of Memory and Language*, 54, 541-553.
- Fedorenko, E., Piantadosi, S., Frank, M., & Gibson, E. (submitted). The interaction of syntactic and lexical information sources in language processing: The case of the noun-verb ambiguity.
- Ferreira, F. & Clifton, C. (1986) The independence of syntactic processing. *Journal of Memory and Language*, 25(3) 348-368.
- Ferreira, F., & Henderson, J.M. (1990). Use of verb information in syntactic parsing: Evidence from eye movements and word-by-word self-paced reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 555-568.
- 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.
- Ford, M. (1983) A method of obtaining measures of parsing complexity throughout sentences. *Journal of Learning and Verbal Behavior*, 22, 203-218.
- Frazier, L. (1979). *On comprehending sentences: Syntactic parsing strategies*. Ph.D.

- dissertation, Indiana University Linguistics Club, University of Connecticut.
- Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, 14, 178-210.
- Frazier, L. (1987). Sentence processing: A tutorial review. In: M. Coltheart (ed.), *Attention and performance XII: The psychology of reading*, Hillsdale, NJ: Erlbaum, 559-586.
- Garnsey, S.M., Pearlmutter, N.J., Myers, E. & Lotocky, B. (1997). The relative contributions of verb bias and plausibility to the interpretation of temporarily ambiguous sentences. *Journal of Memory and Language*, 37, 58-93.
- Gelman, A. (2008) Scaling regression inputs by dividing by two standard deviations. *Statistics in Medicine*, 27, 2865-2873.
- Gennari, S. & MacDonald, M. (2008). Semantic indeterminacy in object relative clauses. *Journal of Memory and Language*, 58, 161-187.
- Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68, 1-76.
- 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.
- Gordon, P.C., Hendrick, R., & Johnson, M. (2001). Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 27(6), 1411-1423.
- Gordon, P.C., Hendrick, R., & Levine, W.H. (2002). Memory-load interference in syntactic processing. *Psychological Science*, 13, 425-430.
- Grodner, D. & Gibson, E. (2005). Consequences of the serial nature of linguistic input. *Cognitive Science*, 29, 261-290.
- Hale, J. (2001) A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of NAACL*, 2, 159-166.
- Just, M.A., Carpenter, P.A., & Woolley, J.D. (1982). Paradigms and processing in reading comprehension. *Journal of Experimental Psychology: General*, 111, 228-238.
- King, J. & Just, M.A. (1991). Individual differences in syntactic processing: the role of working memory. *Journal of Memory and Language*, 30, 580-602.
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, 95(2), 163-182.
- Konieczny, L. (2000). Locality and parsing complexity. *Journal of Psycholinguistic Research*, 29(6), 627-645.
- Levy, R. (2008) Expectation-based syntactic comprehension. *Cognition*, 106(3), 1126-1177.
- Luce, P., Pisoni, D., & Goldinger, S. (1990). Similarity neighbourhoods of spoken words. In G. Altmann (ed.), *Cognitive Models of Speech Processing*. Cambridge, MA: MIT Press, 122-147.
- MacDonald, M., Pearlmutter, N. & Seidenberg, M. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review*, 103(4), 676-703.
- MacDonald, M., & Christiansen, M. (2002). Reassessing Working Memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review*, 109(1), 35-54.
- MacWhinney, B., & Pleh. (1988). The processing of restrictive relative clauses in

- Hungarian. *Cognition*, 29, 95-141.
- Mak, W., Vonk, W., & Schriefers, H. (2002). The influence of animacy on relative clause processing. *Journal of Memory and Language*, 47, 50-68.
- Marslen-Wilson, W. (1990). Activation, competition, and frequency in lexical access. In G. Altmann (ed.), *Cognitive Models of speech processing*, Cambridge, MA: MIT Press, 148-172.
- Marslen-Wilson, W.D., & Tyler, L.K. (1987). Against modularity. In J. Garfield (ed.), *Modularity in knowledge representations and natural language understanding*. Cambridge, MA: MIT Press, 37-62.
- McClelland, J.L. (1987). The case for interactionism in language processing. In M. Coltheart (ed.), *Attention & performance XII: The psychology of reading*. London: Erlbaum. 1-36.
- 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, 283-312.
- Meyer, A.A., & Schvaneveldt, S. (1971). Facilitation in recognizing pair of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, 90, 227-234.
- Mitchell, D.C. (1987). Lexical guidance in human parsing: Locus and processing characteristics. In M. Coltheart (Ed.), *Attention and performance XII*. Hillsdale, N J: Erlbaum, 601-681.
- R Core Development Team, (2008). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. <http://www.R-project.org>
- Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning and Verbal Behavior*, 22, 358-374.
- Sheldon, A. (1974). On the role of parallel function in the acquisition of relative clauses in English. *Journal of Verbal Learning and Verbal Behavior*, 13, 272-281.
- Spivey, M.J & Tanenhaus, M.K., 1998. Syntactic ambiguity resolution in discourse: Modeling the effects of referential context and lexical frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24(6), 1521-1543.
- Tabor, W., Juliano, C. & Tanenhaus, M.K. (1997). Parsing in a dynamical system: An attractor-based account of the interaction of lexical and structural constraints in sentence processing. *Language and Cognitive Processes*, 12, 211-272.
- Tabossi, P., Colombo, L., & Job, R., (1987). Accessing lexical ambiguity: Effects of context and dominance. *Psychological Research*, 49, 161-172.
- 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
- Traxler, M., Morris, R., & Seely, R. (2002). Processing subject and object relative clauses: Evidence from eye movements. *Journal of Memory and Language*, 47, 69-90.
- Traxler, M., Williams, R., Blozis, S., & Morris, R. (2005). Working memory, animacy, and verb class in the processing of relative clauses. *Journal of Memory and Language*, 53, 204-224.
- Trueswell, J., Tanenhaus, M., & Kello, C. (1993). Verb specific constraints in sentence processing: Separating effects of lexical preference from garden paths. *Journal of*

- Experimental Psychology: Learning, Memory and Cognition, 19, 528-553.
- Trueswell, J.C. (1996). The role of lexical frequency in syntactic ambiguity resolution. *Journal of Memory and Language*, 35, 566-585.
- Trueswell, J.C., Tanenhaus, M.K. & Garnsey, S.M. (1994). Semantic influences on parsing: use of thematic role information in syntactic disambiguation. *Journal of Memory and Language*, 33, 285-318.
- van Gompel, R.P.G., Pickering, M.J., & Traxler, M.J. (2000). Unrestricted race: A new model of syntactic ambiguity resolution. In A. Kennedy, R. Radach, D. Heller, & J. Pynte (eds.), *Reading as a perceptual process*. Oxford: Elsevier, 621-648.
- van Gompel, R.P.G., Pickering, M.J., & Traxler, M.J. (2001). Reanalysis in sentence processing: Evidence against current constraint-based and two-stage models. *Journal of Memory and Language*, 45(2), 225-258.
- Wanner, E. & Maratsos, M. (1978). An ATN approach to comprehension. In Halle, M., Miller, G. & Bresnan, J. (eds.), *Linguistic theory and psychological reality*. Cambridge, MA: MIT Press.

Appendix: Materials

The subject-extracted versions with the two verb versions are shown below for each of the 24 items. The object-extracted versions can be generated as exemplified in (1) below.

1. It was Brandon who congratulated Dustin when the commencement ended.
It was Brandon who Dustin congratulated when the commencement ended.
It was Brandon who felicitated Dustin when the commencement ended.
It was Brandon who Dustin felicitated when the commencement ended.
2. It was Armando who criticized / belittled Jacob during the presentation.
3. It was Jared who impressed / awed Elvin with a wonderful performance.
4. It was Cedric who insulted / affronted Julius when the game was over.
5. It was Zachary who bothered / badgered Carlton in the library reading room.
6. It was Hector who praised / lauded Dillon for getting an A in the difficult math exam.
7. It was Marcela who calmed / placated Dora before the show started.
8. It was Pauline who aggravated / peeved Natalie because of a previous dispute.
9. It was Tara who hated / disdained Stefanie since early childhood.
10. It was Mollie who humiliated / abased Lauren in front of everyone in the audience.
11. It was Kylie who mesmerized / allured Selena during the prom dance.
12. It was Janelle who paid / remunerated Gladys after an argument.
13. It was Daniel who charmed / beguiled Kendra from the first date.
14. It was Jeremy who disciplined / chastised Francesca because of unacceptable behavior.
15. It was Brendan who pleased / assuaged Amanda with a radiant gift.
16. It was Andrew who punished / castigated Shawna following an inadmissible mistake.
17. It was Timothy who signaled / beckoned Samantha the directions to the park.
18. It was Tyler who murdered / annihilated Kathryn that dreadful night.
19. It was Sofia who strangled / asphyxiated Joshua in the first scene of the play.
20. It was Genevieve who teased / pestered Sebastian because of the weird shoes.
21. It was Tania who confused / vexed Byron explaining the procedure.
22. It was Vivian who lectured / chided Terrence for always being late.
23. It was Isabel who informed / apprised Garrett of the importance of the exam.
24. It was Chloe who rejected / spurned Edgar for lack of confidence.

Acknowledgements:

We are grateful to Daniel Jaramillo for his help in constructing the experimental materials, and to Frank Keller and the audience at the 21st CUNY Conference at UNC (Chapel Hill) for helpful comments on this work. We also thank the editor and two anonymous reviewers for their insightful comments, which helped improve this manuscript.