The processing complexity of English relative clauses

Edward Gibson, Hal Tily, Evelina Fedorenko

Introduction

Sentences that contain non-local dependencies between words have long been known to cause comprehension difficulty, relative to sentences where all the dependencies are local. Relative clause (RC) structures have been used extensively to evaluate hypotheses about the source of these complexity effects because different word order configurations can be compared while keeping the words the same and the meaning similar across conditions. The most frequently contrasted RC structures are subject-extracted and object-extracted RCs (SRCs and ORCs), the two most common types of RCs cross-linguistically (Keenan & Comrie, 1977). For example, consider the English RCs modifying the subject noun phrase (NP) "the reporter" in (1):

(1) a. Subject-extracted relative clause (SRC)
The reporter who __ attacked the senator admitted the error.
b. Object-extracted relative clause (ORC)
The reporter who the senator attacked __ admitted the error.

The wh-element "who" is the subject and agent of the embedded verb "attacked" in the subject-extracted condition (1a), and it is the object and patient of "attacked" in the object-extracted condition (1b). In English, the dependency is local between the subject NP and the embedded verb in an SRC (e.g., the pronoun "who" is adjacent to "attacked" in (1a)). In contrast, the dependency is non-local between the object NP and the embedded verb in an ORC: another NP intervenes (e.g., "the senator", occurs between "who" and "attacked" in (1b)).

In addition to subject-modifying RCs, as in (1), researchers sometimes compare SRCs and ORCs that modify other syntactic positions. In English, the extraction effect appears to be the same regardless of the syntactic role played by the NP that the RC is modifying (Gibson et al., 2005).

Across languages, for RCs where both the subject and the object are animate, ORCs have generally been observed to be more difficult (English: Wanner & Maratsos, 1978; King & Just, 1991; Gibson, 1998; Grodner & Gibson, 2005; French: Holmes & O'Regan, 1981; Baudiffier, Caplan, Gaonac'h & Chesnet, 2011; German: Mecklinger, Schriefers, Steinhauer, & Friederici, 1995; Schriefers, Friederici, & Kuhn, 1995; Dutch: Frazier, 1987a; Mak, Vonk & Schriefers, 2002; 2006; Japanese: Miyamoto & Nakamura 2003; Ishizuka, Nakatani & Gibson, 2003; Ueno & Garnsey, 2008; Korean: O'Grady, Lee & Choo, 2003; Kwon, Polinsky & Kluender, 2006; Kwon et al., 2010; cf. *Basque*: Carreiras et al. 2010; *Chinese*: Hsiao & Gibson, 2003; Gibson & Wu, 2011, but see Lin & Bever 2006). For example, in English, ORCs with animate subjects and objects are more difficult to process than corresponding SRCs according to a number of dependent measures, including on-line lexical decision for a word presented during the relative clause, reading times, and response accuracy to probe questions (King & Just, 1991;

among others; see Gibson, 1998, for a review). In reading time experiments in English the difficulty in ORCs is typically observed at the embedded verb (e.g., "attacked" in (1b)) often persisting in the following word(s).

Most proposals that have been advanced to account for the extraction complexity effects fall into three general categories: (1) reanalysis-based theories; (2) experience-/surprisal-based theories; and (3) working-memory-based theories. Memory-based theories further include two sub-classes, each of which has several variants. We summarize these proposals below. (See also O'Grady, 2011, for a good summary of recent work in the processing and acquisition of relative clauses.)

1. Reanalysis-based theories.

The development of detailed models of human parsing by Bever (1970), Kimball (1973) and others led to reanalysis-based theories of reading comprehension difficulty (e.g., Frazier, 1978, 1987b; Frazier & Fodor, 1978; Frazier & Rayner, 1982; Clifton & Frazier, 1989) in which difficulty arises when the parser uses a certain strategy in online comprehension that yields an incorrect analysis, and that analysis has to be repaired. For instance, Frazier & colleagues (Frazier & Flores d'Arcais, 1987; Clifton & Frazier, 1989) suggest that upon encountering a relativizer ("that"), comprehenders always adopt an SRC interpretation, treating the head noun as the subject of both the main and the relative clause. Upon encountering the RC subject ("the reporter" in (1b)), reanalysis is required to interpret the head noun as the object of the RC instead, leading to additional processing cost (see also Traxler et al., 2002, 2005, for more recent discussions of this proposal).

2. Experience-/surprisal-based theories.

Experience-based accounts postulate that comprehenders' difficulty in processing incoming linguistic input is a function of their experience with similar input in the past: the more common a word or a construction is, the less difficulty it will cause in comprehension. This general idea finds a lot of support in the lexical processing literature: more frequent words are indeed easier to process (e.g., Morton, 1969; see Monsell, 1991, for a review). However, the extension of these same principles to phrase, clause, and sentence-level units is more controversial.

The earliest account of RC complexity differences that can be considered experience-based was proposed by Keenan & Comrie (1977), who argued that a) syntactic positions differ in how "accessible" they are (with the subject position being most accessible, followed by the direct object position, etc.), and b) extracting elements from less accessible positions is more costly. One explanation for accessibility is in terms of experience: extractions from less accessible positions are less frequent in the input than extractions from more accessible positions (Keenan, 1975; Hawkins, 2004; see Roland, Dick & Elman, 2007, for relevant English statistics) leading to greater difficulty in processing the former.

A more general version of this theory derives from the constraint-based processing

literature, in which the frequency of structures or interpretations are among the probabilistic cues that comprehenders pay attention to (e.g., Trueswell et al., 1994; Garnsey et al., 1997; MacDonald, Pearlmutter & Seidenberg, 1994). Gennari & MacDonald (2008, 2009) propose that difficulty arises when multiple competing continuations are highly activated, a situation they call "indeterminacy". Both the SRC as well as the passive ORC are argued to be highly activated continuations given the first few words of an active ORC, leading to just such indeterminacy.

Finally, a group of theories formalize this idea using quantities from information theory: surprisal, a measure of the amount of information conveyed by a word in context (Hale, 2001; Levy, 2008) and entropy, a measure of the uncertainty over possible continuations (Hale, 2003). For example, according to surprisal-based accounts, comprehenders form detailed expectations about upcoming elements at many levels of linguistic structure (including rich syntactic expectations; Boston et al., 2008, 2011; Demberg and Keller, 2008, 2009; Hale, 2001; Levy, 2008; Roark et al., 2009; Smith and Levy, 2008), and the difficulty of processing an incoming element is a function of its probability given the preceding context, with more probable elements being easier to process (Hale, 2001; Levy, 2008; Smith & Levy, 2008).

3. Memory-based theories.

According to memory-based theories, ORCs cause greater comprehension difficulty because they require more working memory resources. Two kinds of mental operations in sentence processing have been argued to require working memory resources, leading to two sub-classes of memory-based accounts: (a) keeping track of the predictions about upcoming syntactic elements ("storage" costs); and b) retrieving earlier encountered representations from memory when forming dependencies ("retrieval" or "integration" costs).

a. Storage cost accounts. According to storage cost accounts, maintaining incomplete dependencies is costly (e.g., Yngve, 1960; Chomsky & Miller, 1963; Wanner & Maratsos, 1978; Stabler, 1994: Lewis, 1996; Gibson, 1991; 1998; 2000; Chen, Gibson & Wolf, 2005; Nakatani & Gibson, 2010). For instance, after encountering the embedded subject in an ORC structure (e.g., "the senator" in (1b)), the comprehender is holding onto three incomplete dependencies. Specifically, the main clause subject ("the reporter") is dependent on the predicted main verb; additionally, the wh-element "who" and the embedded subject ("the senator") depend on a predicted object and subject position respectively of the embedded verb. In contrast, in an SRC structure (1a), there are a maximum of two incomplete dependencies at any point in the sentence.

b. Retrieval cost accounts. According to retrieval cost accounts, processing an incoming word entails retrieving its non-local syntactic dependents from memory, and the cost of this retrieval operation may vary depending on the type and/or number of other items stored in memory since the time when the to-be-retrieved dependent was encountered. The greater processing difficulty for ORCs compared to SRCs can thus be explained as due to the linear distance between dependents in an ORC being longer than that in an

SRC. Two non-mutually-exclusive explanations of retrieval difficulty have been proposed, with several concrete proposals building on each of these explanations:

i. According to the *decay-based* explanation, retrieving non-local dependents is difficult because their activation decays as additional elements are being processed. Within this decay-based framework, Gibson (1998, 2000) and Warren & Gibson (2002) have proposed that distance between syntactic dependents is measured in terms of the number of new discourse referents (nouns and verbs) that intervene between those dependents. Two other decay-based distance metrics have been proposed: a metric in terms of the *type* of intervening noun phrases (e.g., pronouns vs. names vs. definite descriptions; Warren & Gibson, 2002), and a metric in terms of the number of intervening words (Gibson, 1998; cf. Hawkins, 1994, for a word-based production difficulty metric). Another influential decay-based proposal is due to Lewis and colleagues (Lewis & Vasishth, 2005; Vasishth & Lewis, 2006; Lewis, Vasishth & Van Dyke, 2006). Lewis and colleagues hypothesize that the activation level of the non-local element, which decays over time, may also be increased with subsequent retrieval(s) (cf. Gibson, 1998, for a related idea).

ii. According to the *interference-based* explanation, retrieving non-local dependents is costly because elements that intervene between the two ends of a dependency may interfere with the representation of the first, to-be-retrieved, element (Gordon, Hendrick & Johnson, 2001, 2004; McElree, Foraker & Dyer, 2003; Lewis & Vasishth, 2005; Lewis, Vasishth & Van Dyke, 2006). For example, retrieving the head noun associated with the object position in an ORC will cause difficulty as a function of how well the intervening elements satisfy the constraints imposed by verb (i.e., the retrieval cue). A better fit between an intervening element and the retrieval cue will lead to more interference, hence more difficulty in retrieving the target element (e.g., Van Dyke & McElree, 2006).

4. Evidence for each theory

The main classes of syntactic complexity accounts – reanalysis-, experience- and working-memory-based – do not have to be construed as mutually exclusive: many researchers agree that some aspects of each type of account are probably correct (e.g., Boston et al., 2008, 2011; Demberg and Keller, 2008; Gibson, 2007; Vasishth & Drenhaus, 2011). Next we review some key differences in the predictions that the different accounts make and the available evidence.

a. The locus of the difficulty effect in ORCs.

Predictions: Reanalysis-, experience- and retrieval-cost memory-based accounts make clear and distinct predictions about where difficulty should occur during the processing of ORCs. In particular, both reanalysis- and experience-based theories predict difficulty to arise at the point where the comprehender knows that an ORC structure is being processed, i.e., at the embedded subject ("the senator" in (1b)). In the reanalysis theory, this is when the comprehender realizes that the default parsing strategy has failed, and in

the expectation theory this is when the comprehender realizes that a less frequent / probable structure is being processed. These accounts predict little / no difficulty at the embedded verb because the verb is the expected continuation following the embedded subject (Hale, 2001; though see Hale, 2003). In contrast, retrieval-cost memory-based accounts predict little difficulty at the embedded subject, where no non-local dependents need to be retrieved, predicting difficulty instead at the embedded verb ("attacked" in (1b)), when retrieval occurs.

Evidence: Consistent with the predictions of retrieval-based accounts, in word-by-word lexical decision (Ford, 1983), self-paced reading (e.g., King & Just, 1991; Gordon, Hendrick & Johnson, 2001; Grodner & Gibson, 2005) and eye-tracking (e.g., Holmes & O'Regan, 1981; Gordon et al., 2006; Lowder & Gordon, submitted) paradigms, difficulty has been shown to occur at the embedded verb with little or no difficulty observed at the embedded subject (cf. Staub, 2010, who finds some evidence of difficulty at the embedded subject in early eye-tracking measures). However, in a task where participants had to incrementally select which of two words was a valid continuation of the sentence, ORC difficulty is entirely experienced at RC onset, in line with expectation and reanalysis models (Forster, Guerrera & Elliot, 2009). Thus it appears that significant cognitive effort is exerted at the verb in online reading, although the processing that is typically done at the verb can in principle be completed earlier if people are forced to parse the input more fully as it arrives.

One caveat to consider when evaluating the evidence from online reading tasks is the possibility that difficulty is actually experienced upon processing the embedded subject, but the behavioral correlate that is being measured occurs slightly later, falling by chance at the embedded verb (so-called "spillover" effects, Mitchell, 1984). However, Grodner & Gibson (2005) have ruled out this possibility, by demonstrating that the slowdown occurs at the embedded verb even when the embedded subject is modified with a prepositional phrase (e.g., "the nurse at the clinic").

b. The types of the subject and object NPs.

Predictions: As discussed above, interference-based memory-based accounts predict that the better the intervening elements in a non-local dependency fit the selection criteria of the retrieval cue, the more difficult it should be to retrieve the target element. So, for example, with respect to ORCs, the better the match between the embedded subject and the object selection requirements of the embedded verb, the harder it should be to retrieve the object noun. Working out exactly which kinds of features (syntactic, lexico-semantic, phonological, etc.) are considered in computing the match between the retrieval cue and the potential to-be-retrieved elements is a matter of active research. Different versions of the decay-based memory-based accounts differ with respect to their predictions: a word-based distance metric (e.g., Gibson, 1998) makes no prediction for the effects of NP type, but the new-discourse-referent-based and NP-type-based metrics (Gibson, 1998; Warren & Gibson, 2002) predict that when the intervening elements are pronouns, and therefore discourse-old, retrieval difficulty should be reduced. The predictions of reanalysis- and experience-based accounts depend on the assumptions made about the granularity of the

linguistic representations we tabulate. If we assume that our knowledge is highly abstract (i.e., we only keep track of how frequently particular phrase-structure rules, or sequences of syntactic categories have been encountered), then the types of NPs should not matter. If, on the other hand, we assume that we store details of which types of nouns and verbs co-occur in different syntactic contexts – a more plausible assumption given the available evidence – then all sorts of lexico-semantic properties of the relevant nouns and verbs are predicted to matter, and the more frequent configurations are predicted to cause less processing difficulty. However, with respect to these accounts, it is not clear how to tabulate the relevant frequencies of different linguistic events (see Mitchell, Cuetos, Corley & Brysbaert, 1995). For example, do we keep track of the different frequencies of the animacy configurations (e.g., animate subject / animate object, animate subject / inanimate object, etc.) across all clauses or for main vs. relative clauses separately, or do we keep track of the frequencies of particular verbs in noun-verb relationships, or do we collapse across some syntactic/semantic properties of those verbs, etc.? In summary, predictions of experience-based accounts are highly dependent on the assumptions about exactly what linguistic experience comprehenders are sensitive to.

Evidence: Several studies have shown that the relative difficulty of ORCs vs. SRCs depends on the types of head noun and embedded NPs. For example, as predicted by the interference-based memory-based accounts, Gordon et al. (2001; also Gordon et al., 2002, 2004; Van Dyke & Lewis, 2003; Van Dyke & McElree, 2006) showed that the retrieval difficulty increases when the intervening elements are similar to the target element. As predicted by some versions of the interference-based memory-based accounts, Gordon et al. (2001) and Warren & Gibson (2002) showed that ORCs with a pronoun in the embedded subject position are easier to process than those with a full NP. However, Fedorenko et al. (2012) have shown that these effects are unlikely to be due to the fact that pronouns are old to the discourse: a robust extraction effect is observed with the full NPs even in cases where the critical sentences are presented in highly supportive contexts and all the NPs are introduced a priori. As a result, Warren & Gibson's (2002) results are most likely due to the lexical properties of the pronouns and/or to the frequencies of the pronouns in the relevant syntactic contexts. Consistent with this idea, Reali & Christiansen (2007) found that ORCs with pronominal subjects (e.g., "the barber that you admired") are actually easier to process than SRCs with pronominal objects (e.g., "the barber that admired you"). They explained these results in terms of higher frequencies of ORCs with pronominal subjects, compared to SRCs with pronominal objects (see also Troyer et al., 2011). Finally, as predicted by experience-based accounts, more frequent animacy configurations have been shown to lead to a smaller extraction effect. For example, Traxler, Morris & Seeley (2002) demonstrated that ORC difficulty is low when the RC modifies an inanimate NP and contains an animate subject (e.g., "the rock that the boy threw"), but high when the RC modifies an animate NP and contains an inanimate subject (e.g., "the mountaineer that the boulder hit"; see also Traxler et al., 2005; see Mak et al., 2002, 2006, for similar results in Dutch). In contrast, Traxler et al. (2002) found that when mere plausibility, rather than animacy, was manipulated (e.g., "The thief that the policeman arrested…") reduction in difficulty was minimal. Additionally, Gordon et al. (2006) find no reduction of ORC difficulty associated with definite as opposed to indefinite subjects, despite higher ORC frequency with definite

subjects. Together, these findings support the experience-based claim that more frequent configurations of some, if not all, syntactic/lexico-semantic properties are processed more readily.

Summary

The evidence for both the locus of the extraction effect and for the NP type manipulations is mixed. Although the bulk of the difficulty in reading is observed at the embedded verb (e.g., Grodner & Gibson, 2005), as predicted by working-memory-based accounts, some difficulty is observed at the embedded subject (e.g., Staub, 2010). With respect to the NP type manipulations, some evidence is compatible with interference-based working-memory theories. In contrast, the effects of animacy configuration reported by Traxler et al. (2002) and others are hard to account for within existing working-memory theories, and yet easily explained in an expectation-based theory that assumes abstract representations of animacy configurations. However, Traxler et al. (2002) do observe some ORC difficulty for the frequent animacy configuration in some of the eye-tracking measures, and Tily et al. (2011, in revision) find robust extraction effects in self-paced reading across all four animacy configurations. This suggests that – at least in English – ORC difficulty cannot be eliminated completely even when the relevant NPs are highly dissimilar and the event conforms to the most common animacy configuration (i.e., an animate entity acting upon an inanimate entity). One possibility then is that there are multiple independent sources of difficulty in the ORC: one immutable source originating in retrieval from working memory, and a partially or entirely independent source associated with infrequent lexico-semantic configurations. The two may appear at different positions within the structure (Lowder & Gordon, submitted; Tily, Fedorenko & Gibson, in revision; see also Bornkessel & Schlesewsky, 2006; Staub, 2010; Baudiffier, Caplan, Gaonac'h & Chesnet, 2011).

We now report two experiments that further investigate some predictions of the three kinds of theories discussed above.

Experiment 1

In Experiment 1, we used materials that consist of one RC embedded in another RC: *doubly-nested* RCs. It has long been known that doubly-nested RC materials like (2) and (3) are very difficult for people to understand (e.g., Yngve, 1960; Chomsky & Miller, 1963; Bever, 1970; Gibson, 1991; 1998; Lewis, 1996). Although several studies have investigated doubly-nested RCs using off-line complexity measures, there have been few studies of the on-line complexity of such materials (cf. Vasishth et al., 2010; see Grodner & Gibson, 2005, for some data on closely related structures).

Critically, doubly-nested RC structures allow us to investigate different theories of RC processing, including reanalysis-based theories, experience-based theories and retrieval-based memory-based theories. Here, we manipulate whether the NPs in the initial component of a doubly-nested structure could plausibly serve multiple roles with respect

to the verbs to come (resulting in reversible clauses as in (2)) or not (resulting in less or non-reversible clauses as in (3)).

(2) reversible doubly-nested RCs
a. The vandal that the thief that the policeman wounded on the leg accused with some certainty was known to the authorities.
b. The girl that the boy that the dog chased down the street liked for her smile was in sixth grade.

(3) non-reversible doubly-nested RCs
a. The jewels that the thief that the policeman arrested on the weekend stole from the vault were worth a lot.
b. The book that the boy that the dog bit on the hand read during school recess was one of the Harry Potter volumes.

The examples in (3) are instances of nested structures with non-reversible clauses: the verb "arrested" only plausibly takes "the policeman" as agent and "the thief" as patient; and the verb "stole" most plausibly takes "the thief" as agent and "the jewels" as patient. We can contrast these non-reversible examples with the highly reversible examples in (2), where all the NPs are animate and many are plausible as either the agent or the patient of the two embedded verbs.

Predictions

The theories make differing predictions at the most embedded verb and at the second verb in the doubly-nested sentences. According to both reanalysis- and experience-based theories, RTs should increase at the point where the unexpected / surprising syntactic structure occurs (see Grodner & Gibson, 2005, for a detailed discussion). Because doubly-embedded RCs are very rare in natural production, the point of highest surprisal in these structures is at the start of the most embedded RC (e.g., "that the policeman" in (2a)/(3a), "that the dog" in (2b)/(3b)). Thus these theories predict that RTs should be slowest at this point in these sentences. Critically, reanalysis- and experience-based theories predict that RTs should be fast for the most embedded verb region ("wounded"/"chased" in (2); "arrested"/"bit" in (3)) and the second embedded verb region ("accused"/"liked" in (2); "stole"/"read" in (3)) because a verb is the most expected element based on the preceding context.

The surprisal-based theory's prediction of fast RTs at the second verb depends on the reader fully understanding the structure for the input up to that point. Because these kinds of structures are very difficult, it is possible that on some trials, people may fail to fully understand them. Indeed, previous research has shown that people sometimes have only a partial representation for the initial sequence of NPs in these kinds of structures, such that the second NP may not be fully connected to the first noun phrase by the time the first verb has been processed. This may lead to the expectation for only two verb phrases following the initial sequence of three NPs: one for the most embedded NP, and one for the sentence-initial NP. Evidence in support of this kind of disjoint

representation for doubly-nested RC structures was originally provided by Frazier (1985) who suggested that ungrammatical examples like (4b) are more acceptable than grammatical but complex examples like (4a) (an intuition attributed to Janet Fodor):

(4) a. The patient who the nurse who the clinic had hired admitted met Jack.
b. The patient who the nurse who the clinic had hired met Jack.

In (4b), the second NP "the nurse" is not the subject of any verb to follow, and hence this sentence is ungrammatical. Yet, intuitively this sentence sounds surprisingly like a possible English sentence, in spite of the fact that it lacks a complete structure. Gibson & Thomas (1999) experimentally investigated modified versions of examples like (4) with verb phrases that were plausible with only one of the preceding subject NPs. They observed that omitting the intermediate VP resulted in materials that were most acceptable among the ungrammatical conditions (see also Gibson & Fedorenko, 2011, for corroborating evidence from a sentence completion task).[1]

Given that people sometimes don't fully represent the initial sequence of NPs in doubly-nested structures, we need to consider the predictions of the theories when such structures are not fully represented. It turns out that surprisal-based theories make different predictions under such circumstances for the reversible (2) vs. non-reversible (3) versions of the materials. In particular, a reader who fails to keep track of all the NPs in the preceding context will maintain some expectation for a main verb for the initial subject NP at the point when s/he is reading the second verb. The second verb differs across the two conditions with respect to its predictability from the initial NP: in the reversible condition, this verb is semantically compatible with the initial NP (e.g., "accused" is a plausible verb for the subject NP "the vandal" in (2a)), but in the non-reversible condition the second verb is not semantically compatible with the initial NP (e.g., "stole" is not an appropriate verb for the subject NP "the jewels" in (3a)). Consequently, if the reader does not keep a full representation of the intermediate NP, surprisal-based theories predict relatively fast RTs for the second verb for the reversible condition, but slower RTs for the non-reversible condition. Alternatively, if the reader does manage to keep a full representation for the initial sequence of NPs, then surprisal-based theories predict relatively fast RTs for the second verb independent of the reversibility of the materials.

---

[1] Fodor (this volume) attributes part of the complexity of doubly-nested examples like (2)-(4) to the relative phonological length of the NPs and VPs in the examples, leading to potential difficulties in intonational phrasing of the examples. While intonational phrasing may be a factor that contributes to the complexity of such examples, we think that its effect must be weak relative to other sources of difficulty, because it does not explain the large asymmetry between SRCs and ORCs in the embedded clauses in doubly-nested structures. For example, there is a large complexity difference between ORC/ORC examples like (2a), and ORC/SRC examples like (i) (Gibson & Thomas, 1996, 1997; Gibson & Fedorenko, 2011):

i. The vandal that the thief that wounded the policeman on the leg accused with some certainty was known to the authorities.

The most embedded RC in these examples consists of the same words across ORC and SRC versions, and yet the ORC/ORC examples are much more complex than the ORC/SRC versions. It seems difficult for an intonational phrasing based account to explain this kind of difference.

Let us now consider the predictions of retrieval-based theories. Retreival-based theories predict that the greatest complexity should occur at the embedded verbs in the RCs in each of (2) and (3), e.g., at "arrested"/"stole" in (2a); at "bit"/"read" in (2b); at "wounded"/"accused" in (3a); and at "chased"/"liked" in (3b). Furthermore, interference-based retrieval theories predict differential difficulty according to reversibility at both verbs: non-reversible materials should be faster than reversible ones because the lexical semantic properties of the items make retrieval of the appropriate subject and object NPs easier for the non-reversible versions compared to the reversible versions. In contrast, the decay-based retrieval theory predicts difficulty at the embedded verbs, but not according to reversibility.

Finally, memory-based theories make differing predictions regarding the relative difficulty of processing the most embedded verb vs. the second verb in these structures. Simple linear-distance-based retrieval theories predict that the second embedded verb ("stole") should be substantially harder to process than the most embedded verb ("arrested") since both the object and subject of the second verb are non-local, whereas only the object of the first verb is non-local. However, these theories also may appeal to storage costs, which are greater for the most embedded verb compared to the second verb, because of the additional open dependencies in the more embedded positions (Chen, Gibson & Wolf, 2005; Grodner, Gibson & Tunstall, 2002). Thus, a theory with both storage cost and linear distance retrieval cost would be consistent with any possible pattern of RTs, depending on the relative strength of the storage and retrieval costs (and possibly how they interact). In contrast, the activation decay-based theory proposed by Lewis and colleagues predicts little or no difference between the two verbs, since the second verb's subject ("thief") is reactivated at the most embedded verb, making the dependency effectively local.

Methods

Participants

We recruited 100 participants with IP addresses within the United States using Amazon.com's Mechanical Turk service. Participants were asked to indicate their native language and their country of residence, but payment was not contingent on their responses to these questions. They were paid $1.50 for their participation, which took on average 12 minutes per participant.

Materials

Eighteen sets of sentences were constructed in six conditions, as in (5) below. In particular, each item had four singly-embedded object-extracted RC versions – two non-reversible, and two reversible – and each pair of singly-embedded RCs were combined to form a doubly-embedded object-extracted RC, which was either non-reversible or reversible:

(5)

a. non-reversible, inanimate initial noun, outer clause in doubly-embedded version:
The jewels that the thief stole from the vault were worth a lot.

b. non-reversible animate initial noun, inner clause in doubly-embedded version:
The thief that the policeman arrested on the weekend was known to carry a weapon.

c. non-reversible doubly-nested RC:
The jewels that the thief that the policeman arrested on the weekend stole from the vault were worth a lot.

d. reversible, outer clause in doubly-embedded version:
The vandal that the thief accused with some certainty was known to the authorities.

e. reversible, inner clause in doubly-embedded version:
The thief that the policeman wounded on the leg was known to carry a weapon.

f. reversible doubly-nested RC:
The vandal that the thief that the policeman wounded on the leg accused with some certainty was known to the authorities.

In (5a), the only plausible object of the verb "stole" is "jewels". In contrast, in (5d), either "thief" or "vandal" is a plausible object for "accused". There is a similar contrast between (5b) and (5e): in (5b) the only plausible object of the verb "arrested" is "thief", whereas in (5e), either "thief" or "policeman" is a plausible object for "wounded".

The non-reversible doubly-embedded versions (5c) were constructed so that the most embedded verb would only plausibly take the immediately preceding NP as its subject, and the second NP as its object. For example, only "policeman" can plausibly be the subject of "arrested" and only "thief" can plausibly be its object. Furthermore, the materials were constructed so that the second NP is most plausible as the subject of the second verb, and the initial NP is the only NP that is plausible as the object of this verb. For example, it is most plausible for "thief" to be the subject of "stole" in (5c) and only "jewels" can plausibly be the object of this verb. In contrast, in the reversible doubly-nested versions in (5f), all the NPs are animate and many are plausible to be the either the subject or the object of the two embedded verbs.

To ensure that the participants understood the sentences, a yes/no comprehension question about the propositional content of the sentence was asked at the end of each trial. For the doubly-nested RC conditions (c and f above), the questions were distributed evenly among three types of questions with 6 questions of each type: about the outer clause (3 yes, and 3 no), about the middle clause (3 yes, and 3 no), and about the inner clause (3 yes, and 3 no). For the singly-nested RC conditions (a-b and d-e above), the questions were distributed evenly between two types of questions with 9 questions of each type: about the outer clause (5 yes, and 4 no), and about the inner clause (4 yes, and 5 no).

The experiment also included 36 filler sentences (with corresponding comprehension questions), which were similar to the critical sentences in their length, but were generally less complex.

Procedure

The sentence-reading task used self-paced moving-window word-by-word reading (Just, Carpenter & Woolley, 1982), which ran in the participants' web browser. Words were presented one at a time. Following each sentence, participants were asked a simple comprehension question.

The web-based self-paced reading software has been shown in previous research to replicate results from the laboratory (Tily, Fedorenko & Gibson, 2011, in preparation).

Results

Before analyzing the self-paced reading data, we excluded participants (a) that didn't complete the survey (6 participants); (b) that didn't identify themselves as native speakers of American English (6 additional participants); and (c) that didn't answer at least 80% of the comprehension questions correctly (1 additional participant: mean accuracy for this participant = 74%). This left 87 participants for the self-paced reading analyses, all of whom answered 83% or more of the comprehension questions correctly (mean = 93%). Average accuracies across conditions are shown in Table 1.

Single-embedded conditions	Accuracy
Non-reversible, inanimate initial noun, outer clause	.967 (.01)
Non-reversible, animate initial noun, inner clause	.980 (.01)
Reversible, animate initial noun, outer clause	.926 (.02)
Reversible, animate initial noun, inner clause	.948 (.01)
Double-embedded conditions	
Non-reversible	.924 (.02)
Reversible	.802 (.03)
Distractor items	.931 (.01)

Table 1. Comprehension accuracy across the six conditions in Experiment 1 (standard errors of the mean in the parentheses).

Before analyzing the reading time (RT) data, we first excluded extremely fast (<50 msec) and extremely slow (>5 sec) RTs. We then excluded RTs that were more than three standard deviations faster or slower than mean RTs for each word position by condition, across participants and items. These exclusion procedures affected less than 2.3% of the data. We first present the data from the singly-embedded conditions (5a/b/d/e), and then the data from the critical doubly-embedded conditions (5c/f).

Analysis of singly-embedded conditions

Mean reading times (RTs) per word are presented in Figure 1. The critical region is the embedded clause. Consequently, we consider the RTs at the embedded verb and the regions immediately before and after for comparison.

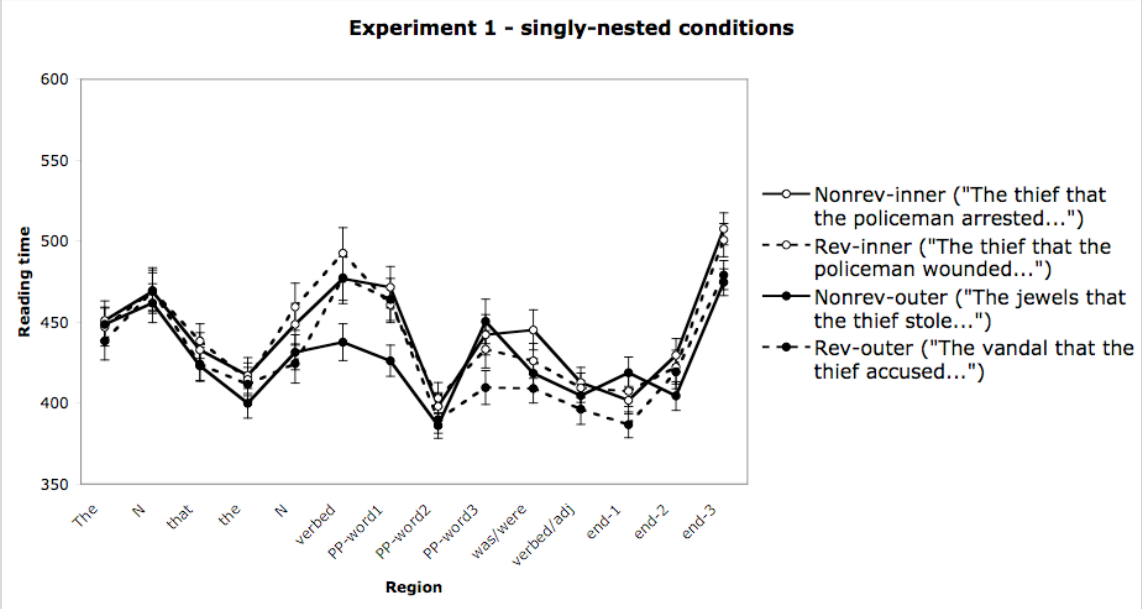

Figure 1: Average reading times for the singly-embedded relative clause conditions in Experiment 1. Error bars indicate standard errors of the mean across participants.

At the embedded verb ("arrested"/"accused") we found slower reading times as compared to the two previous regions ($F1(1, 86)=29.83$, $p<.001$; $F2(1, 17)=59.90$, $p<.001$) and a significant interaction between reversibility and region ($F1(1, 86)=8.36$, $p=.004$; $F2(1, 17)=6.98$, $p=.02$). The animacy manipulation was marginally significant over the entire region in the by-subjects analysis ($F1(1, 86)=3.23$, $p=.08$) but not in the by-items analysis or in interaction with region or reversibility (all $Fs<.42$, $ps>.53$). Looking just at RTs at the embedded verb, we found a similar pattern, with the only significant effect being that of reversibility ($F1(1, 86)=7.32$, $p=.008$; $F2(1, 17)=5.33$, $p=.03$; all other $Fs<1.76$, $ps>.19$). These results are most compatible with working memory-based theories: difficulty is experienced primarily beginning at the verb, and keeping confusable NPs in memory increases difficulty at that point. At the word following the embedded verb, we found an interaction between reversibility and animacy that was reliable in the

participants analysis and marginal in the items analysis (F1(1, 86)=7.90, p=.006; F2(1, 17)=3.26, p=.09). No other effects were significant (all Fs<2.08, ps>.15).

Analysis of doubly-embedded conditions.

The critical conditions in this experiment were the doubly-nested conditions. Mean reading times per word are presented in Figure 2.

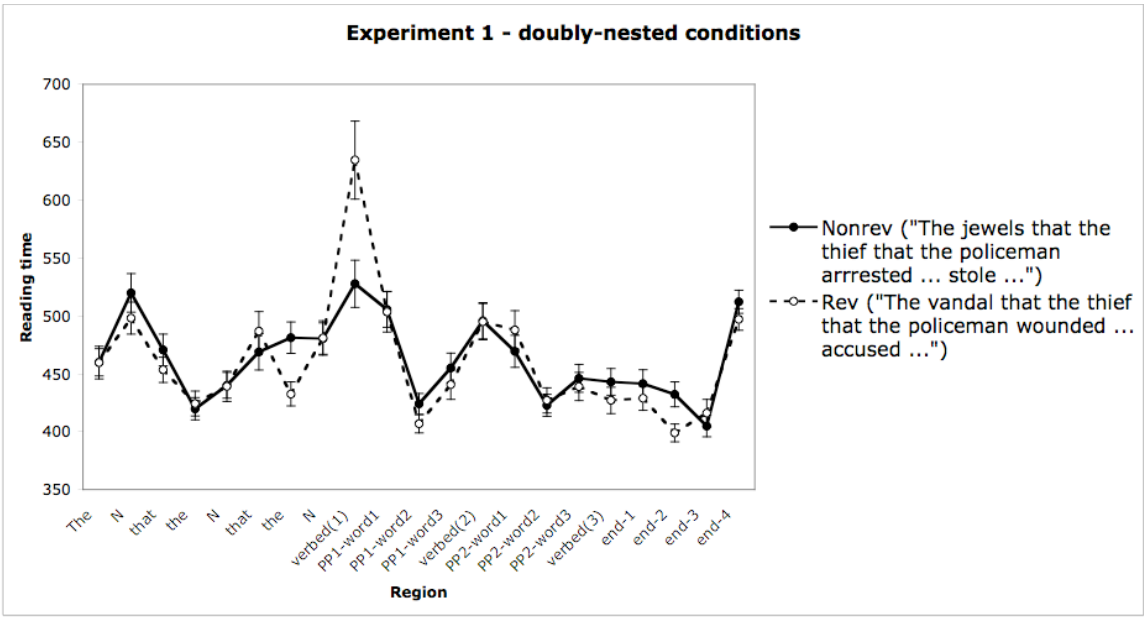

Figure 2: Average reading times for the doubly-embedded relative clause conditions in Experiment 1. Error bars indicate standard errors of the mean across participants.

We considered two critical regions: the first embedded verb, and the second embedded verb region. We also examine reading times in the regions immediately before and after these regions for comparison.

At embedded verb region 1 (“wounded”/“arrested”) we observed a substantial RT slow-down as compared with the preceding two words (F1(1, 86)=27.93, p<.001; F2(1,17)=39.37, p<.001). As can be seen from the plot, this effect was driven by substantially slower RTs at the verb as compared to either word in the embedded subject NP. This locus of effect is predicted by memory-based theories, and not by reanalysis- or expectation-based theories. Additionally, we observed an interaction between reversibility and region which was significant in the participants analysis and marginal in the items analysis (F1(1, 86)=5.35, p=.02; F2(1, 17)=4.19, p=.06). This interaction is driven by the condition effect emerging primarily at the embedded verb: in a separate analysis of data from only the embedded verb we found a similar pattern (F1(1, 86)=4.74, p=.03; F2(1, 17)=3.37, p=.08). Reversible materials led to higher RTs at the embedded verb, as predicted by interference-based theories: in reversible sentences, the embedded subject (“thief”) is similar to the modified noun (“vandal”) and therefore may interfere with retrieval of the dependent from working memory. We also analyzed the word

following the verb, but found no significant effect of condition ($F1(1, 86)=.88$, $p=.35$; $F2(1, 17)=.24$, $p=.63$). Thus it appears that the locus of retrieval difficulty occurred primarily at the verb itself and did not lead to substantial slow-down thereafter.

Next, we looked at RTs at the second verb ("accused"/"stole"). Here again we found an increase in RTs compared to the two previous regions ($F1(1, 86)=20.15$, $p<.001$; $F2(1, 17)=43.55$, $p<.001$) but no interaction with condition and no main effect of condition at any of the regions (all $Fs<.53$, $ps>.48$).

Finally, we compared RTs for the first verb region with RTs for the second verb region, and we found that RTs on the first verb were significantly slower than RTs on the second verb ($F1(1, 86)=20.41$, $p<.001$; $F2(1, 17)=27.79$, $p<.001$) and there was an interaction, such that RTs on the first verb showed a significantly larger reversibility effect ($F1(1, 86)=6.05$, $p=.02$; $F2(1, 17)=4.84$, $p=.04$). The greater RTs for the most embedded verb are predicted by storage costs: when people have to keep track of extra RC dependency, RTs increase, especially for complex structures (Chen et al., 2005; Nakatani & Gibson, 2010). This result is not consistent with a simple linear distance retrieval theory, which would predict RTs for the second verb to be slower than RTs on the first verb, because the retrieval distances are longer at the second verb. The result is also not predicted by the activation decay-based theory proposed by Lewis and colleagues, which predicts little or no difference between the two.

Experiment 2

Although the results from Experiment 1 generally support the predictions from retrieval-based theories, the evidence comes from the processing of doubly-nested object-extracted RCs, sentence materials that people have a great deal of difficulty understanding. Thus the RT pattern is potentially somewhat hard to interpret, if people sometimes do not have a complete representation for such structures, as discussed above. Evidence that participants understood the doubly-nested materials in the experiment is provided by the relatively high accuracy rates, 80.2% for the reversible versions and 92.4% for the non-reversible versions. The accuracy rate is especially high for the non-reversible doubly-nested versions, approaching the highest levels of accuracy that one finds in self-paced reading experiments, even for much simpler materials. The accuracy rate for the reversible doubly-nested versions is lower, but still far above chance, indicating that participants understood a lot of the content in these materials.

Even so, it would be useful to complement these results by an investigation of materials that are somewhat easier. We opted for a design in which the subject of a singly-embedded object-extracted RC was modified by increasingly more material, but never another object-extracted RC (see Grodner & Gibson, 2005, for a similar design). In particular, we compared singly-embedded materials (as those in Experiment 1) and materials in which the embedded subject is modified by a prepositional phrase (PP) vs. by a subject-extracted RC whose predicate consisted of the same PP as in the PP conditions. There was therefore more material between the embedded subject and its verb across the three conditions, with plausibility controlled across the PP and SRC

versions, because their content was the same. We crossed the embedding manipulation (bare, PP modification, SRC modification) with the reversibility of the outer ORC (non-reversible, reversible), to result in six conditions as in (6):

(6)
a. non-reversible, bare embedded NP:
The jewels that the thief stole from the vault were worth a lot.
b. non-reversible, embedded NP modified by PP:
The jewels that the thief *from New York* stole from the vault were worth a lot.
c. non-reversible, embedded NP modified by SRC:
The jewels that the thief *who was from New York* stole from the vault were worth a lot.
d. reversible, bare embedded NP:
The vandal that the thief accused with some certainty was known to the authorities.
e. reversible, embedded NP modified by PP:
The vandal that the thief *from New York* accused with some certainty was known to the authorities.
f. reversible, embedded NP modified by SRC:
The vandal that the thief *who was from New York* accused with some certainty was known to the authorities.

Although the embedded SRC conditions (6c)/(6f) are technically doubly-embedded RCs, they are much less complex than doubly-nested object-extracted RCs, as evidenced by significantly lower complexity ratings in Gibson & Thomas's (1996, 1997) studies and by much higher syntactically correct completion rates in Gibson & Fedorenko's (2011) sentence completion studies. For example, in Gibson & Fedorenko's (2011) completion studies, participants were able to complete inanimate / animate materials like "The manuscript which the writer who…" correctly (with three VPs) 64.6% of the time (usually with a subject-extracted RC in the most embedded clause), and they were able to complete animate / animate materials like "The reporter who the professor who …" correctly 57.4% of the time. A plausible source for the lower complexity of these ORC/SRC materials compared with ORC/ORC materials is that the most embedded clause does not have a lexical NP subject (which is usually animate), which will often interfere with the retrieval of earlier NP subjects in the ORC/ORC versions, as in (5c)/(5f) (Gordon, Hendrick & Johnson, 2001, 2004; McElree, Foraker & Dyer, 2003; Lewis & Vasishth, 2005; Lewis, Vasishth & Van Dyke, 2006). Thus we thought participants should be able to understand these materials well.

Predictions

As in Experiment 1, the theories make differing predictions at the verb associated with the second RC ("stole"/ "accused" in (6)). (Unlike Experiment 1, there is no second object-extracted RC across the conditions to compare RTs on.) As discussed above, both reanalysis- and experience-based theories predict that RTs should increase only at points where unexpected / surprising syntactic structures occur. The embedded verb "stole"/ "accused" is always a highly expected element in each of the conditions in (6), so RTs should not increase at this point relative to the preceding regions, where more surprising

material occurs. In contrast, retrieval-based theories predict relatively slow RTs at the verb "stole"/ "accused", because this is the point at which a long-distance dependency must be integrated. The decay-based retrieval theory predicts increasing difficulty across the types of NP modification, such that the bare NP condition should be fastest, followed by the PP condition, with the SRC condition slowest, because it includes the longest modifiers of the embedded NP separating the head NP with its verb. On the other hand, the inference-based retrieval theory predicts differential difficulty according to reversibility: non-reversible conditions should be faster than reversible ones because the lexical semantic properties of the items make retrieval of the appropriate subject and object NPs easier for the non-reversible versions compared to the reversible versions.

Finally, we can further evaluate the activation decay-based theory proposed by Lewis and colleagues using these materials. No difference among the three NP conditions is predicted at the verb retrieval site under this proposal, because the embedded NP is reactivated during/after processing either the PP or the SRC, making all three conditions equally local according to this proposal. Thus between-condition differences for the distance manipulation would be difficult for this proposal to explain.

Methods

Participants

We recruited 100 new participants with IP addresses within the United States using Amazon.com's Mechanical Turk service. They were paid $1.50 for their participation, which took on average 11 minutes per participant.

Materials

Eighteen sets of sentences were constructed in six conditions, as in (6) above. The bare embedded NP versions consisted of the simple outer clause conditions from Experiment 1 (e.g., (5a) for the non-reversible bare embedded NP condition and (5d) for the reversible bare embedded NP condition). To form the PP modifier conditions, we simply added a 3-word PP whose NP object was a proper name of a location (such as "New York" or "the South") or an inanimate NP which could not plausibly be the subject of the verbs to come (e.g., "in a hurry", "during school recess", "in his twenties"). Thus this NP should not interfere with retrievals of NP subjects, in a retrieval-based model of RC processing. To form the SRC versions from the PP versions, the words "who was" were added. Thus the PP and SRC versions are controlled for plausibility.

To ensure that the participants understood the sentences, a yes/no comprehension question about the propositional content of the sentence was asked at the end of each trial. For the bare embedded NP conditions, an equal number of questions were asked about the inner and outer clauses, balanced as closely as possible between yes and no answers for each (and balanced overall). For the other conditions, there were 6 questions about each of the main clause, the first embedded clause and the most embedded SRC/PP, balanced yes/no in each.

The experiment also included 36 filler sentences (with corresponding comprehension questions), which were similar to the critical sentences in their length.

Procedure

The procedure was the same self-paced moving-window word-by-word reading over Amazon.com's Mechanical Turk as in Experiment 1.

Results

Before analyzing the self-paced reading data, we excluded participants (a) that didn't complete the survey (2 participants); (b) that didn't identify themselves as native speakers of American English (6 additional participants); and (c) that didn't answer at least 80% of the comprehension questions correctly (3 additional participants: mean accuracy for these participants = 75.9%, 79.6%, 79.6%). This left 89 participants for the self-paced reading analyses, all of whom answered 83% or more of the comprehension questions correctly (mean = 92%). Average accuracies across conditions are shown in Table 1.

	Accuracy
Non-reversible, bare embedded NP	.927 (.02)
Non-reversible, embedded NP modified by PP	.896 (.02)
Non-reversible, embedded NP modified by SRC	.872 (.02)
Reversible, bare embedded NP	.880 (.02)
Reversible, embedded NP modified by PP	.837 (.02)
Reversible, embedded NP modified by SRC	.859 (.02)
Distractor items	.936 (.01)

Table 2. Comprehension accuracy across the six conditions in Experiment 2 (standard errors of the mean in the parentheses).

Before analyzing the RT data, we first excluded extremely fast (<50 msec) and extremely slow (>5 sec) RTs. We then excluded RTs that were more than three standard deviations faster or slower than mean RTs for each word position by condition, across participants and items. These exclusion procedures removed 2.3% of the data.

Mean RTs per word are presented in Figure 3. We consider the RTs at the embedded verb and the regions immediately before and after for comparison.

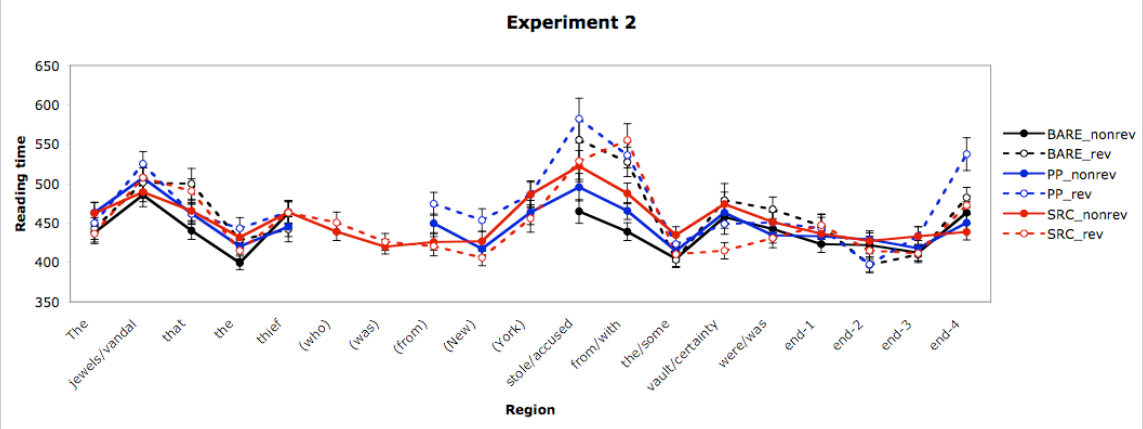

Figure 3: Average reading times for the conditions in Experiment 2. Error bars indicate standard errors of the mean across participants.

We analyzed the three word positions up to and including the verb; one subject was removed from the by-subjects analysis due to all observations in one condition being missing due to trimming of extreme values. As can be seen from the figure, the RTs spike at the embedded verb "stole"/ "accused" for most of the conditions. We see this as a main effect of position looking across these three word positions ending in the verb ($F1(2,174)=27.84$, $p<.001$; $F2(2, 34)=49.09$, $p<.001$). We also see interactions between position, embedding and reversibility over these three positions (position * embedding * reversibility: $F1(4, 348)=3.48$, $p<.01$; $F2(4, 68)=2.43$, $p=.06$). This effect seems to be driven by the elevated RTs for all the reversible conditions and the SRC non-reversible.

Looking at the verb region, we find marginal effects of reversibility ($F1(1, 87)=3.23$, $p=.08$; $F2(1, 17)=.63$, $p=.44$), embedding ($F1(2, 174)=3.10$, $p=.05$; $F2(1, 17)=1.56$, $p=.22$), and an interaction between the two ($F1(2, 174)=5.21$, $p<.01$; $F2(1, 17)=2.89$, $p=.07$). However, it appears that there is some spillover with elevated RTs occurring also on the word following the verb. Consequently, we analyzed this average word RTs in this two-word region. See Figure 4 for a bar graph of the RTs in this region.

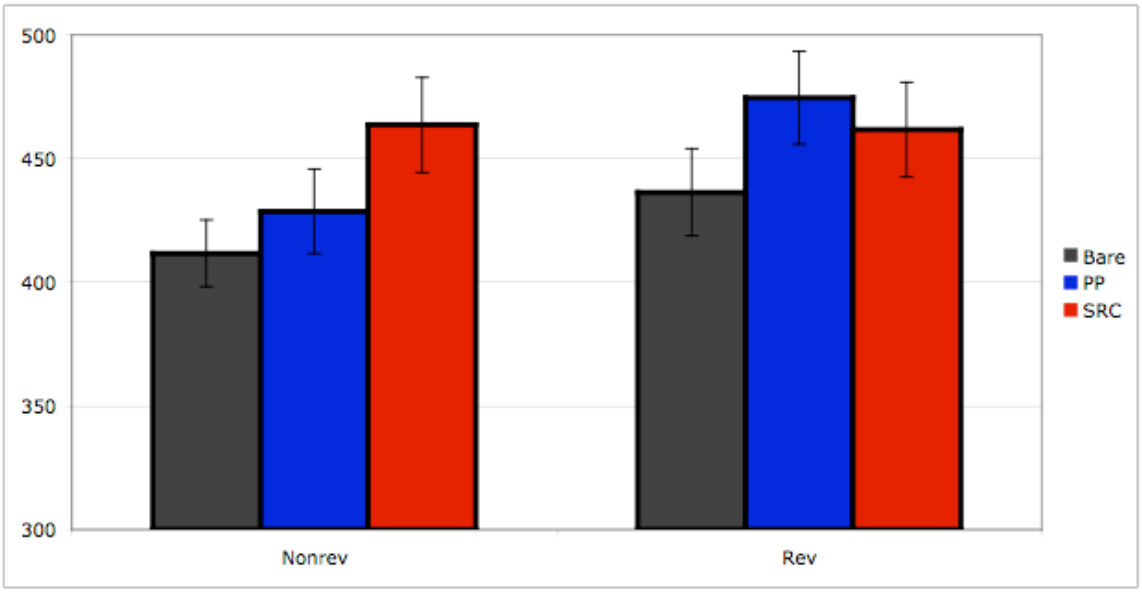

Figure 4: Average reading times for the embedded verb and the following word in Experiment 2. Error bars indicate standard errors of the mean across participants.

Two subjects were removed from the by-subjects analysis due to all observations in one condition being missing. As can be seen from the RTs in Figure 4, there is a main effect of reversibility at this region, such that reversible materials are processed more slowly than non-reversible materials in the by-subjects analysis and marginally so by items ($F1(1,85)=12.53$, $p<.001$; $F2(1, 17)=3.99$, $p=.06$). Furthermore, there is a main effect of the length of the intervener, such that longer intervening elements lead to higher RTs on average (NP with SRC is slowest across conditions; NP with PP next; bare NP is fastest) ($F1(2, 170)=11.08$, $p<.001$; $F2(2, 34)=4.16$, $p=.02$). Finally, the two factors interact significantly by subjects and marginally by items ($F1(2, 170)=3.91$, $p=.02$; $F2(2, 34)=2.32$, $p=.11$) such that the complexity is at its maximum with reversible materials when only a PP intervenes, but for non-reversible materials it is only when the SRC intervenes when the maximal RTs occur.

Overall, these results are highly consistent with retrieval-based theories of relative clause processing. The effect of reversibility is as predicted by the interference-based theories, and the linear-distance effect is as predicted by the decay-based theories. Thus it appears that there are both interference and decay components to retrieval, as expected under models like those of Boston et al. (2008, 2011) and Lewis, Vasishth & Van Dyke (2006), but in order to explain these results, it is critical that the activation decay-based model does not reactivate the embedded noun at the completion of processing the PP and SRC modifiers, otherwise none of the decay effects could be accounted for.

General Discussion

Across two self-paced reading experiments, we have provided evidence in support of retrieval-based theories of English RC processing. In Experiment 1, we critically showed

that RTs in English doubly-nested object-extracted RCs peak at the embedded verbs and immediately after, as predicted by retrieval-based theories, and in contrast to experience-based and reanalysis-based theories. We also showed evidence for the interference-based retrieval theories, in the form of reversibility effects at the most embedded verb. Finally, we provided some evidence that RTs were highest at the most embedded verb in doubly-nested RC structures, which is most consistent with storage-based memory theories of RC processing. In Experiment 2, we provided further evidence for interference-based retrieval theories, in the form of reversibility effects, such that reversible materials were processed more slowly at the embedded verb than non-reversible materials. Furthermore, we provided a lot of evidence in support of decay-based theories, such that longer distance retrievals were more costly even when there were no additional interfering elements. We therefore find evidence for both interference- and decay-based theories of retrieval in sentence processing.

We close this paper with a brief discussion of how retrieval-based theories might play a role in the processing of languages with word orders which are different from English in RCs. English is an SVO language, whose RCs follow their head nouns (which is typical for SVO languages). There has been a lot of recent discussion of how RCs are processed in SOV (head-final) languages, such as Japanese and Korean. In these languages, it has often been stated that linear-distance-based theories like the dependency locality theory of Gibson (1998, 2000) (which are retrieval-based theories) predict that object-extracted RCs should be easier to process than subject-extracted RCs in these kinds of languages (e.g., Kwon et al., 2006, 2010; Miyamoto & Nakamura 2003; Ishizuka, Nakatani & Gibson, 2003; Ueno & Garnsey, 2008). The argument goes as follows. Consider the word order for an SRC vs. an ORC in a language like Korean or Japanese:

(7) Head-final relative clause word order:
a. Subject-extracted RC: [RC __i NP-acc V] NP_i
b. Object-extracted RC: [RC NP-nom __i V] NP_i

The connection between the empty position (notated as "__i" in (7)) and the verb is local in the ORC (7b) but non-local in the SRC in (7a), where this dependency crosses the accusatively-marked object NP. Thus, it has been argued that linear-distance-based theories predict that ORCs should be less complex than SRCs in head-final languages. But this version of a linear-distance-based theory is not a retrieval theory. A retrieval theory predicts no difference between the two structures, because the empty position is not retrieved at the verb: this is the position that the empty position is first posited in each structure, and there is no retrieval difference in this construction. When the head noun is processed next, an activation-based retrieval theory predicts no difference between the retrieval of subject or object position, because each was last activated at the verb. The typical result in verb-final languages is that ORCs are more complex than SRCs (Kwon et al., 2006, 2010; Miyamoto & Nakamura 2003; Ishizuka et al., 2003; Ueno & Garnsey, 2008). This result is not explained by retrieval theories, but may be explained by surprisal-based theories, given that SRCs are more common than ORCs in these languages.

In contrast to verb-final languages like Japanese and Korean, however, retrieval theories predict that a language with SVO word order in main clauses and head-final RCs should result in more complex SRCs compared to ORCs. An example of such a language is Chinese: whereas Chinese relative clauses are prenominal (like Japanese and Korean RCs), the basic word order in main clauses is subject-verb-object, as in English and in contrast to Japanese, Korean and other head-final languages. Consider (8) for example (from Hsiao & Gibson, 2003):

(8)
a. Subject-extracted relative clause
__ yaoqing fuhao de guanyuan xinhuaibugui
__ invite tycoon REL official have bad intentions
'The official who invited the tycoon had bad intentions.'
b. Object-extracted relative clause
fuhao yaoqing __ de guanyuan xinhuaibugui
tycoon invite __ REL official have bad intentions
'The official who the tycoon invited had bad intentions.'

Because of the different position of RCs with respect to their head nouns and the word order within the RCs, retrieval-based theories predict that SRCs should be more complex than ORCs in Chinese. In particular, consider the processing steps during the RC, and at the relative clause marker "de" (roughly corresponding to English "that" or "who") following the RC and the head noun for the RC. During the RC ("__ invite tycoon" or "tycoon invite __") the integrations are local and matched across the two structures. Next, the RC marker "de" and the head noun for the RC are processed. The head noun for the RC needs to be linked with the empty NP position in the RC. This is a local integration in the ORC structure, because the object position and the verb are the most recent positions that have been processed. In contrast, this is a more distant integration in the SRC structure, because the object noun phrase intervenes. Thus retrieval-cost memory-based theories predict that SRCs should be more complex than ORCs in Chinese. Indeed, a few studies have provided evidence for SRCs being more complex than ORCs in Chinese (e.g., Hsiao & Gibson, 2003; Gibson & Wu, 20011; Lin & Garnsey, 2007; cf. Lin & Bever, 2006; Kuo and Vasishth, 2006).

In summary, we reported two self-paced reading experiments investigating syntactically complex English structures. Evidence from both experiments provides support for working-memory-based accounts of syntactic complexity, including both decay-based and interference-based accounts, and this evidence is not easily explained by reanalysis- or experience-based accounts. As discussed in the introduction, however, neither a purely memory-based or a purely experience-based account appears to account for all previously reported patterns of data. As a result, the data reported here should not be viewed as evidence for memory-based and against experience-based accounts, but rather as another set of results from complex syntactic structures that needs to be explained by any broad-coverage account of processing complexity. Such an account would almost certainly have to include both i) a mechanism for keeping track of our linguistic experiences and then using this knowledge to form expectations about the likelihoods of various upcoming elements, and ii) a memory component engaged when dependencies

are formed in the course of online comprehension. As suggested by Boston et al. (2008, 2011), Demberg & Keller (2009), Staub (2010) and Lowder & Gordon (to appear), an adequate account of relative clause comprehension will incorporate multiple distinct components, rather than relying on a single mechanism to explain all sources of comprehension difficulty.

References

Baudiffier, V., Caplan, D., Gaonac'h, D. & Chesnet, D. (2011). The effect of noun animacy on the processing of unambiguous sentences: Evidence from French relative clauses. *Quarterly Journal of Experimental Psychology,* 64(10), 1896-1905.

Bever, T.G. (1970). The cognitive basis for linguistic structures. In: J.R. Hayes, Editor, Cognition and the development of language, Wiley, New York, pp. 279–362.

Bornkessel, I. & Schlesewsky, M. (2006). The Extended Argument Dependency Model: A neurocognitive approach to sentence comprehension across languages. *Psychological Review*, 113, 787-821.

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

Boston, M.F., Hale, J.T., Vasishth, S. & Kliegl, R. (2011). Parallel processing and sentence comprehension difficulty. *Language and Cognitive Processes*, 26(3):301-349.

Carreiras, M., Duñabeitia, J.A., Vergara, M., de la Cruz-Pavía, I., Laka, I. (2010) Subject relative clauses are not universally easier to process: Evidence from Basque. Cognition, 115, 19-92.

Chen, E., Gibson, E. & Wolf. F. (2005). Online syntactic storage costs in sentence comprehension. *Journal of Memory and Language*, 52, 144-169.

Chomsky, N., & Miller, G. A. (1963). Introduction to the formal analysis of natural languages. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.). Handbook of mathematical psychology (Vol. 2, pp. 269–321). New York: Wiley.

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

Demberg, V. and Keller, F. (2009). A computational model of prediction in human parsing: Unifying locality and surprisal effects. In Proceedings of the Cognitive Science Society.

Frazier, L. & Clifton, C., Jr. (1989) Successive cyclicity in the grammar and the parser. *Language and Cognitive Processes*, 4(2), 93-126.

Fedorenko, E., Piantadosi, S. & Gibson, E. (2012). Processing relative clauses in supportive contexts. *Cognitive Science*.

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.

Forster, K. I., Guerrera, C. & Elliot, L. (2009). The maze task: Measuring forced incremental sentence processing time. *Behavioral Research Methods*, 41, 163-171.

Frazier, L. (1985). Syntactic complexity. In: Dowty, D., Karttunen, L., Zwicky, A. (Eds.), Natural Language Processing: Psychological, Computational and Theoretical Perspectives. Cambridge University Press, Cambridge, UK, pp. 129–189.

Frazier, L. (1987a). Syntactic processing evidence from Dutch. *Natural Language and Linguistic Theory*, 5, 519–559.

Frazier, L. (1987b). Sentence processing: a tutorial review. In: Coltheart, M. (Ed.), Attention and Performance XII. Erlbaum, Hillsdale, NJ, 559–585.
Frazier, L. & Flores d'Arcais, G.B. (1989). Filler-driven parsing: A study of gap filling in Dutch. *Journal of Memory and Language*, 28, 331-344.
Frazier, L. & Fodor, J.D. (1978). The sausage machine: A new two-stage parsing model. *Cognition,* 6, 291-325.
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.
Garnsey, S.M., Pearlmutter, N.J., Myers, E., Lotocky, B. (1997). The relative contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language*, 37, 58–93.
Gennari, S. & MacDonald, M. (2008). Semantic indeterminacy in object relative clauses. *Journal of Memory and Language*, 58, 161-187.
Gennari, S. & MacDonald, M. (2009). Linking production and comprehension processes: The case of relative clauses. *Cognition*, 111, 1-23.
Gibson, E. & Fedorenko, E. (2011). The domain-generality of working memory resources for language. AMLaP conference. Paris, France.
Gibson, E. & Thomas, J. (1999). Memory limitations and structural forgetting: The perception of complex ungrammatical sentences as grammatical. *Language and Cognitive Processes*, 14, 225-248.
Gibson, E., Thomas, J., 1996. The processing complexity of English center-embedded and self-embedded structures. In: Schu¨tze, C. (Ed.), Proceedings of the NELS 26 Sentence Processing Workshop, *MIT Occasional Papers in Linguistics*, 9, Cambridge, MA, 45–71.
Gibson, E. & Thomas, J. (1997). The complexity of nested structures in English: Evidence for the syntactic prediction locality theory of linguistic complexity. Unpublished Manuscript.
Gibson, E. & Wu, I. (2011). Processing Chinese relative clauses in context. *Language and Cognitive Processes*.
Gibson, E., Desmet, T., Grodner, D., Watson, D. & Ko, K. (2005). Reading relative clauses in English. *Cognitive Linguistics*, 16, 313-354.
Gibson, E. (1991). A computational theory of human linguistic processing: Memory limitations and processing breakdown. Unpublished Ph.D. Thesis.
Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68, 1-76.
Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In Miyashita, Y., Marantz, A., & O'Neil, W. (Eds.), Image, language, brain. MIT Press, Cambridge, MA.
Gibson, E. (2007). Locality and anti-locality effects in sentence comprehension. Workshop on processing head-final languages, Max-Planck Institute for Cognitive Neuroscience, Leipzig, Germany.
Gordon, P., Hendrick, R., & Johnson, M. (2001). Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 27(6), 1411-1423.

Gordon, P., Hendrick, R., & Levine, W. (2002). Memory-load interference in syntactic processing. *Psychological Science*, 13, 425-430.
Gordon, P., Hendrick, R., & Johnson, M. (2004). Effects of noun phrase type on sentence complexity. *Journal of Memory and Language*, 51, 97-114.
Gordon, P., Hendrick, R., Johnson, M. & Lee, Y. (2006). Similarity-Based Interference During Language Comprehension: Evidence from Eye Tracking During Reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(6), 1304-1321.
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.
Hale, J. (2003). The Information Conveyed by Words in Sentences. *Journal of Psycholinguistic Research*, 32(2), 101-123.
Hawkins, J. (1994). A performance theory of order and constituency. Cambridge, UK: Cambridge University Press.
Hawkins, J.A. (2004). Efficiency and Complexity in Grammars. Oxford: Oxford University Press.
Holmes, V.M., O'Regan, J.K. (1981). Eye fixation patterns during the reading of relative clause sentences. *Journal of Verbal Learning and Verbal Behavior,* 20, 417–430.
Hsiao, F. & Gibson, E. (2003). Processing relative clauses in Chinese. *Cognition*, 90, 3-27.
Ishizuka, T., Nakatani, K. & Gibson, E. (2003). Relative clause extraction complexity in Japanese. Poster presented at the 16th annual CUNY conference on human sentence processing. Massachusetts Institute of Technology, Cambridge, MA.
Just, M., Carpenter, P., & Woolley, J. (1982). Paradigms and processing in reading comprehension. *Journal of Experimental Psychology: General*, 111, 228-238.
Keenan, E. & Comrie, B. (1977). Noun Phrase Accessibility and Universal Grammar. *Linguistic Inquiry*, 8, 63-99.
Keenan, E. (1975). Variation in Universal Grammar. In R. Fasold & R. Shuy (Eds.), Analyzing Variation in Language. Washington, D.C.: Georgetown University Press.
Kimball, J. (1973). Seven principles of surface structure parsing in natural language. *Cognition*, 2, 15–47.
King, J. & Just, M. (1991). Individual differences in syntactic processing: the role of working memory. *Journal of Memory and Language*, 30, 580-602.
Kuo, K., & Vasishth, S. (2006). Processing Chinese relative clauses: Evidence for the universal subject preference (Unpublished manuscript). University of Potsdam.
Kwon, N., Gordon, P.C., Lee, Y. & Kluender. R. (2010). Cognitive and linguistic factors affecting subject/object asymmetry: An eye-tracking study of prenominal relative clauses in Korean. *Language*, 86, 546-582.
Kwon, N., Polinsky, M. & Kluender. R. (2006). Subject preference in Korean. In D. Baumer, D. Montero, & M. Scanlon (Eds.), Proceedings of the 25th West Coast Conference on Formal Linguistics (WCCFL 25), 1-14. Somerville, MA: Cascadilla Press.

Levy, R. (2008) Expectation-based syntactic comprehension. *Cognition*, 106(3), 1126-1177.
Lewis, R., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, 29, 375-419.
Lewis, R. (1996). A theory of grammatical but unacceptable embeddings. *Journal of Psycholinguistic Research,* 25, 93–116.
Lewis, R., Vasishth, S. & Van Dyke, J. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences*, 10(10), 447-454.
Lin, Y. B., & Garnsey, S. M. (2007). Plausibility and the resolution of temporary ambiguity in relative clause comprehension in Mandarin. Poster presented at the 20th CUNY Conference on Human Sentence Processing, University of California, San Diego, CA, March 2007.
Lin, C.-J. C., & Bever, T. G. (2006). Subject preference in the processing of relative clauses in Chinese. In D. Baumer, D. Montero, & M. Scanlon (Eds.), Proceedings of the 25th West Coast conference on formal linguistics (WCCFL 25), 254-260. Sommerville, MA: Cascadilla Press.
Lowder & Gordon, P.C. (submitted). The pistol that injured the cowboy: Difficulty with Inanimate Subject-Verb Integration is Reduced by Structural Separation. University of North Carolina manuscript.
MacDonald, M., Pearlmutter, N. & Seidenberg, M. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review*, 103(4), 676-703.
Mak, W. M., Vonk, W., & Schriefers, H. (2002). The influence of animacy on relative clause processing. *Journal of Memory and Language*, 47, 50–68.
Mak, W. M., Vonk, W., & Schriefers, H. (2006). Animacy in relative clauses: The hikers that rocks crush. *Journal of Memory and Language*, 54, 466-490.
McElree, B., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory and Language*, 48, 67–91.
Mecklinger, K., Schriefers, H., Steinhauer, K., & Friederici, A. (1995). Processing relative clauses varying on syntactic and semantic dimensions. *Memory and Cognition*, 23, 477-494.
Mitchell, D. C. (1984). An evaluation of subject-paced reading tasks and other methods for investigating immediate processes in reading. In Kieras, D. and Just, M. A., editors, New methods in reading comprehension. Hillsdale, NJ: Earlbaum.
Mitchell, D., Cuetos, F., Corley, M., & Brysbaert, M. (1995). Exposure-based models of human parsing: Evidence for the use of coarse-grained (nonlexical) statistical records. Journal of Psycholinguistic Research, 24, 469–488.
Miyamoto, E., & Nakamura, M. (2003). Subject/object asymmetries in the processing of relative clauses in Japanese. In G. Garding & M. Tsujimura (Eds.), Proceedings of the 22nd West Coast conference on formal linguistics, 342-355. Somerville, MA: Cascadilla Press.
Monsell, S. (1991). The nature and locus of word frequency effects in reading. In D. Besner & G. W. Humphreys (Eds.), Basic processes in reading: Visual word recognition, 148–197. Hillsdale, NJ: Erlbaum.
Morton, J. (1969). The interaction of information in word recognition. *Psychological Review*, 76, 165–178.

Nakatani, K., & Gibson, E. (2010). An on-line study of Japanese nesting complexity. *Cognitive Science*, 34, 94-112.
O'Grady, W. (2011). Relative clauses: Processing and acquisition. In Evan Kidd (ed.), *The acquisition of relative clauses: Processing, typology and function*. Amsterdam: John Benjamins. pp. 13-38.
O'Grady, W., Lee, M. & Choo, M. (2003). Asubject-object asymmetry in the acquisition of relative clauses in Korean as a second language. *Studies in Second Language Acquisition*, 25, 433–48.
Reali, F., & Christiansen, M. H. (2007). Processing of relative clauses is made easier by frequency of occurrence. *Journal of Memory and Language, 57*, 1–23.
Roark, B., Bachrach, A., Cardenas, C., and Pallier, C. (2009). Deriving lexical and syntactic expectation-based measures for psycholinguistic modeling via incremental top-down parsing. In Proceedings of EMNLP.
Roland, D., Dick, F. & Elman, J. (2007). Frequency of basic English grammatical structures: A corpus analysis. *Journal of Memory and Language*, 57, 348–379.
Schriefers, H., Friederici, A., & Ku¨hn, K. (1995). The processing of locally ambiguous relative clauses in German. *Journal of Memory and Language*, 34, 499-520.
Smith, N. J. and Levy, R. (2008). Optimal processing times in reading: a formal model and empirical investigation. In Proceedings of the 30th Annual Meeting of the Cognitive Science Society.
Stabler, E.P. (1994). The finite connectivity of linguistic structures. In Clifton, C., Jr., Frazier, L., Rayner, K. (Eds.), Perspectives on Sentence Processing, 303–336. Erlbaum: Hillsdale, NJ.
Staub, A. (2010). Eye movements and processing difficulty in object relative clauses. *Cognition*, *116*, 71-86.
Tily, H., Fedorenko, E. & Gibson, E. (2011). A comprehensive investigation of animacy effects in relative clauses. The 24th CUNY Conference on Human Sentence Processing, Palo Alto CA, March 2011.
Tily, H., Fedorenko, E. & Gibson, E. (in revision). A comprehensive investigation of animacy effects in relative clauses.
Tily, H., Fedorenko, E. & Gibson, E. (in preparation). Self-paced reading over Mechanical Turk.
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.
Troyer, M., O'Donnell, T., Fedorenko, E. & Gibson E. (2011). Storage and computation in syntax: Evidence from relative clause priming. *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society.*
Trueswell, J., Tanenhaus, M & Garnsey, S. (1994). Semantic influences on parsing: use of thematic role information in syntactic disambiguation. *Journal of Memory and Language*, 33, 285-318.
Ueno, M., & Garnsey, S. (2008). An ERP study of the processing of subject and object relative clauses in Japanese. *Language and Cognitive Processes*, 23, 646-688.

Van Dyke, J. & McElree, B. (2006). Retrieval interference in sentence comprehension. *Journal of Memory and Language*, 55, 157–166.
Van Dyke, J., Lewis, R. (2003). Distinguishing effects of structure and decay on attachment and repair: A cue-based parsing account of recovery from misanalyzed ambiguities. *Journal of Memory and Language*, 49, 285-316.
Vasishth, S., & Lewis, R. (2006). Argument-head distance and processing complexity: Explaining both locality and anti-locality effects. *Language*, 82(4), 767-794.
Vasishth, S., Suckow, K., Lewis, R.L. & Kern, S. (2010). Short-term forgetting in sentence comprehension: Crosslinguistic evidence from head-final structures. *Language and Cognitive Processes*, 25(4), 533-567.
Vasishth, S. & Drenhaus, H. (2011). Locality in German. *Dialogue and Discourse*, 1, 59-82.
Wanner, E., & Maratsos, M. (1978). An ATN approach to comprehension. In M. Halle, J. Bresnan, & G. Miller (Eds.), Linguistic Theory and psychological reality (pp. 119–161). Cambridge, MA: MIT Press.
Warren, T. & Gibson, E. (2002). The influence of referential processing on sentence complexity. *Cognition, 85,* 79-112.
Yngve, V.H. (1960). A model and an hypothesis for language structure. *Proceedings of the American Philosophical Society*, 104, 444–466.