COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science (2012) 1–27

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/j.1551-6709.2011.01217.x

Processing Relative Clauses in Supportive Contexts

Evelina Fedorenko,* Steve Piantadosi, Edward Gibson*

Department of Brain and Cognitive Sciences, MIT

Received 4 July 2008; received in revised form 21 June 2011; accepted 27 June 2011

Abstract

Results from two self-paced reading experiments in English are reported in which subject- and object-extracted relative clauses (SRCs and ORCs, respectively) were presented in contexts that support both types of relative clauses (RCs). Object-extracted versions were read more slowly than subject-extracted versions across both experiments. These results are not consistent with a decay-based working memory account of dependency formation where the amount of decay is a function of the number of new discourse referents that intervene between the dependents (Gibson, 1998; Warren & Gibson, 2002). Rather, these results support interference-based accounts and decay-based accounts where the amount of decay depends on the number of words or on the type of noun phrases that intervene between the dependents. In Experiment 2, presentation in supportive contexts was directly contrasted with presentation in null contexts. Whereas in the null context the extraction effect was only observed during the RC region, in a supportive context the extraction effect was numerically larger and persisted into the following region, thus showing that extraction effects are *enhanced* in supportive contexts. A sentence completion study demonstrated that the rate of SRCs versus ORCs was similar across null and supportive contexts (with most completions being subject-extractions), ruling out the possibility that an enhanced extraction effect in supportive contexts is due to ORCs being less expected in such contexts. However, the content of the RCs differed between contexts in the completions, such that the RCs produced in supportive contexts were more constrained, reflecting the lexical and semantic content of the preceding context. This effect, which we discuss in terms of expectations/lexico-syntactic priming, suggests that the enhancement of the extraction effect in supportive contexts is due to the facilitation of the subject-extracted condition.

Keywords: Sentence processing; Working memory; Supportive context; Relative clauses; Predictability

Correspondence should be sent to Evelina Fedorenko, MIT 46-3037, Cambridge, MA 02139; Edward Gibson, MIT 46-3035, Cambridge, MA 02139. E-mails: evelina9@mit.edu, egibson@mit.edu
*These authors contributed equally to this work.

1. Introduction

On-line sentence comprehension involves the rapid use of many information sources (Gibson & Pearlmutter, 1998; Tanenhaus & Trueswell, 1995), including (a) the syntactic rules of the language; (b) the syntactic/semantic properties of words; (c) the real-world plausibility of the described events; (d) the local discourse context; and (e) for spoken language, the prosody of the utterance. The interpretation process is further constrained by the amount of working memory resources available to the comprehender. A useful construction for investigating constraints on sentence processing is the *relative clause* (RC), a clause that modifies a noun and usually serves to pick a relevant element from a set. It has long been observed that in English, certain object-extracted relative clauses (ORCs) such as (1a) are more difficult to process than subject-extracted relative clauses (SRCs) such as (1b), in a null context:

(1) a. The reporter that the senator attacked admitted to making an error.
b. The reporter that attacked the senator admitted to making an error.

In (1a), the RC pronoun ''that,'' which is coindexed with the subject noun phrase (NP) ''the reporter,'' is interpreted as the object of the verb ''attacked.'' In (1b), the RC pronoun ''that'' is interpreted as the subject of the verb ''attacked.'' When both the main clause subject (''the reporter'') and the embedded clause NP (''the senator'') are animate and each NP is plausible as both the subject and the object of the embedded verb, people read ORCs more slowly than SRCs and perform less well on memory tasks and comprehension questions after an ORC (e.g., Fedorenko, Gibson, & Rohde, 2006; Ford, 1983; Gibson, 1998; Gordon, Hendrick, & Johnson, 2001; Gordon, Hendrick, & Levine, 2002; Just & Carpenter, 1992; King & Just, 1991; Wanner & Maratsos, 1978). Some of this difference in processing difficulty may be due to differences in working memory demands associated with the processing of the two structures. In particular, it has been argued that ORCs cause more processing difficulty than SRCs because ORCs contain a non-local syntactic dependency, between the matrix NP and the embedded verb (i.e., ''the reporter'' and ''attacked'' in [1a]). This is in contrast to SRCs, in which both of the dependencies in the RC—between the matrix subject and the verb (i.e., ''the reporter'' and ''attacked'' in [1b]), and between the verb and the embedded object (i.e., ''attacked'' and ''the senator'')—are local. Non-local dependencies have been argued to increase working memory demands because the first element of the dependency (i.e., ''the reporter'' in [1a]) needs to be retrieved from memory at the point when the second element of the dependency (i.e., ''attacked'' in [1a]) is encountered (e.g., Gibson, 1998, 2000; Gordon et al., 2001; Just & Carpenter, 1992; Lewis & Vasishth, 2005).

An alternative class of accounts that attempt to explain the difference in processing difficulty between ORCs and SRCs has to do with syntactic surprisal (Gennari & MacDonald, 2008; Hale, 2001; Levy, 2008; see Traxler, Morris, & Seely, 2002, for a related proposal). According to these accounts, because the ORC structure is less frequent than the SRC

structure (Roland, Dick, & Elman, 2007), an SRC is more expected than an ORC following the word ''that.'' Thus, when the embedded subject NP ''the senator'' is encountered, there may be some difficulty in processing the less expected continuation. Whereas syntactic surprisal can explain potential differences in processing difficulty early in the RCs, surprisal-based accounts cannot explain processing difficulty at the embedded verb in the ORC structures (as e.g., Levy, 2008, acknowledges), which is where most of the processing difficulty is typically observed in reading paradigms such as self-paced reading (e.g., Gordon et al., 2001; Grodner & Gibson, 2005; King & Just, 1991) and eye-tracking (e.g., Gordon, Hendrick, Johnson, & Lee, 2006; Holmes & O'Regan, 1981; Traxler et al., 2002; cf. Staub, 2010, for some evidence of difficulty at the embedded subject in ORCs in regressive saccades). A verb is the most expected continuation following the embedded subject in an ORC structure, which leads the surprisal-based accounts to predict low complexity there, whereas high complexity is observed. Furthermore, Grodner and Gibson (2005) show that the reading time slowdown occurs at the verb even when the ORC subject NP is modified by a prepositional phrase (e.g., ''the nurse at the clinic''). A surprisal-based account would predict difficulty at the embedded subject, or at least soon after it, during the modifier. Such a slowdown does not typically occur (cf. Staub, 2010), contrary to the predictions of surprisal-based accounts. As a result, whereas syntactic surprisal may be an important factor in sentence processing in general, the best accounts of the difference in processing difficulty between SRCs and ORCs in English (and, in particular, of the locus of the greatest processing difficulty at the embedded verb in ORC structures) are working-memory-based accounts (see also results on processing RCs from other languages, e.g., Dutch [Mak, Vonk, & Schriefers, 2002, 2006], German [Mecklinger, Schriefers, Steinhauer, & Friederici, 1995; Schriefers, Friederici, & Kühn, 1995], Korean [Kwon, Gordon, Lee, & Kluender, 2010], Japanese [Ishizuka, Nakatani, & Gibson, 2003; Miyamoto & Nakamura, 2003; Ueno & Garnsey, 2008], and Chinese [Gibson & Wu, in press]). We will therefore focus on working-memory-based accounts in the discussion of our experiments.

Although most working memory accounts are based on the idea that retrieving an element further back in the input stream is more difficult than retrieving a more local element, it is not yet clear how to best quantify retrieval difficulty. Two non-mutually exclusive classes of retrieval metrics have been proposed. According to the first, the difficulty of retrieving the first element of the dependency at the time of processing the second element of the dependency is due to the decay in the activation of the representation of the first element as additional elements are being processed. Under such a decay-based framework, Gibson (1998, 2000) and Warren and Gibson (2002) have proposed that retrieval difficulty might depend on the number of new discourse referents (nouns and verbs) intervening between the two elements of a long-distance dependency. Warren and Gibson also consider a decay-based metric in terms of the surface type of NPs that intervene (e.g., pronouns vs. names vs. definite descriptions),[1] and Gibson (1998) considers a decay-based metric in terms of the number of intervening words (cf. Hawkins, 1994, for a word-based production difficulty metric).

An important variant of the activation decay idea is proposed by Lewis and Vasishth (2005), Vasishth and Lewis (2006), and Lewis, Vasishth, and Van Dyke (2006). According to Lewis and colleagues, retrieval difficulty is determined not by some function of the linear

distance since the item was first activated, but by the history of the item's recent retrievals. In particular, if the item has been re-activated since it was first accessed, then its activation can increase.

According to the second class of proposals, retrieval difficulty is due to interference of the elements intervening between the two ends of a dependency with the representation of the first, to-be-retrieved, element (Gordon et al., 2001; Gordon, Hendrick, & Johnson, 2004; McElree, Foraker, & Dyer 2003; Lewis & Vasishth, 2005; Lewis et al., 2006).

Both the decay- and interference-based proposals successfully predict processing difficulty in ORCs, compared to SRCs. According to the decay-based proposals, processing more new discourse referents/words following the matrix NP ''the reporter'' in an ORC leads to the decay of this NP, making its retrieval more difficult at the verb ''attacked.'' According to the interference-based proposals, the fact that the intervening NP ''the senator'' is similar both syntactically and semantically to the matrix NP ''the reporter'' makes retrieval of ''the reporter'' more difficult at the verb ''attacked.''

In most previous work investigating the processing of SRCs and ORCs, the critical sentences were presented in a null context (cf. Battinich & Levine, 2009). However, because (a) the primary function of RCs is to pick out the relevant referent from a set, and (b) some accounts of RC complexity invoke the notion of discourse context (Gibson, 1998; Warren & Gibson, 2002), it is important to investigate the processing of RCs in contexts that license them. The current experiments therefore examine the effect of supportive contexts on the processing of SRCs and ORCs.

There is much evidence from the sentence processing literature that the context (linguistic, visual, social, etc.) in which an utterance is produced strongly affects processing and interpretation. To summarize some key findings:

1. Words are processed faster in context (e.g., Frisson, Rayner, & Pickering, 2005; Schustack, Ehrlich, & Rayner, 1987); relatedly, in studies using event-related potentials, the N400 component, which is observed in response to each content word and assumed to reflect the relative ease of integrating the word into the evolving meaning representation (e.g., Holcomb, 1993), decreases for words occurring at later sentence positions (Van Petten & Kutas, 1990).
2. Syntactically ambiguous materials that cause processing difficulty in a null context cause much less difficulty or even no difficulty in a supportive context (Altmann & Steedman, 1988; Crain & Steedman, 1985; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995).
3. Referential processing is affected by the local visual context, such that the presence of particular objects in the context affects people's expectations of what elements will be referred to in a discourse (Altmann & Kamide, 1999; Kamide, Altmann, & Haywood, 2003).
4. The processing of RCs is affected by the context, such that restrictive RCs are processed more quickly than non-restrictive RCs in a supportive context, whereas they are processed more slowly than non-restrictive RCs in a null context (Grodner, Gibson, & Watson, 2005).

Understanding how context modulates RC complexity effects can help distinguish between some of the accounts summarized above. One account is particularly relevant to the examination of RCs in supportive contexts: A discourse-referent-based decay account, whereby the difficulty of retrieving the first element of a dependency at the time of processing the second element of the dependency is due to the decay in the activation of the representation of the first element as additional new discourse referents are being processed in the interim (Gibson, 1998, 2000; Warren & Gibson, 2002). This account predicts that the ORC versus SRC complexity effect should be eliminated in a supportive context. For example, consider the context in (2),[2] to be followed by the ORC in (1a) or the SRC in (1b) above:

(2) At the press-conference, a senator and two reporters got into an argument.

The senator attacked one of the reporters and then the other reporter attacked the senator.

The context in (2) establishes the existence of two reporters, one who attacked a senator, and the other who was attacked by the same senator. When the definite NP ''the reporter'' is then processed at the beginning of (1a) or (1b), a modifier is expected in order to distinguish between the two reporters present in the discourse context. An RC is therefore a likely continuation at this point: An ORC like the one in (1a) picks out one reporter, and an SRC like the one in (1b) picks out the other.

According to the discourse-referent-based decay account of retrieval difficulty in sentence processing, no difficulty should be observed for the ORC in (1a) because the referent for the intervening NP ''the senator'' is not new; it has already been introduced to the discourse and is therefore highly activated. Thus, this account predicts no difference between the SRCs and ORCs. In contrast, other working-memory-based accounts of RC complexity predict that the ORC complexity effect should remain in supportive contexts. For example, according to interference-based accounts (e.g., Gordon et al., 2001, 2004; Lewis et al., 2006), processing the verb ''attacked'' should be difficult in the ORC in (1a), independent of the context, because the presence of the intervening NP ''the senator'' interferes with the retrieval of the object NP ''the reporter'' whether or not this NP refers to a new entity in the context or to an entity that has already been introduced into the context. Similarly, the word-based decay account and the NP-type-based decay account predict the same ORC difficulty in a supportive context as in a null context.

2. Experiment 1

Experiment 1 was designed to test whether the local discourse context would eliminate the object- versus subject-extraction complexity effect, as predicted by a discourse-referent-based decay account of retrieval difficulty. The contexts were like (2) above, consisting of two sentences that set up a scenario in which either an SRC or an ORC is appropriate. The two critical sentences were presented in the form of a dialog between two speakers (Mary

and John). The first speaker would produce an utterance involving an SRC and the second speaker would produce an utterance involving an ORC, or vice versa, as shown in (3), with slash-marks (/) indicating presentation regions and double slash-marks (//) indicating line breaks (across which there was always a new region).

(3) At the press-conference, a senator and two reporters got into an argument.//

The senator attacked one of the reporters and then the other reporter attacked the senator.//

Mary: / I heard that / the reporter / that {attacked the senator | the senator attacked} / admitted to / making an error.//

John: / I'm not sure about that. / I heard that / the reporter / that {the senator attacked | attacked the senator} / admitted to / making an error.//

We designed the materials in a way that would also allow us to investigate potential effects on the RC extraction effect of the presence of an RC in the preceding context. In particular, we can see whether the processing of the RC in John's utterance is affected by the RC in Mary's utterance. If the position of an RC (early vs. late) does not affect its processing, then an item could serve as its own control, because each trial includes both an ORC and an SRC. If, however, the position of an RC does matter, then we can learn something about how the presence and/or kind of an earlier-encountered RC affects subsequent RC processing.

The predictions for this experiment are as follows:

1. If retrieval difficulty is determined by discourse-referent-based decay of earlier syntactic positions, then reading times during the ORCs should be no longer than during the SRCs for either the early or the late RC positions, because all the relevant referents corresponding to the NPs in the RCs are present in the context and should therefore be easily accessible. In particular, processing the intervening NP in an ORC should cause little or no decay of the head NP, because there is no work involved in constructing a discourse referent for this intervening NP.
2. In contrast, if retrieval difficulty is determined by interference, such that similarity between the first (to-be-retrieved) element of the dependency and the intervening element(s) gives rise to processing difficulty, then reading times during the ORCs should be longer than during the SRCs. Similarly, if retrieval difficulty is determined by word-based decay or NP-type-based decay, then ORCs should be processed more slowly than SRCs.
3. The position of the RC within an item (early vs. late) may also affect RC processing. Some kind of a modifier is quite likely to follow the definite head noun ''the reporter'' in both Mary's and John's statements in (3), because there are two reporters in the context: ''the reporter'' by itself is not felicitous. A natural way to distinguish the two potential referents is by using an RC in each case. But the form of the RC is unknown following Mary's statement in (3): It could be an active subject-extraction,

an active object-extraction, a passive subject-extraction, or a passive object-extraction. In contrast, at the point of encountering the definite NP ''the reporter'' in John's response to Mary in (3), an RC may be more expected due to re-use of structures in dialog (e.g., Branigan, Pickering, McLean, & Cleland, 2007; Pickering & Branigan, 1999), and furthermore a particular kind of an RC may be expected to contrast with the earlier occurring RC. (Even if participants learn—over the course of the experiment—that passive RCs are never used, the type of RC in John's utterance is still more constrained than the type of RC in Mary's utterance given that the intended referent is now clear.) It is thus possible that any ORC versus SRC effect that might occur in the early position (as predicted by the interference-based and word-/NP-type-based decay retrieval theories) would be diminished or eliminated in the later position because of the higher predictability of the relevant RC in the later position based on the previously encountered RC.

2.1. Methods

2.1.1. Design and materials

Twenty-four items were constructed following the format of (3), with two conditions for each item: subject-extracted-first versus object-extracted-first. There were two RCs in each item: one early (in Mary's utterance) and one late (in John's utterance). The following factors were counter-balanced:

1. The structure of the second context sentence varied with regard to whether the description of the events started with the unique referent or one of the two referents of the same kind: half of the items started with the unique noun (*The NOUN1 VERBed one of the NON-UNIQUE NOUNS and then the other NON-UNIQUE NOUN VERBed the NOUN1*), and the other half started with a non-unique noun (*One of the NON-UNIQUE NOUNS VERBed the NOUN1 and then the NOUN1 VERBed the other NON-UNIQUE NOUN*).
2. Each item had two versions such that the identity of the unique noun changed. For example, there was another version of the item in (3), which started as follows: *At the press-conference, a reporter and two senators got into an argument.* This ensures that any differences that we observe are not due to plausibility differences between the meanings of the subject-extracted structure and the object-extracted structure.

In addition to the target materials, the experiment included 48 fillers. The fillers were carefully constructed to resemble the target materials. Specifically, the first context sentence always introduced three entities (all were unique nouns: e.g., *the model, the photographer, and the hairdresser*). The second sentence contained two predicates that described events happening to the three participants (e.g., the first character doing something to the second character and the second character doing something to the third character; the order in which the nouns from the first sentence were used in these events was varied across the filler items). The two clauses of the second sentence were connected by a variety of connectives,

such as *because, however, when, while, after*, etc. Mary's utterance started with *I heard that,* like in the target sentences, and continued with a statement about one of the three characters (16 fillers—1st character, 16 fillers—2nd character, and 16 fillers—3rd character). John's utterance started with a contradiction in 12/48 fillers: This was similar to the target sentences, but a variety of different expressions preceded ''*I heard that,*'' such as ''*Hmm, that's weird*'' or ''*I am not sure.*'' In the remaining 36/48 fillers John's utterance started with an expression of agreement, such as ''*Yeah, that's right*'' or ''*I think you're right.*'' This way, overall, half of the materials in the experiment involved a disagreement between Mary and John, and the other half—agreement. Critically, Mary's and John's utterances did not involve RCs. A sample filler is presented in (4):

(4) During the photo-shoot, the model, the photographer and the hairdresser discussed possible hairstyles.

The model liked her hair straight; however, the photographer and the hairdresser wanted her hair to be curly.

Mary: I heard that the model is stubborn and hard to work with.

John: Yeah, that's right, and I also heard that she gets paid ten thousand dollars for every photo-shoot.

See Appendix for a complete list of experimental materials (filler materials are available from the authors upon request).

2.1.2. Participants

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

2.1.3. Procedure

The task was self-paced reading, using a moving window display (Just, Carpenter, & Woolley, 1982). LINGER 1.7 by Doug Rohde was the software used to run the experiment. Each trial began with a series of dashes marking the position and length of the words in the sentences, across several lines of text. Participants pressed the spacebar to reveal each fragment of the materials. The amount of time the participant spent reading each fragment was recorded as the time between key-presses.

The first two sentences of each text were presented sentence by sentence. The final two sentences of each text (Mary's utterance and John's response, see [3]) were presented region by region, with the presentation regions indicated in (3). The fillers were broken down into regions in a similar fashion. Most previous experiments from the literature that evaluated RC complexity in self-paced reading in null contexts used a word-by-word presentation. But the current materials involve several sentences per trial, which makes

word-by-word presentation somewhat bothersome/tiring for participants. We therefore chose to use a phrase-by-phrase presentation in this experiment.

The questions for the target materials were about the content of the utterances. Specifically, they either asked about the RCs and were of the form *According to Mary, was it the reporter that {attacked the senator/the senator attacked} that admitted to making an error?*, or they asked about the main verb and were of the form *Did one of the {reporters / senators} admit to making an error?* Half of the questions for the filler materials were about the information contained in the context sentences, and the other half was about the information contained in the utterances.

Participants took approximately 35 min to complete the experiment.

2.2. Results

2.2.1. Analyses

Analyses reported here were conducted with the lme4 package (Bates et al., 2008) for the statistical language R (R Core Development Team, 2008). Recent results have shown that including only random intercepts in linear mixed effects regressions can be anti-conservative, so we also include random slopes for participants in our model. Significance (p) values were estimated from (a) the t-values that were obtained from the lmer function; and (b) conservative estimates of the number of degrees of freedom in the model. The estimates of the number of degrees of freedom in the model consisted of the number of observations minus the number of intercepts fit in the model (the number of participants + the number of items = 32 + 24 = 56) and the number of slopes being fit in the model (the number of participants = 32).

2.2.2. Comprehension question performance

Participants answered the comprehension question correctly 75.7% of the time for the target items. A logistic regression showed that the order of the RC type (SRC occurring in the early position vs. ORC occurring in the early position) did not affect accuracy (SRC-first: 76.2%; ORC-first: 74.9%; $\beta = .069$, $z = .404$, $p = .686$). The somewhat low accuracy on the questions relative to accuracy rates in experiments in the literature on RCs in null contexts (where accuracies are usually above 80%) was likely due to the complexity of the questions that were asked of the participants in the current experiment. In particular, participants were asked not only who did what to whom in the contexts but also which of the two dialog partners (John or Mary) said so. In the filler items, where the comprehension questions were not quite as complex as in the targets, participants achieved a higher accuracy of 85.4% on average across the 48 items, thus supporting this interpretation of the relatively low accuracy on the target items.

2.2.3. Reaction times

Before performing linear mixed-effects analyses on our reading time data, we transformed the raw reading times into residual reading times (Ferreira & Clifton, 1986; see Trueswell, Tanenhaus, & Garnsey, 1994, for discussion). First, a regression equation

predicting reaction times from region length was derived for each participant, using all filler and target items. Then, for each region, the reaction time predicted by the participant's regression equation was subtracted from the actual measured reaction time to obtain a residual reaction time. We analyzed residual reading times because (a) they adjust for differences in region lengths and overall differences in participants' reading rates; and (b) they are more normally distributed than raw reading times (normally distributed data are a requirement for regression analyses, e.g., Jaeger, 2008). Raw reading times > 8 s and residual reading times more than three standard deviations away from the mean for a position within condition were removed from the analyses, excluding 2.1% of the data.

We defined the critical regions as the two occurrences of the RCs (in Mary's utterance [early occurrence], and in John's utterance [late occurrence]), and the two main verb regions following the RCs, because RC complexity effects are often observed on the region immediately following the RC. Fig. 1 presents the mean raw reading times per region across the two orders of the RC type.[3]

We first wanted to evaluate the potential effects of comprehension accuracy and its potential interactions with the critical factors during the critical region, the RC. A linear mixed-effects model including RC type (SRC, ORC), position (early, late) and comprehension question accuracy (correct, incorrect) as sum-coded factors, as well as random slopes and intercepts for participants and random intercepts for items revealed no effect of comprehension question accuracy ($t = -1.15$), and no interactions between comprehension question accuracy and other factors (ts < 0.86). Similarly, there were no reliable effects or interactions with comprehension question accuracy on other regions. Consequently, we do not include comprehension question accuracy in further analyses.

Analyses without this factor revealed a main effect of RC type on the RC region ($\beta = 134.0$, $t = 3.82$, $p < .001$), such that SRCs were processed 268 ms faster than ORCs on

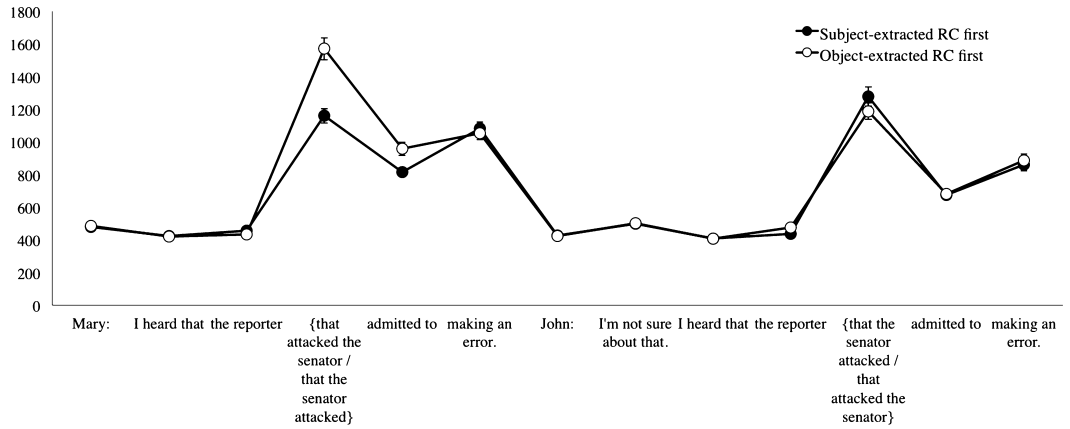

Fig. 1. Mean raw reading times per region for Experiment 1. Error bars indicate standard errors of the mean. Note that the line with the open circles corresponds to the object-extracted relative clause (RC) in the early region, but to the subject-extracted RC in the late region.

average. There was also a marginal main effect of position (β = 62.6, t = 1.80, $p < .08$), such that RCs in the later region were processed 125.2 ms faster than in the early region, and an interaction between RC type and position (β = 77.5, t = 2.20, $p < .05$), such that the RC effect is diminished in the later position.[4]

We also conducted pairwise analyses according to RC type for each region. We first report the results of these analyses for the target regions, regions 6 and 7 (early RC) and regions 13 and 14 (late RC). In regions 6 and 7—the first RC region and the following main verb region—there was an effect of RC type (RC region: β = 423.5, t = 3.81, $p < .001$; main-verb region: β = 144.2, t = 2.64, $p < .05$), such that SRCs were processed 423.5/144.2 ms faster than ORCs in the respective RC/main-verb regions. In regions 13 and 14, the second instance of the RC and the following main verb region, the RTs were not significantly different (ts < 1.5). For completeness, we analyzed the other 11 regions. There were no reliable differences according to condition.

2.3. Discussion

The results of this experiment demonstrated a robust extraction effect such that ORCs were read more slowly than SRCs in supportive contexts. This result is not consistent with the discourse-referent-based decay account of object-extraction difficulty, which predicts that there should be no extraction effect in supportive contexts. Rather, this result is consistent with interference-based accounts and word- and NP-type-based decay accounts of retrieval difficulty. The experiment also provided suggestive evidence that the position of the RC in the context (early vs. late) affects its processing. In particular, the extraction effect appears to be stronger in the early position than in the late position, plausibly because the type of RC was less predictable in the early position than in the late position, as discussed above.

It is interesting to compare the RC complexity effects in the current experiment to the RC complexity effects that have been previously observed in self-paced reading experiments in null contexts. In particular, we will compare the current results to the results from Gibson, Desmet, Grodner, Watson, and Ko (2005) and from Grodner and Gibson (2005), both of which were run in the same lab as the experiment reported here, using the same self-paced reading presentation software, similar materials, and similar data-trimming procedures. There are two potentially important differences between these earlier experiments and the current experiment: (a) the earlier experiments presented RC materials in null contexts, and (b) the presentation was word by word, not phrase by phrase. Gibson et al. (2005) reported an effect of about 100 ms total during the RC region (approximately 35 ms per word over the three words in the RC), such that ORCs were read more slowly than SRCs. This effect was robust across four different comparisons. Grodner and Gibson (2005) report a similar effect size for a comparison between ORCs and SRCs: ORCs were read approximately 120–150 ms more slowly than SRCs. In the current experiment, we observed a 325 ms effect during the RC region and a further 156 ms effect in the following region. Thus, it appears that the presence of a supportive context is *enhancing* the complexity difference between SRCs and ORCs (which, interestingly, is the opposite of what we had predicted based on

the literature on processing sentences in context). We investigate this possibility further in Experiment 2.

3. Experiment 2

In Experiment 1, we observed a larger difference between ORCs and SRCs in supportive contexts than the differences typically observed in the literature for similar comparisons in null contexts. The larger difference could be due to the presence of a supportive context. However, it could also be a result of different experimental procedures used in Experiment 1 compared to those used in previous studies. In particular, the self-paced reading experiments from the literature that have investigated SRCs and ORCs in null contexts have usually used a word-by-word presentation, whereas in Experiment 1 we used a region-by-region presentation, and RCs were presented as single regions. Region-by-region presentation in self-paced reading might lead to larger complexity effects than word-by-word presentation because it is possible that people perform more regressions when they have multiple words available to the left of the word that they are currently fixating, compared with the word-by-word presentation when regressions are less likely. Experiment 2 was conducted in order to directly investigate potential differences between processing RCs in null and supportive contexts using self-paced reading. In order to do this, we used the materials from Experiment 1 for the supportive context conditions, and we removed the context from each of these items to create the null context conditions. Furthermore, because the strongest extraction effects in Experiment 1 occurred during the first RC region, we omitted the second RC region in the items for Experiment 2. An item for Experiment 2 therefore had the form of (5):

(5) Supportive context (not present in the null context versions of the items):

At the press-conference, a senator and two reporters got into an argument.//

The senator attacked one of the reporters and then the other reporter attacked the senator.//

Target sentence:

Mary: / I heard that / the reporter / that {attacked the senator | the senator attacked} / admitted to / making an error.//

John: / I'm not sure about that.//

A discourse-referent-based decay account of retrieval difficulty predicts a contrast in the RC complexity effect across the two contexts. In particular, the RC complexity effect should be strong in the null context, because a new discourse referent is introduced in subject position of the RC (''the senator'' in [5]), between the pronoun *that* and the verb of the RC.

This effect should be weakened or eliminated in the supportive context, because this discourse referent has been introduced in the context immediately preceding the RC. Thus, a discourse-referent-based decay account predicts an interaction between context and RC type, such that the RC complexity effect should be weaker/non-existent in supportive contexts.

In contrast, other working memory accounts of RC complexity predict that the ORCs should be more complex than SRCs independent of the context, supportive or null.[5] Finally, if the results of Experiment 1 are directly comparable with the results of previous experiments where RCs were presented in null contexts, then the RC complexity effect may actually be enhanced in supportive contexts.

3.1. Methods

3.1.1. Design and materials

The experiment had a 2×2 design, crossing the presence of context (context, no context), and extraction (SRC, ORC). The 24 items from Experiment 1 were modified to form the items for Experiment 2, as follows. First, the second utterance (from John) was shortened to ''I'm not sure about that,'' omitting the second sentence. Second, the null context conditions were created by omitting the supportive contexts, so that the items in the null context condition started with Mary's utterance.

Twenty-four of the 48 filler items from Experiment 1 were used without any changes for the current experiment. The other 24 filler items were modified for the current experiment by omitting the context before Mary and John's dialog.

3.1.2. Participants

Thirty-two participants from MIT and the surrounding community were paid for their participation. All were native speakers of English and were naive as to the purposes of the study. None had participated in Experiment 1.

3.1.3. Procedure

We used a phrase-by-phrase presentation in this experiment, as in Experiment 1. Participants took approximately 25 min to complete the experiment.

3.2. Results

3.2.1. Comprehension question performance

Participants answered the comprehension question correctly 73.7% of the time for the target items. Although there were numerical trends for differences between the conditions (SRC, context: 79.1%; ORC, context: 72.9%; SRC, no context: 73.3%; ORC, no context: 69.3%), a logistic regression showed that neither RC type (SRC, ORC; $\beta = 0.135$, $z = 1.61$, $p = .107$) nor the presence of a supportive context (context, no context; $\beta = .125$, $z = 1.49$, $p = .135$) reliably affected accuracy, nor was there an interaction ($\beta = -.038$, $z = -.456$, $p = .648$). As in Experiment 1, the somewhat low accuracy rate overall is probably due to

the complexity of the questions. In the filler items, where the comprehension questions were not as complex as in the targets, participants achieved a higher accuracy of 83.5% on average across the 48 items.

3.2.2. Reaction times

As in Experiment 1, raw reading times > 8 s and residual reading times more than three standard deviations away from the mean for a position within condition were removed from the analyses, excluding 2.0% of the data.

As in Experiment 1, the critical regions were the RC region and the main verb region immediately following the RC. Fig. 2 presents the mean raw reading times per region across the four conditions.[6]

We first wanted to evaluate the potential effects of comprehension accuracy and its potential interactions with the critical factors during the critical region. A linear mixed effects model including RC type (SRC, ORC), context (context, no context), and comprehension question accuracy (correct, incorrect) as sum-coded factors, as well as random slopes and intercepts for subjects and random intercepts for items revealed no effects (main effects or interactions) of comprehension question accuracy (ts < 1.8). Similarly, there were no reliable effects or interactions with comprehension question accuracy on other regions. Consequently, we do not include comprehension question accuracy in further analyses. Analyses without this factor on the RC region revealed two main effects with no interaction: (a) a main effect of RC type (β = 106.6, t = 3.22, p < .005), such that SRCs were processed 213 ms faster than ORCs; (b) a main effect of context (β = 159.0, t = 4.01, p < .001), such that the context conditions were processed 318.0 ms faster than the no-context conditions; but no interaction between RC type and context (β = −10.0, t = −0.35). Analyses on the main-verb region following the RC revealed (a) a main effect of RC type (β = 54.1, t = 2.81, p < .01), such that SRCs were processed 108 ms faster than ORCs; (b) a main effect of context (β = 86.5, t = 3.32, p < .005), such that the context conditions were

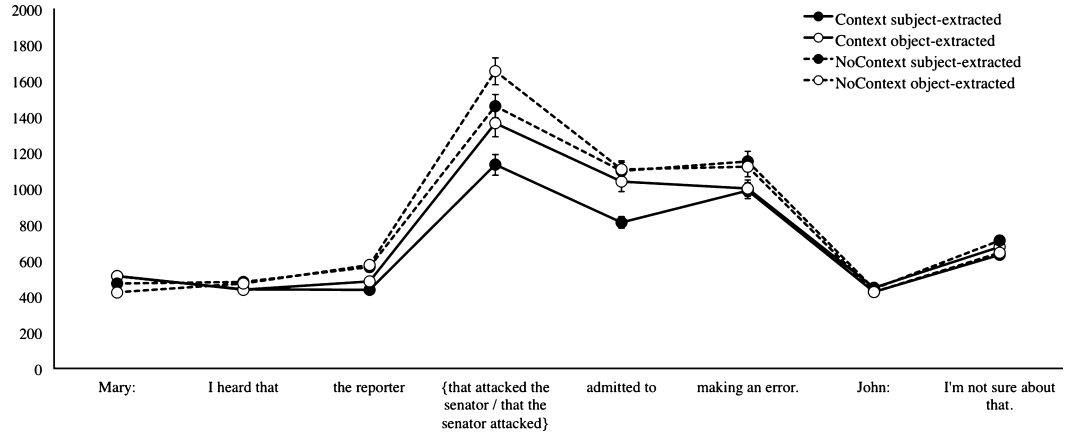

Fig. 2. Mean raw reading times per region for Experiment 2. Error bars indicate standard errors of the mean.

processed 173 ms faster than the no-context conditions; and (c) an interaction between RC type and context ($\beta = -49.3$, $t = -2.15$, $p < .05$), such that the RC extraction effect is reduced in the null context. Further analysis of this region shows that there was no RC effect for null context conditions ($t = 0.19$), but a reliable effect of RC for supportive context conditions ($\beta = 104.6$, $t = 3.33$, $p < .005$).

For completeness, we analyzed the other eight regions. These analyses revealed three reliable effects: (a) a reliable effect of context in region 3, the first word in the no-context conditions, such that the no-context conditions were read more quickly than the context conditions ($\beta = -31.9$, $t = -4.85$, $p < .001$); (b) a reliable effect of context in region 4, such that the context conditions were processed faster than the no-context conditions ($\beta = 20.0$, $t = 3.01$, $p < .005$); and (c) a reliable effect of context in region 5, such that the context conditions were processed faster than the no-context conditions ($\beta = 59.1$, $t = 5.23$, $p < .001$). There were no other significant differences in these analyses.

3.3. Discussion

The results of this experiment replicated the large ORC complexity effect in supportive contexts that was observed in Experiment 1. In addition, the current experiment demonstrates that the RC complexity effect is larger in supportive contexts in some sentence regions than in null contexts. In particular, during the RC region there was an RC complexity effect for both supportive and null context conditions, although it was numerically larger for the supportive context conditions (221 ms vs. 148 ms). However, during the following region, there was a large RC complexity effect in the supportive context conditions, but no such effect in the null context conditions, leading to the observed interaction. Thus, it appears that there is a longer lasting processing cost for ORCs in the supportive context relative to the null context.

Because during the RC region the effect sizes were comparable for the supportive and null-context conditions, it is plausible that smaller effect sizes for the extraction effect in the null-context experiments discussed above (Gibson et al., 2005; Grodner & Gibson, 2005) are due to differences in the presentation procedure—that is, region-by-region versus word-by-word presentation—as discussed above. Nevertheless, taking into account the region following the RC, the extraction effect is still enhanced in supportive compared to null contexts.

These results support interference-based and word- and NP-type-based decay working memory accounts of RC complexity, and they are not consistent with a decay-based account where decay is a function of constructing new discourse referents for elements that intervene between the relevant syntactic dependents. Such an account predicts that RC complexity effects should be eliminated (or, at least, greatly reduced) in supportive contexts compared with null contexts, but the reverse is true.

One point is worth making with respect to the NP-type-based decay account. Although the current experiments do not directly evaluate the predictions of this account (as they do not manipulate the type of the embedded NP), they nevertheless are informative with respect to its conceptual basis. In particular, as discussed in footnote 1, the NP-type-based decay

account is conceptually related to the discourse-referent-based decay account. In discussing the NP-type-based account, Warren and Gibson (2002) discuss the relative difficulty of constructing referents corresponding to different types of NPs, with this process being the easiest for pronouns (especially first- and second-person pronouns because those are always implicitly present in the discourse), harder for names and definite descriptions, and hardest for indefinite descriptions, according to an accessibility hierarchy (Ariel, 1988, 1990; Gundel, Hedberg, & Zacharski, 1993). The current experiments demonstrate that the difficulty in ORCs is not due to the difficulty of constructing a new discourse referent for the element intervening between the relevant syntactic dependents. As a result, although the effects of NP type in the embedded subject position of ORCs are plausibly real, it is unlikely that they have anything to do with the difficulty of constructing discourse referents. Instead, they are probably due to the relative frequencies of different types of NPs in the embedded subject position (e.g., Reali & Christiansen, 2007) and/or to the lexical or phonological properties of the relevant NPs (frequency, length, etc.)

4. General discussion

Across two experiments we have observed a robust ORC processing complexity effect in supportive contexts. These results provide evidence against a decay-based working memory account of RC complexity, where decay is a function of constructing new discourse referents for elements that intervene between the relevant syntactic dependents. In contrast, these results are consistent with interference-based accounts and word- or NP-type-based decay accounts of RC complexity (although they do constrain the range of interpretations of NP-type effects in ORCs, as discussed above).

Moreover, somewhat surprisingly, we observed a larger ORC complexity effect in supportive contexts than in null contexts in Experiment 2. This result is not straightforwardly predicted by any of the working-memory-based accounts discussed here. It is therefore interesting to speculate on why supportive contexts might enhance the RC complexity effect. We consider several possibilities here.

First, one possibility that we considered initially was that the enhanced effect might be due to minimizing readers' individual differences with regard to (a) lexical properties of the words used in the sentences (such as the words' frequencies in an individual's lexicon); and (b) world knowledge, which is used in assessing the plausibility of the events described in the sentences. Providing all of this information in the context should make all of the words in the critical sentences highly accessible for all readers. Furthermore, whereas using the plausibility of an event (Did X do Y to Z, or did Z do Y to X?) may often help to arrive at the meaning of a sentence in parallel with syntactic computation (e.g., Culicover & Jackendoff, 2005; Kim & Osterhout, 2005; Kuperberg, Sitnikova, Caplan, & Holcomb, 2003), using plausibility is not helpful in the supportive contexts in the current experiment, because both possible relationships among the entities in the discourse have already been established as equally likely. Thus, people are forced to use syntactic cues in order to understand the meaning of the sentence. By this reasoning, all readers may be using a more similar strategy

(i.e., relying on the syntactic dependency structure) to understand the sentences in supportive contexts.

However, these observations cannot explain the *larger effect size* for the RC complexity effect in supportive contexts. At most, forcing the comprehenders to rely on a more restricted set of cues for interpretation could result in reduced variance in reading times in supportive contexts compared with null contexts. This would result in increased reliability of the effects in supportive contexts, but not their size. (In fact, variance is similar across the supportive and null context conditions in our data.)

A second potential explanation for the enhanced RC complexity effect in supportive contexts is based on working memory considerations, such that the contextually supported RC conditions are perhaps more complex than the null context RC conditions. In particular, there are more events and relationships between entities to keep track of in the supportive context conditions, due to the presence of the information in the context. This is information that the reader needs to keep track of, because s/he is required to answer questions about this context. Perhaps the larger RC complexity effect in a supportive context is thus caused by readers having to remember aspects of the earlier context when reading the RCs. However, contrary to the prediction of this kind of hypothesis, people were actually numerically *better* at answering the comprehension questions following the contextually supported materials in Experiment 2 than after the null context materials (79.1% and 72.9% accuracy in supportive contexts vs. 73.3% and 69.3% in null contexts). If the contextually supported RC conditions were more complex than the null context RC conditions, then we would expect to see suggestions of comprehension question accuracy effects in the reverse direction of what we actually see. Consequently, we consider this possibility unlikely.

The third possibility is that the presence of the contexts changes the comprehenders' expectations for the upcoming structures. For example, it is possible that ORCs are particularly unexpected following contexts used in our materials. If this were the case, then syntactic-surprisal-based accounts (e.g., Gennari & MacDonald, 2008; Hale, 2001; Levy, 2008) would correctly predict an increased complexity effect for supportive context conditions. In order to investigate this possibility, we conducted a completion study for our experimental materials.

We presented our 24 items to 36 participants on Amazon.com's Mechanical Turk, a marketplace interface that can be used for collecting behavioral data over the Internet.[7] Each participant was asked to complete materials corresponding to the initial parts of our materials in Experiment 2, like (6), either with or without the supportive context:

(6) Supportive context (not present in the null context versions of the items):

At the press-conference, a senator and two reporters got into an argument.

The senator attacked one of the reporters and then the other reporter attacked the senator.

Mary: I heard that the reporter that …

Each participant completed a different randomized order of 24 target items (12 in each condition) along with 48 filler items. The results of this study showed no difference in completion rates for SRCs versus ORCs across the context manipulation: People completed 406/432 trials (94.0%) as SRCs in each condition, and they completed 11/432 trials (2.55%) as ORCs in the null context condition versus 8/432 trials (1.85%) in the supportive context condition, with the remainder of the trials being errors or not completed by participants. These results demonstrate a strong bias for SRC completions in each condition, but no difference between supportive and null contexts. Consequently, syntactic surprisal by itself cannot explain the larger ORC complexity effect observed in supportive contexts: ORCs are equally unexpected regardless of the presence of a supportive context. However, a closer look at the content of the completions provides a potential explanation of the larger complexity effect in supportive contexts. Specifically, in the null context condition, the SRC completions were highly variable, but in the supportive context condition, the SRC completions were much more uniform. In particular, 42% of the SRCs *exactly* matched the RC that was presented as the critical SRC completion in the materials for Experiment 2. For example, for (6), many of the continuations were ''attacked the senator . . . ,'' the completion that was present in the corresponding item in Experiment 2. Hence, the enhanced RC complexity effect in supportive contexts in Experiment 2 is plausibly due to people having a strong expectation for the *particular SRC* continuation after reading the context.

The specific nature of this effect cannot be determined based on these data alone. This effect could, for example, be due to expectation-related processes where expectations are more fine-grained than simply expectations for strings of syntactic categories, or it could be due to lexico-syntactic priming of the SRC from the material in the context (see e.g., Branigan, 2007; Pickering & Branigan, 1999 for reviews of the relevant literature). Either an expectation-based or a priming-based account includes several possibilities which would have to be evaluated in future work.

With respect to priming, for example, there are at least two levels at which priming may occur. It is not likely that it is simply a sequence of words (*V the N* vs. *the N V*) that is being primed, because the SRC sequence of words and the ORC sequence of words are both present in the contexts (e.g., ''attacked the senator'' vs. ''the senator attacked''), and they are also balanced for presentation order (and, thus, for distance to the RC) across items. Thus, if it were a structure-less sequence of words that gave rise to the facilitation effect in the reading times, then we would have seen similar facilitation effects in SRCs and ORCs, but the facilitation is greater for SRCs. However, one possibility is that a verb phrase constituent is being primed in the context for the SRCs (e.g., ''attacked the senator'' in [6]), but there is no corresponding syntactic constituent in the context that can be primed for the sequence of words corresponding to the ORC (e.g., ''the senator attacked''). Another possibility is that sequences of syntactic positions are primed together with their associated lexical items. The context contains subject-verb-object sequences (e.g., ''senator attacked reporter,'' ''reporter attacked senator'')—the syntactic position sequence that is present in SRCs—but no object-subject-verb sequences, the syntactic position sequence that is present in ORCs. This imbalance could thus be responsible for the difference in facilitation effects across SRCs and ORCs.

In conclusion, we have observed a robust RC complexity effect in supportive contexts, such that ORCs are processed more slowly than SRCs. This effect replicates the RC complexity effect in English and extends it to supportive contexts. These results are consistent with interference-based and word- or NP-type-based decay working memory accounts of RC complexity (although they constrain the interpretation of the effects of NP type in ORCs, as discussed above). They provide evidence against a discourse-referent-based decay account (Gibson, 1998; Warren & Gibson, 2002). Furthermore, the results of Experiment 2, together with the results of the completion study, provide evidence of fine-grained expectations beyond expectations for sequences of syntactic categories, and/or for lexico-syntactic priming that takes into account constituency information or syntactic positions of the relevant words.

Notes

1. It is worth pointing out that this account (whereby the difficulty of retrieving a previously encountered syntactic dependent is a function of the types of NPs—pronouns, names, definite descriptions—that intervene between the dependents) is not conceptually independent from the decay-based framework where distance is measured in terms of the number of new discourse referents, because the type of NP used to refer to, for example, the embedded subject of an ORC is highly dependent on the preceding context (e.g., a pronoun can only felicitously be used when there is a strongly salient referent in the discourse). We treat these accounts separately throughout the paper but come back to the issue of their conceptual non-independence in the Discussion section after Experiment 2.
2. The current materials and experimental design were developed in parallel with similar work in Japanese (Ishizuka, Nakatani, & Gibson, 2006).
3. We did not include the two context sentences in the graph, because the RTs in these positions were substantially longer than in other positions (on average, 3,784 ms for the first sentence and 4,063 ms for the second sentence) because these sentences were presented as single regions. Including them in the same figure makes it difficult to see differences between conditions for the shorter length regions, including the critical regions. There were no differences between the two conditions in either of the context sentences ($ps > .5$), which is to be expected given that the materials are identical across conditions.
4. In addition to analyzing residual RTs, we also analyzed raw RTs, log RTs, and log residual RTs. Although log RTs and log residual RTs are more normally distributed than raw or residual RTs, Roger Levy has observed that some interactions that are present in analyses of raw or residual RTs can disappear in analyses of log RTs (see the R-lang discussion group, June 10, 2010; http://pidgin.ucsd.edu/pipermail/r-lang/). Consequently, we present detailed analyses of residual RTs. In any case, all of our analyses (raw RTs, residual RTs, log RTs, log residual RTs) resulted in the same statistical patterns.

5. It is possible that some version of an interference-based account would predict a larger RC complexity effect in supportive contexts. Context plausibly makes both of the entities corresponding to the NPs in the RC highly salient. If more salient elements produce greater interference, then we might expect that ORCs in supportive contexts would lead to greater processing complexity than ORCs in null contexts. However, because context makes *both* entities salient, such an account would have to postulate that it is the absolute level of salience of the intervening element that matters, rather than the level of salience of the intervening element relative to that of the to-be-retrieved element. The latter seems more plausible.
6. As in Fig. 1, we did not include the two context sentences in the graph, because the RTs in these positions were substantially longer than in other positions, and including them in the same figure makes it difficult to see differences between conditions for the shorter length regions, including the critical regions. There were no reliable differences between the two context conditions in these initial regions (as expected given that the materials are identical across conditions).
7. There now exist several replications of results from in-lab studies using Amazon's Mechanical Turk, thus establishing the viability of this method for obtaining experimental linguistic data (e.g., Frank, Tily, Arnon, & Goldwater, 2010; Gibson & Fedorenko, in press; Munro et al., 2010).

Acknowledgments

This work was supported by a National Science Foundation Grant 0844472 from the linguistics program to E.G. We would like to thank Michael Frank, Florian Jaeger, Roger Levy, Shravan Vasishth, an anonymous reviewer, and the audience at the CUNY conference on sentence processing in 2007 for helpful comments on this work. Special thanks go to Tomoko Ishizuka, who collaborated on a similar set of studies in Japanese and conducted a pilot experiment in English.

References

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*, 247–264.

Altmann, G. T. M., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition*, *30*, 191–238.

Ariel, M. (1988). Referring and accessibility. *Journal of Linguistics*, *24*, 65–87.

Ariel, M. (1990). *Accessing noun-phrase antecedents*. London: Routledge.

Bates, D., Maechler, M., & Dai, B. (2008). lme4: Linear mixed-effects models using S4 classes. Retrieved December 2011, from http://lme4.r-forge.r-project.org/

Battinich, W. R., & Levine, W. H. (2009). *Production and processing of restrictive relative clauses in pragmatically-appropriate context*. A poster presented at the Annual Meeting of the Psychonomic Society.

Branigan, H. P. (2007). Syntactic priming. *Language and Linguistics Compass*, *1-2*, 1–16.

Branigan, H. P., Pickering, M. J., McLean, J. F., & Cleland, A. A. (2007). Syntactic alignment and participant role in dialogue. *Cognition*, *104*, 163–197.

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

Culicover, P. W., & Jackendoff, R. (2005). *Simpler syntax*. Oxford, England: Oxford University Press.

Fedorenko, E., Gibson, E., & Rohde, D. (2006). The nature of working memory capacity in sentence comprehension: Evidence against domain specific resources. *Journal of Memory and Language*, *54*(4), 541–553.

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

Ford, M. (1983). A method for obtaining measures of local parsing complexity throughout sentences. *Journal of Verbal Learning and Verbal Behavior*, *22*, 203–218.

Frank, M. C., Tily, H., Arnon, I., & Goldwater, S. (2010). *Beyond transitional probabilities: Human learners impose a parsimony bias in statistical word segmentation*. Paper presented at *The Cognitive Science Society,* Portland, OR.

Frisson, S., Rayner, K., & Pickering, M. J. (2005). Effects of contextual predictability and transitional probability on eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 862–877.

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

Gibson, E., Desmet, T., Grodner, D., Watson, D., & Ko, K. (2005). Reading relative clauses in English. *Cognitive Linguistics*, *16*, 313–354.

Gibson, E., & Fedorenko, E. (In press). The need for quantitative methods in syntax and semantics research. *Language and Cognitive Processes*. DOI:10.1080/01690965.2010.515080.

Gibson, E., & Pearlmutter, N. (1998). Constraints on sentence comprehension. *Trends in Cognitive Science*, *2*, 262–268.

Gibson, E., & Wu, I. (In press). Processing Chinese relative clauses in context. *Language and Cognitive Processes*. DOI:10.1080/01690965.2010.536656.

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

Gordon, P., 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.

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*(2), 261–290.

Grodner, D., Gibson, E., & Watson, D. (2005). The influence of contextual contrast on syntactic processing: Evidence for strong-interaction in sentence comprehension. *Cognition*, *95*, 275–296.

Gundel, J., Hedberg, H., & Zacharski, R. (1993). Referring expressions in discourse. *Language*, *69*, 274–307.

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of NAACL* (Vol. 2, pp. 159–166). NAACL.

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

Holcomb, P. J. (1993). Semantic priming and stimulus degradation: Implications for the role of the N400 in language processing. *Psychophysiology*, *30*, 47–61.

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.

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.

Ishizuka, T., Nakatani, K., & Gibson, E. (2006). Processing Japanese relative clauses in context. Paper presented at the 19th Annual CUNY Conference on Human Sentence Processing, CUNY, New York, March, 2006.

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

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*, 122–149.

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processing in reading comprehension. *Journal of Experimental Psychology: General*, *111*, 228–238.

Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, *49*, 133–156.

Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language*, *52*, 205–225.

King, J. & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language*, *30*, 580–602.

Kuperberg, G. R., Sitnikova, T., Caplan, D., & Holcomb, P. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cognitive Brain Research*, *17*, 117–129.

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.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*, 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., Vasishth, S., & Van Dyke, J. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences*, *10*(10), 447–454.

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.

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* (pp. 342–355). Somerville, MA: Cascadilla Press.

Munro, R., Bethard, S., Kuperman, V., Lai, V. T., Melnick, R., Potts, C., Schnoebelen, T., & Tily, H. (2010, June). *Crowdsourcing and language studies: The new generation of linguistic data*. NAACL workshop on Creating Speech and Language Data with Amazon's Mechanical Turk, Los Angeles, CA.

Pickering, M. J., & Branigan, H. P. (1999). Syntactic priming in language production. *Trends in Cognitive Science*, *3*, 136–141.

R Core Development Team (2008). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

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.
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., & Kühn, K. (1995). The processing of locally ambiguous relative clauses in German. *Journal of Memory and Language*, *34*, 499–520.
Schustack, M. W., Ehrlich, S. F., & Rayner, K. (1987). Local and global sources of contextual facilitation in reading. *Journal of Memory and Language*, *26*, 322–340.
Staub, A. (2010). Eye movements and processing difficulty in object relative clauses. *Cognition*, *116*, 71–86.
Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*, 1632–1634.
Tanenhaus, M. K., & Trueswell, J. C. (1995). Sentence comprehension. In J. Miller & P. Eimas (Eds.), *Speech, language, and communication* (pp. 217–262). New York: Academic Press.
Traxler, M. J., Morris, R. K., & Seely, R. E. (2002). Processing subject and object relative clauses: Evidence from eye movements. *Journal of Memory and Language*, *47*, 69–90.
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.
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 Petten, C., & Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Memory & Cognition*, *18*, 380–393.
Vasishth, S., & Lewis, R. (2006). Argument-head distance and processing complexity: Explaining both locality and anti-locality effects. *Language*, *82*(4), 767–794.
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.

Appendix

Each of the 24 items had two versions, as shown in (1). The second version of each item can be generated in a similar way.

1. At the press-conference a senator and two reporters got into an argument. The senator attacked one of the reporters and then the other reporter attacked the senator.

Mary: I heard that the reporter that attacked the senator admitted to making an error.

John: I am not sure about that. I heard that the reporter that the senator attacked admitted to making an error.

At the press-conference a reporter and two senators got into an argument. The reporter attacked one of the senators and then the other senator attacked the reporter.

Mary: I heard that the senator that attacked the reporter admitted to making an error.

John: I am not sure about that. I heard that the senator that the reporter attacked admitted to making an error.

2. During the interview a newscaster and two musicians had a brief discussion. The newscaster insulted one of the musicians and then the other musician insulted the newscaster.

Mary: I heard that the musician that insulted the newscaster left the building after the interview.

John: I am not sure about that. I heard that the musician that the newscaster insulted left the building after the interview.

3. In the store a customer and two cashiers talked about recent events. The customer thanked one of the cashiers and then the other cashier thanked the customer.

Mary: I heard that the cashier that thanked the customer supported the new bill about immigration rules.

John: I am not sure about that. I heard that the cashier that the customer thanked supported the new bill about immigration rules.

4. After the lecture a scientist and two interns went over the problem set. The scientist confused one of the interns and then the other intern confused the scientist.

Mary: I heard that the intern that confused the scientist worked at a famous lab at Harvard University.

John: I am not sure about that. I heard that the intern that the scientist confused worked at a famous lab at Harvard University.

5. After the incident an officer and two detectives talked about possible suspects. The officer approached one of the detectives and then the other detective approached the officer.

Mary: I heard that the detective that approached the officer had a good record in solving similar cases

John: I am not sure about that. I heard that the detective that the officer approached had a good record in solving similar cases

6. In the office a secretary and two co-workers got along very well. The secretary welcomed one of the co-workers and then the other co-worker welcomed the secretary.

Mary: I heard that the co-worker that welcomed the secretary brought some flowers to the office.

John: I am not sure about that. I heard that the co-worker that the secretary welcomed brought some flowers to the office.

7. Over the weekend a dentist and two pediatricians talked about various medications. The dentist called one of the pediatricians and then the other pediatrician called the dentist.

Mary: I heard that the pediatrician that the dentist called left a message about the recommended dosage.

John: I am not sure about that. I heard that the pediatrician that called the dentist left a message about the recommended dosage.

8. Before the surgery a physician and two neurologists had a long discussion. The physician helped one of the neurologists and then the other neurologist helped the physician.

Mary: I heard that the neurologist that the physician helped worked at MGH for the last ten years.

John: I am not sure about that. I heard that the neurologist that helped the physician worked at MGH for the last ten years.

9. During the event an author and two critics argued about the point of a book. The author annoyed one of the critics and then the other critic annoyed the author.

Mary: I heard that the critic that the author annoyed had strong opinions about many things.

John: I am not sure about that. I heard that the critic that annoyed the author had strong opinions about many things.

10. Before the class a teacher and two students went over the homework. The teacher greeted one of the students and then the other student greeted the teacher.

Mary: I heard that the student that the teacher greeted gave an interview to the school newspaper recently.

John: I am not sure about that. I heard that the student that greeted the teacher gave an interview to the school newspaper recently.

11. After the interview a politician and two journalists argued about the new law. The politician criticized one of the journalists and then the other journalist criticized the politician.

Mary: I heard that the journalist that the politician criticized left the room around 3pm.

John: I am not sure about that. I heard that the journalist that criticized the politician left the room around 3pm.

12. At the convention a researcher and two inventors displayed innovative technology. The researcher praised one of the inventors and then the other inventor praised the researcher.

Mary: I heard that the inventor that the researcher praised patented several inventions over the last five years.

John: I am not sure about that. I heard that the inventor that praised the researcher patented several inventions over the last five years.

13. At the debate a congressman and two governors argued about the national election. One of the governors antagonized the congressman and then the congressman antagonized the other governor.

Mary: I heard that the governor that antagonized the congressman apologized for being too aggressive.

John: I am not sure about that. I heard that the governor that the congressman antagonized apologized for being too aggressive.

14. Earlier this month an interpreter and two ambassadors planned a trip. One of the ambassadors contacted the interpreter and then the interpreter contacted the other ambassador.

Mary: I heard that the ambassador that contacted the interpreter lived in Africa for many years.

John: I am not sure about that. I heard that the ambassador that the interpreter contacted lived in Africa for many years.

15. After the meeting an administrator and two managers examined the accounting books. One of the managers questioned the administrator and then the administrator questioned the other manager.

Mary: I heard that the manager that questioned the administrator had a problem with the company.

John: I am not sure about that. I heard that the manager that the administrator questioned had a problem with the company.

16. During the lecture an anthropologist and two historians discussed the article. One of the historians challenged the anthropologist and then the anthropologist challenged the other historian.

Mary: I heard that the historian that challenged the anthropologist published a famous book on the same topic.

John: I am not sure about that. I heard that the historian that the anthropologist challenged published a famous book on the same topic.

17. Before the meeting an instructor and two counselors talked about teaching methods. One of the counselors offended the instructor and then the instructor offended the other counselor.

Mary: I heard that the counselor that offended the instructor regretted the comment after the presentation.

John: I am not sure about that. I heard that the counselor that the instructor offended regretted the comment after the presentation.

18. In the boardroom a manufacturer and two analysts talked about the deal. One of the analysts consulted the manufacturer and then the manufacturer consulted the other analyst.

Mary: I heard that the analyst that consulted the manufacturer was involved in a scandal not long ago.

John: I am not sure about that. I heard that the analyst that the manufacturer consulted was involved in a scandal not long ago.

19. At the laboratory a mathematician and two engineers talked about the project. One of the engineers impressed the mathematician and then the mathematician impressed the other engineer.

Mary: I heard that the engineer that the mathematician impressed received an award at a recent conference.

John: I am not sure about that. I heard that the engineer that impressed the mathematician received an award at a recent conference.

20. A week ago a writer and two artists met at the museum. One of the artists upset the writer and then the writer upset the other artist.

Mary: I heard that the artist that the writer upset left the museum in a bad mood.

John: I am not sure about that. I heard that the artist that upset the writer left the museum in a bad mood.

21. At the gym a gymnast and two wrestlers were training for an upcoming meet. One of the wrestlers observed the gymnast and then the gymnast observed the other wrestler.

Mary: I heard that the wrestler that the gymnast observed attended the college on a scholarship.

John: I am not sure about that. I heard that the wrestler that observed the gymnast attended the college on a scholarship.

22. Before the reunion a chef and two caterers talked about the food. One of the caterers frustrated the chef and then the chef frustrated the other caterer.

Mary: I heard that the caterer that the chef frustrated was famous for his mushroom soup recipe.

John: I am not sure about that. I heard that the caterer that frustrated the chef was famous for his mushroom soup recipe.

23. At a construction site, a bricklayer and two carpenters worked near one another. One of the carpenters assisted the bricklayer and then the bricklayer assisted the other carpenter.

Mary: I heard that the carpenter that the bricklayer assisted had twenty years of experience.

John: I am not sure about that. I heard that the carpenter that assisted the bricklayer had twenty years of experience.

24. Before the press-conference a lawyer and two legislators talked among themselves. One of the legislators cautioned the lawyer and then the lawyer cautioned the other legislator.

Mary: I heard that the legislator that the lawyer cautioned was reported to have considered taking bribes.

John: I am not sure about that. I heard that the legislator that cautioned the lawyer was reported to have considered taking bribes.